Structural and biochemical studies between the T-cell receptor p14 and MHC class I complexes loaded with the LCMV derived antigen gp33
and
The structure and function of the catalytic domain of *Streptococcus pneumoniae* major autolysin LytA

Master Thesis

submitted by

Bakk. techn. Markus Bernhard TOMEK

to be conferred with the academic degree "Diplom-Ingenieur"

accomplished at the

Department of Microbiology, Tumor- and Cell Biology at the Karolinska Institutet, Stockholm

under the supervision of
Ass. Prof. Adnane ACHOUR, PhD

and in cooperation with the

Department of Biotechnology at the
University of Natural Resources and Applied Life Sciences, Vienna

under the supervision of
A.o. Univ. Prof. Dr. Florian RÜKER

Vienna and Stockholm, May 2011

For my Dad.

Acknowledgement

First of all I would like to thank my supervising professor Florian Rüker from the BOKU for making it possible for me to perform my Master thesis abroad. Thank you for your supporting and friendly attitude!

I am deeply thankful to Adnane Achour too, who gave me the opportunity to work in his group at the Karolinska Instituted in Sweden. I enjoyed my freedoms in the lab and learned a lot, not only about my projects, from you. In the same way I thank Tatyana Sandalova who taught me almost everything I know about crystallography. Another big tack så mycket to Eva Allerbring - with you it was always a pleasure to work in the lab and spending some hours in front of the Biacore. Thanks for answering so many of my questions, helping me in the lab and also for spending some time outside of the lab, my dear lab and former lab colleagues, Adil, Hannes, Tim, Marcos and Evi!

Since the thesis is one of the last steps I have to take before finishing my master program, it is also a good time to reflect about the last years. Therefore I would like to say thank you to all my colleagues and friends at the Boku and especially those from IAESTE Boku!

For your support all over the years and in various ways I say thanks to all members of my big family!

Markus Bernhard Tomek
May 2011

"Borders? I've never seen one, but I've heard they exist in some people's minds."
Thor Heyerdahl

Abbreviations

°C	degree Celsius
CDR	Complementarity Determining Region
ΔC_P	specific heat capacity
ΔG	free Gibbs energy
ΔH	free Enthalpy
ΔS	free Entropy
DTT	Dithiothreitol
H-2D^b	H-2 class I histocompatibility antigen
HLA	Human Leucocyte Antigen
IEC	Ion Exchange Chromatography
IPTG	Isopropyl-1-thio-β-D-galactopyranoside
K	Kelvin
k_a	rate of formation or on-rate
K_D	equilibration dissociation constant or binding constant
k_d	rate of dissociation or off-rate
mβ2m	murine β-2-microglobulin
MHC	Major Histocompatibility Complex
MWCO	Molecular Weight Cut Off
pMHC	peptide MHC
PMSF	Phenylmethylsulfonyl fluoride
R	universal gas constant, 8.314472 [J K^{-1} mol^{-1}]
RU	Resonance Unit
SDS-PAGE	Sodium Dodecyl Sulfate – Poly Acrylamid Gel Electrophorese
SEC	Size Exclusion Chromatography
SPR	Surface Plasmon Resonance
T	Temperature in Kelvin
TCR	T Cell Receptor

Abstract

For this Master thesis two independent projects were performed. The first project dealt with the detailed recognition by the T-cell receptor (TCR) of peptide and altered peptide ligands (APL) loaded to Major Histocompatibility Complex class I molecules (pMHC class I). The crystalstructure of the catalytic domain of *Streptococcus pneumoniae* major autolysin LytA was solved and its function was described during performing the second project.

The molecular basis of recognizing pMHCs by TCRs is known, however no predictions on the outcome of the recognition can presently be made when investigating escape mutants and APL. In this study the lymphocytic choriomeningitis virus (LCMV) derived antigen gp33 and variants with single amino acid mutations at the main TCR recognition position 4 (gp33-Y4F, gp33-Y4S and gp33-Y4A) were analysed. Binding affinities were determined to 8.9 µM and 54.5 µM for gp33 and gp33-Y4A respectively. No binding was determined between the TCR p14 and the APL gp33-Y4F or gp33-Y4S. The thermodynamic signatures revealed an enthalpically driven recognition of gp33 driven by the formation of several hydrogen bonds around position 4 of the peptide and an entropically driven partial recognition of the APL gp33-Y4A. In the case of gp33-Y4A hydrogen bond interactions are found around position 1 forming a secondary recognition hotspot, but not around position 4. The loss of several hydrogen bonds and the expulsion of water molecules from the peptide binding cleft are possibly responsible for the large entropic term for this interaction.

The diametrically opposed recognition of gp33 and gp33-Y4A can structurally be explained, the non-recognition of the viral escape mutant gp33-Y4F and the APL gp33-Y4S with even smaller changes at position 4 than compared to gp33-Y4A, remains unclear. Thus, at the present time, no clear-cut conclusion can be drawn regarding the strategy used by TCRs in order to recognize their cognate ligands.

The catalytic domain of *Streptococcus pneumoniae* major autolysin LytA is an N-acetylmuramoyl-L-alanine amidase breaking peptidoglycan strands. Crystallization conditions were established at 0.1 M HEPES p*H* = 6.8-7.1, 1.0 M Lithiumchloride

and 9-11 % PEG 6000. Crystals diffracted to 1.0 Å resolution and the solved crystal structure revealed a typical α/β hydrolase fold with six mixed β-sheets and seven flanking α-helices. The conserved peptidoglycan binding groove, containing a zinc atom in the active site, could be determined and a mechanism for catalysis within the catalytic domain was proposed. The position of the catalytic domain in relation to the choline binding domain and the mechanisms of activation of the autolysin still remain elusive and need to be further investigated.

The presented structure of the catalytic domain is a major step forward in understanding the pneumococcal autolysin. Future works should investigate the activation of the autolysin which can lead to the development of novel antibacterial drugs.

Zusammenfassung

Zwei Projekte wurden unabhängig voneinander zum Verfassen dieser Diplomarbeit durchgenommen. Das erste Projekt behandelte die Fragestellung wie T-Zell-Rezeptoren (TCR) mit Peptiden oder designten Peptiden (APL) geladene Haupthistokompatibilitätskomplex (MHC Klasse 1) Moleküle, erkennt. Während des zweiten Projektes wurde die Kristallstruktur der katalytischen Untereinheit von *Streptococcus pneumoniae* Autolysin LytA aufgelöst und deren Funktionsweise beschrieben.

Der generelle Ablauf zwischen der Erkennung von pMHCs durch TCRs ist bekannt, allerdings können keine Vorhersagungen auf den Ausgang gemacht werden, wenn modifizierte Peptide, wie sie z.B. als Fluchtmutanten natürlich vorkommen, von den MHC Molekülen präsentiert werden. Hier wurde ausgehend vom lymphozytären-Choriomeningitis-Virus Antigen gp33 und an vierter Aminosäureposition veränderten Peptiden (gp33-Y4F, gp33-Y4S und gp33-Y4A) die T-Zell Interaktionen untersucht. Bindungsaffinitäten wurden dabei auf 8.9 µM für gp33 und 54.5 µM für gp33-Y4A gemessen. Keine Bindung trat auf wenn gp33-Y4F oder gp33-Y4S dem T-Zell-Rezeptor p14 präsentiert wurden. Durch die Ausbildung von mehreren Wasserstoffbrückenbindungen nahe dem T-Zell-erkennungs Hot-spot bei Peptidposition 4, konnte eine enthalpisch getriebene Erkennung von gp33 erklärt werden. Eine verminderte Erkennung zeigte sich auch wenn gp33-Y4A präsentiert wurde. Allerdings wurden hier keine Wasserstoffbrückenbindungen bei Position 4 sondern um Position 1, die hier als zweiter Erkennungs Hot-spot auftritt, gefunden. Die thermodynamische Signatur zeigte eine entropisch getriebene Erkennung, wobei sich der entropische Term durch das Nichtzustandekommen von Wasserstoffbrückenbindungen bei Position 4 und dem Ausschluss von Wassermolekülen aus der Peptidbindungsspalte erklären lässt.

Die unterschiedliche Erkennung von gp33 und gp33-Y4A kann strukturell erklärt werden, die Nichterkennung von gp33-Y4F und gp33-Y4S, welche nur kleine strukturelle Veränderungen im Gegensatz zur Wildtypenform gp33 aufweisen, bleibt ungeklärt. Daher können nach wie vor keine grundlegenden

Mechanismen bekanntgegeben werden, die T-Zell-Rezeptoren verwenden um ihre zugehörigen pMHCs zu erkennen.

Die katalytische Untereinheit von *Streptococcus pneumoniae* Autolysin LytA ist eine Amidase, welche Peptidoglycanstränge zwischen N-Acetylmuramin Säuren und L-Alanin aufbrechen kann. Kristalle bildeten sich in 0.1 M HEPES pH = 6.8-7.1, 1.0 M Lithiumchlorid und 9-11 % PEG 6000 und konnten bis zu einer Auflösung von 1.0 Å gebeugt werden. Die so erhaltene Kristallstruktur zeigte eine typische α/β Hydrolase Faltung mit sechs β-Faltblättern und sieben umgebenden α-Helices. Die konservierte Peptidoglycanbindungsstelle beinhaltet ein Zinkatom in der aktiven Seite, weiters konnte ein Mechanismus, der zum Aufbrechen des Peptidoglycanstranges führt, vorgeschlagen werden. Die Position der katalytischen Untereinheit in Relation zum Gesamtprotein sowie die molekularen und strukturellen Vorgänge, die zur Aktivierung des Proteins führen, bleiben vorerst unbekannt.

Trotzdem ist die vorliegende Kristallstruktur ein wichtiger Schritt um die Funktionsweise des Autolysins zu verstehen. Sollte die Funktionsweise der Aktivierung des Proteins in Erfahrung gebracht werden, könnte dieses Wissen in die Entwicklung von neuen antibakteriellen Medikamenten und Impfstoffen führen.

1 Introduction

1.1 The human immune system

The immune system is crucial to human survival. Without having a defence mechanism against pathogens (bacteria, viruses, fungi and parasites) serious infections can easily occur. Beside of physical and chemical barriers such as the skin and the mucosal surfaces or the acidic environment in the stomach, humans have two methods to react against infections, the innate and the adaptive immune system.

1.1.1 The innate and the adaptive immune system

The innate immune system shows a fast and effective response in the recognition and destruction of most infections at an early stage. Conserved components among various microorganisms, which are not made by the host, are recognized and destroyed by phagocytic cells (neutrophils, macrophages, natural killer cells) and via the complement system. Conserved components such as microbial surface proteins (e.g. lipopolysaccharides from gram negative bacteria) or viral double stranded RNA are recognized by pattern recognition receptors (e.g. Toll like receptors) of phagocytes. After recognition, phagocytes use a combination of degrading enzymes and antimicrobial peptides to kill the invading pathogens. The complement system can target not recognized pathogens and help phagocytes to destroy them (opsonization). The complement cascade can also form a transmembrane channel causing osmotic lysis of the host cell. The innate immune response is also required to activate the adaptive immune system giving a more specialized immune response to a specific infection.

The adaptive immune system gives a highly specialized defence against one pathogen. It is necessary if the innate immune system is not able to get hold of an infection. The mechanisms of the adaptive immune system can also provide a long term immunological memory to one pathogen after first infection (acquired immune system). Activation of the adapted immune response is mediated through antigens which can be any substance from foreign origin.

There are two ways for the adaptive immune system to react; the antibody response (humoral immune response) and the T-cell mediated immune response.

They are carried out by B-lymphocytes and T-lymphocytes respectively. Lymphocytes belong to the white blood cells and have their origin in undifferentiated hematopoietic stem cells, located in the bone marrow. B-lymphocytes differentiate in the bone marrow to B-cells and secrete antibodies which specifically bind to the foreign antigens making them unable to bind to their receptors. Antibody binding also marks invading pathogens which are then recognized by phagocytes.

T-cell mediated immune response is based on the direct interaction between activated T-lymphocytes and foreign antigens which are presented on the surface of a host cell. T-lymphocytes originate in the bone marrow and develop in the thymus to T-cells. Upon recognition of a foreign antigen by the T-cell it can either directly destroy the infected cell as a cytotoxic T-cell (CTL) or the T-cell can develop to T-helper cells (T_H1 or T_H2). While T_H1 cells can activate macrophages to destroy the pathogens by phagocytosis, T_H2 cells can activate B-cells leading to the production of specific antibodies against the invading pathogen.[9, 58]

1.1.2 Major Histocompatibility Complex class I (MHC class I) antigens and peptide presentation

During adaptive immune response antigens are recognized by highly variable receptor molecules (immunoglobulines) present on B-cells and T-cells. T-cells can detect the presence of intracellular pathogens because peptide fragments from the pathogen are displayed on the surface of the infected host cell. The delivery of the antigens is carried out by specialized host-cell glycoproteins called Major Histocompatibility Complex (MHC) molecules. Antigens derived by intracellular processing of viral, bacterial and endogenous proteins are presented by MHC class I molecules (MHC class I). MHC class I molecules are expressed on all nucleated cells in the human body and present the cytosolic derived antigens to $CD8^+$ T-cells, also known as cytotoxic T-lymphocytes (CTL). Cytotoxic T-cells initiate direct killing of the infected antigen presenting cell.

MHC class II molecules are expressed on B-cells, macrophages and dendritic cells. They bind exogenous peptide antigens within endosomal compartments and present them to $CD4^+$ helper T-cells (T_H). Upon interaction macrophages or B-cells are activated.

The Major Histocompatibility Complex is found on chromosome 6 in humans and on chromosome 17 in mice. Many different genes involved in the immune system are found and have two special features. Firstly, the MHC is polygenic which means that multiple similar genes encode the MHC class I heavy chains and MHC class II α and β chains (isotypes). Secondly, the MHC is polymorphic, which is the presence within the population of multiple alternative forms of a gene (allotypes).

Human MHC class I, also called human leukocyte antigen (HLA), can be classified into six isotypes and various allotypes indicated in brackets; HLA-A (218), HLA-B (439), HLA-C (96), HLA-E (4), HLA-F (1) and HLA-G (6).[9, 58]

MHC class I molecules are molecular complexes consisting of a membrane-anchored heavy chain (~45 kDa), a non-covalently attached β_2-microglobulin (β_2m) unit (~12 kDa) and a 8-10 amino acids long peptide. The polymorphic heavy chain has three extracellular domains (α_1, α_2, α_3), a transmembrane segment and a cytoplasmic tail. Each of the domains α_2, α_3 and β_2m contain two cysteine residues involved in the formation of disulphide bonds. The β_2m and α_3 domains have a folded structure that resembles the one of immunglobulins. The α_1 and α_2 domains form two α-helices topping a sheet of eight antiparallel β-strands. This high polymorphic structure forms a cleft where the peptide antigen can bind. The position of β_2m is well conserved in all known structures of MHC class I molecules. The α_3 domain does not have a major impact on peptide binding and is variable among MHC class I molecules. (Figure 1)[58, 62]

Each α-helical domain contributes to half of the eight stranded β-sheet floor of the peptide binding cleft and of the α-helical wall. The peptide binding cleft of MHC class I molecules is closed at both ends, restricting the size of bound peptides to 8-10 amino acids. Within the peptide binding cleft, the most variable residues point into the groove and up from the tops of both helices to confer unique peptide and TCR binding properties to each MHC class I molecule. The majority of these variable residues are located in the central portion of the cleft. Clusters of highly conserved residues occur at the ends of the binding groove to stabilize and anchor the peptide termini. The relative arrangement of the polymorphic residues within the peptide binding cleft creates pockets that may accommodate predominant amino acid side chains of the peptide. In this way the peptide is anchored on the

MHC class I molecule. In MHC class I structure, HLA-A2, six major pockets were identified and labelled A through F. (Figure 1)[62, 63]
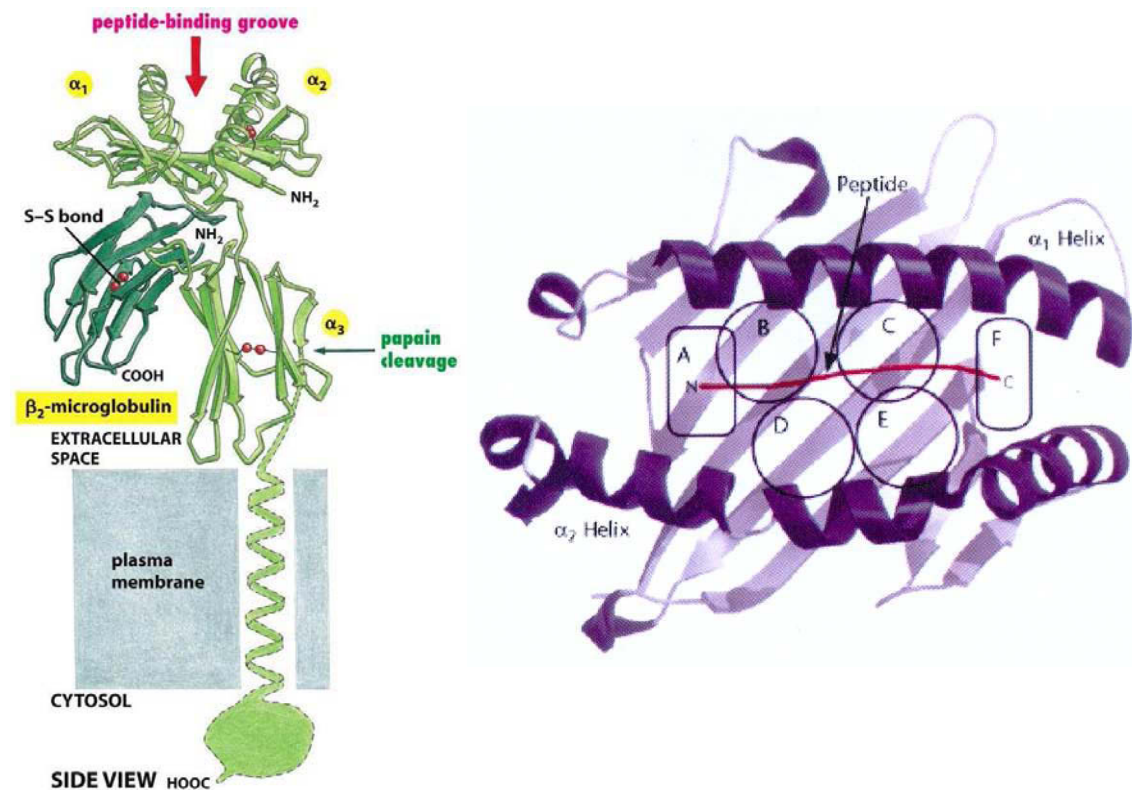
Figure 1 Left The MHC class I molecule consists of three extracellular domains (α_1, α_2, α_3), a transmembrane segment and a cytoplasmic tail. A non-covalently bound β_2-microglobulin (β_2m) unit is attached. The α_1 and α_2 domains form two α-helices topping a sheet of eight antiparallel β-strands creating the peptide-binding groove.[9] **Right** Inside the peptide binding groove the six major peptide binding pockets, labelled A to F, can be seen.[62]

The A and F pockets of MHC class I molecules are used to bind the N and C termini of the peptides independently from peptide sequences. This orientation (amino to carboxyl) of the peptide is fixed for all MHC class I molecules due to the unique chemical environment of these pockets. Hydrogen bonds between main chain atoms along the peptide and MHC class I residues from the peptide binding site provide a second component to the sequence independent binding interactions.

Sequence dependent binding is essential for the specificity of the peptides presented in the peptide binding groove. The peptides bound to MHC class I proteins have allele-specific binding motifs characterized by strong preferences for a few side chains at some positions in the peptide and wide tolerance for many side chains at other positions. The pockets created by the arrangement of

polymorphic residues complement a small number of specific amino acids from the peptide, thereby selecting the peptides which are allowed to bind. Residues that protrude into specificity pockets are termed anchors. MHC class I molecules preferentially bind peptide binding sequence motifs defined by the second (P2) and the last residue (P8, P9 or P10) in humans and by the fifth (P5) and the last residue (Pω) in mice. Additionally, secondary pockets can enhance the affinity of a particular peptide.[62]

In order for an antigenic peptide to be presented by MHC class I molecules, several steps have to be followed. This multi-step process is called antigen processing and starts with the uptake of a virus via endocytosis and delivery to the endosome in dentritic cells. In the endosome the virus fuses with the membrane releasing viral RNA into the cytosol, where replication and translation of viral RNA and proteins start. Proteolysis of viral protein molecules takes place in the proteasome, releasing viral peptides. These peptides are delivered into the lumen of the endoplasmatic reticulum (ER) via transporter for antigen processing (TAP) translocase. TAP is one subunit of the MHC class I loading complex, which resembles the MHC class I molecules. Binding of the antigenic peptide to the α-chain stabilizes the complex formed of α-chains α_1, α_2, α_3 and $\beta_2$microglobulin. The peptide MHC class I complex is then transferred to the cell surface via the standard secretory route using the Golgi apparatus where it is presented to the immune system. (Figure 2)[9]
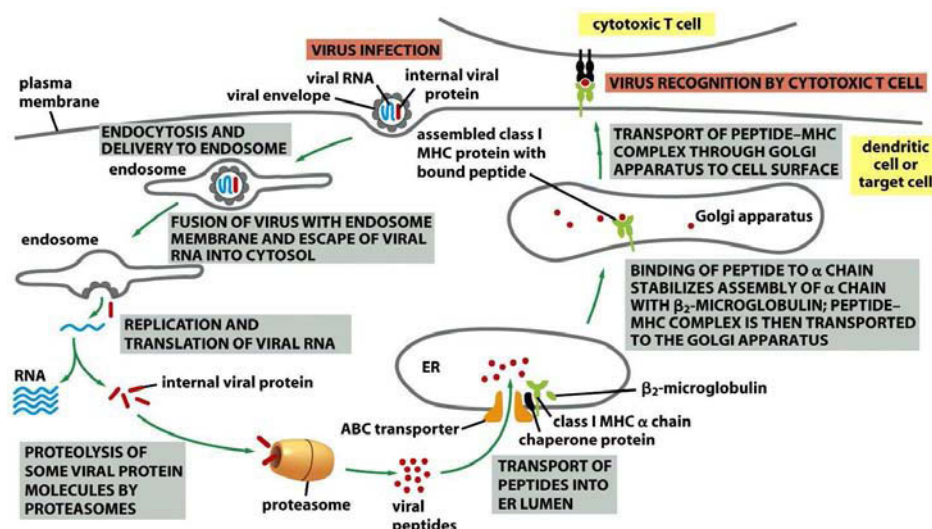

Figure 2 Antigen processing upon viral infection leads to the presentation of an antigenic peptide through MHC class I molecules.[9]

1.1.3 T-Cell Receptor (TCR) and the recognition of peptide-MHC class I complexes

αβ T-cell receptors (TCR) consist of one α and one β chain, linked together via disulphide bonds. The extracellular part is folded into two immunoglobulin-like domains, a constant (C) and a variable (V) domain. Both variable domains (Vα and Vβ) have three complementarity-determining regions (CDR1, CDR2 and CDR3) building the antigen binding site. CDRs are hypervariable loops located at the tip of the TCR that recognize the antigenic surface of the peptide-MHC complex. TCRs have a cleft between the two CDR3 loops, where peptide residues can protrude and interact with the so called "functional hot spot" of the TCR. These hot spots allow the TCR to finely discriminate between the peptides presented.[9, 58, 62]

The formation of the TCR/pMHC-I complex is the central event of antigen recognition in the adaptive immune response. (Figure 3) The highly specific TCR is genetically restricted to recognize the MHC class I with bound antigenic peptide complex (MHC-I restriction). Allotypes of MHC class I molecules can differ from each other by only a single amino acid or by even more than 30 amino acids. Those small variations can have a profound impact on TCR recognition. Some generalizations regarding the interactions between TCR and pMHC-I complexes were made. i) The interaction is of very weak affinity and is characterized by a slow association and fast dissociation rate. ii) The TCR interfaces the pMHC-I complex in a diagonal orientation parallel to the β-sheets which form the pedestal of the pMHC binding cleft. iii) Relative to the peptide, the MHC-I dominates the interaction with the TCR. iv) The CDR1 and CDR2 loops interact primarily with the MHC α-helices α_1, α_2, or the ends of the peptide, while both CDR3 loops contact the peptide. v) The interaction is characterized by induced fit of CDR3 loops. vi) The interaction is enthalpically driven, consistent with the CDR loops being stabilized upon ligation. However, a few TCR/pMHC interactions are found to be entropically driven.[62, 64, 65, 68]
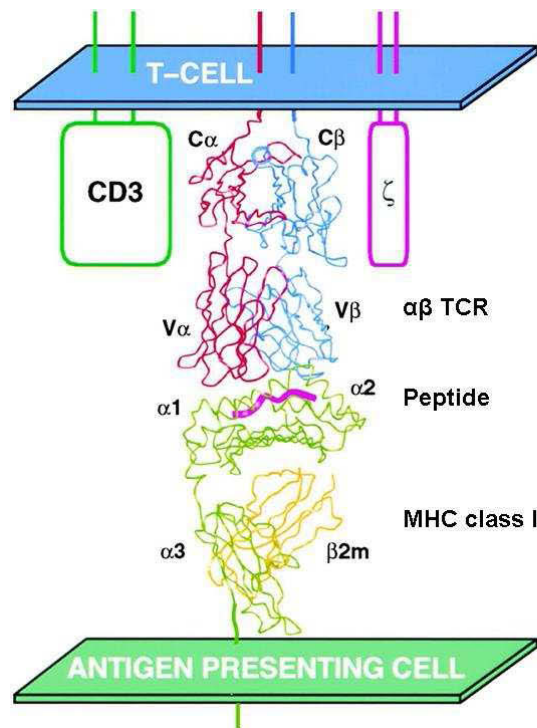
Figure 3 The TCR/pMHC complex. The TCR is composed of two membrane anchored polypeptides, α and β (red and blue) that each contain one constant (C) and one variable (V) domain. Both variable domains contain the hypervariable CDR1, CDR2 and CDR3 loops which interact with the antigenic peptide (purple) and the helices α_1 and α_2 (green) from the MHC class I molecule. MHC class I molecules contain helices α_1, α_2 and α_3 (green) and the non-covalently attached β_2-microglobulin (β_2m) unit (yellow). The multiunit signalling apparatus CD3 and the ζζ heterodimer are indicated.[65]

Upon T-cell recognition the information about the antigenic peptide must be transmitted into the CTL in order to response and kill the infected cell. Signals are transmitted via the conserved multisubunit signalling apparatus CD3, consisting of the αβ TCR heterodimer, a CD3εγ heterodimer a CD3εδ heterodimer and a CD3ζζ homodimer.[9, 64]

However, the activation (or lack of activation) of T-cells seems to be determined by kinetic and thermodynamic parameters inherent to TCR/pMHC recognition. Antagonistic peptides inhibit T-cell activation and weak agonists elicit some of the response. Crystal structures do not explain the large biological differences that can arise when altered peptide ligands (APLs) are bound. Kinetic data loosely correlate slower dissociation rates with increased agonism, but exceptions are found too. The structural basis of how TCRs distinguish between agonist, weak agonist and antagonist ligands remains elusive.[65]

1.1.4 Thermodynamics of T-cell recognition

Immunological recognition is a dynamic process. Therefore biophysical studies are necessary for understanding T-cell activity in addition to time and space averaged structural insights. The most popular method is the surface plasmon resonance (SPR) technique, followed by isothermal titration calorimetry (ITC) and ultracentrifugation (UC). It must be taken into consideration that T-cell signalling is a very complex process where many co-stimulatory molecules are involved. To relate biophysical data from simplified model systems with biological activity is therefore difficult.

Binding affinity (K_D) values can be calculated as the ratio of dissociation (k_{off}) and association (k_{on}) rate constants when it is assumed that reaction binding follows a one-step mechanism. TCR/pMHC interactions have low binding affinities (K_D = ~0.1-500 µM) and are characterized by slow association and fast dissociation reactions. Fast dissociation rates range between ~0.01-5 s^{-1} and correspond to half-lifes of 70-0.1 s. Half-lifes have been found to correlate with the potency of T-cells, but exceptions have also been found. When assuming the TCR/pMHC binding reaction as a first order reaction, the half-life is calculated as

$$t_{1/2} = \ln2/k_{off}$$

Despite low binding affinities and fast dissociation reactions, high specificity is observed in T-cell signalling. Three models are proposed to explain how T-cells distinguish between low concentrations of antigenic peptides (non-self peptides) and high concentrations of self-peptides. 1) Allostery: conformational changes for signal transmission across the cell membrane. 2) Kinetic proofreading: TCR/pMHC binding is coupled to another biochemical reaction such as phosphorylation. 3) Cross-linking: TCR and signalling molecules are relocated to form a supramolecular complex, the immunological synapse. In all models the dissociation competes with the rate of conformational changes and the formation of a supramolecular complex. A threshold time of productive TCR/pMHCs interactions which is needed for the activation of T-cells is also discussed.[65, 67, 68]

Thermodynamic analyses of TCR/pMHC binding give an insight about changes occurring during these interactions. Changes can be conformational adjustments or reductions in flexibility, the amount of buried hydrophobic/hydrophilic surfaces,

the formation of hydrogen bonds and salt bridges, and incorporation or release of water molecules or ions. To different extents these changes are reflected in the enthalpic (ΔH) and entropic (ΔS) terms of the interaction, present in the Gibbs free energy equation;

$$\Delta G = -RT\ln K_D = \Delta H - T\Delta S$$

Changes in binding enthalpy are thought to be a composition of individual enthalpies arising from hydrogen and van der Waals bond rearrangements, conformational changes and proton exchanges with the buffer. Changes in the entropic term may be a sign of the flexibility of CDR loops and the displacement of ordered water molecules. But there are no clear correlations seen between structural or dynamical features and TCR-binding enthalpy or entropy changes. A clearer parameter for observing changes in conformation is the heat capacity change (ΔC_p). Heat capacity changes are strongly influenced by changes in solvation. Burial of hydrophilic surface contributes positively while burial of hydrophobic surface contributes negatively to ΔC_p.[65, 68] Van 't Hoff analysis, which is the analysis of measurements of affinities as a function of temperature, is used to extract estimates of ΔH, ΔS, and ΔC_P for the binding reaction between the pMHC complex and the TCR.[71] A non-linear form of the Van 't Hoff equation is

$$\Delta G = \Delta H_{T0} - T\Delta S_{T0} + \Delta C_p (T - T_0) - T\Delta C_p\ln(T - T_0)$$

which is the same as

$$Y = A - X*(B - C+C*\ln(X / 298.15)) - C*298.15$$

where Y = ΔG, X = T, A = ΔH_{T0}, B = ΔS_{T0}, C= ΔC_p and T_0 = 298.15K.

1.1.5 Lymphocytic Choriomeningitis Virus (LCMV) derived immunogenic antigen gp33 and its recognition by the αβ TCR p14

Lymphocytic choriomeningitis virus (LCMV) is a noncytopathic RNA virus that infects cells of the immune system in its natural host, the mouse. It suppresses cellular and humoral immune responses and causes persistent infections. It can be transmitted to humans from direct contact with murine excreta, but cases are rare and usually non-fatal. Murine H-2^b MHC class I molecules generate a CD8$^+$ CTL response mainly directed towards three immunodominant epitopes. More than 50 % of the total LCMV-specific CTL activity is directed towards gp33 (KAVYNFATM), derived from the viral glycoprotein 1. The full agonist gp33 induces complete CD8$^+$ T-cell response. Crystal structures of H-2D^b/gp33

revealed that the side chains of the tyrosine (p4Y) and phenylalanine (p6F) peptide residues project out of the peptide binding cleft, likely acting as main TCR contacts. (Figure 4, 5)) Upon CTL selection pressure, p4Y can be mutated to a phenylalanine (pY4F) lowering the affinity of the TCR p14 to H-2D^b/Y4F by a 100-fold and abolishing T-cell mediated immune responses against the LCMV infection. Similarly, substitution of p4Y to a serine (pY4S) reduced significantly recognition by p14. In contrast, H-2D^b/Y4A in which p4Y was substituted to an alanine was still recognized by the TCR p14 although it carried a more significant structural alteration when compared to both Y4F and Y4S.[7, 20]
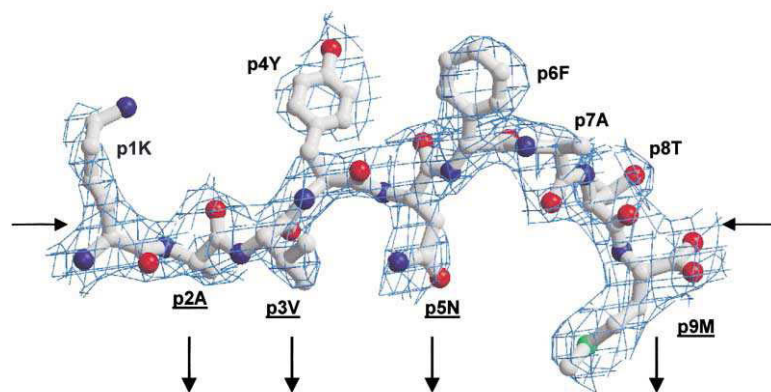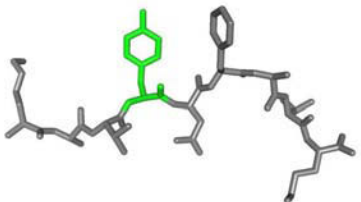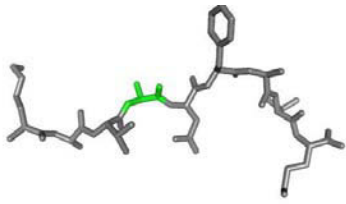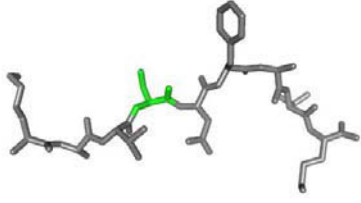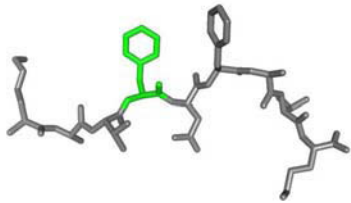
Figure 4 The LCMV derived immunodominant peptide gp33 (KAVYNFATM) is depicted with its N termini to the left and its C termini to the right, illustrating the anchor positions (underlined and vertical arrows). The parts of the peptide that protrude above the surface of the complex are indicated by the horizontal arrows on each side of the peptide. Amino acid side chains at position 4 (p4Y) and position 6 (p6F) project out of the peptide binding cleft, likely acting as the main TCR contacts while amino acids at position 5 (p5N) and 9 (p9M) act as the main anchoring and at position 3 (p3V) as the secondary anchoring position to H-2D^b.[20]
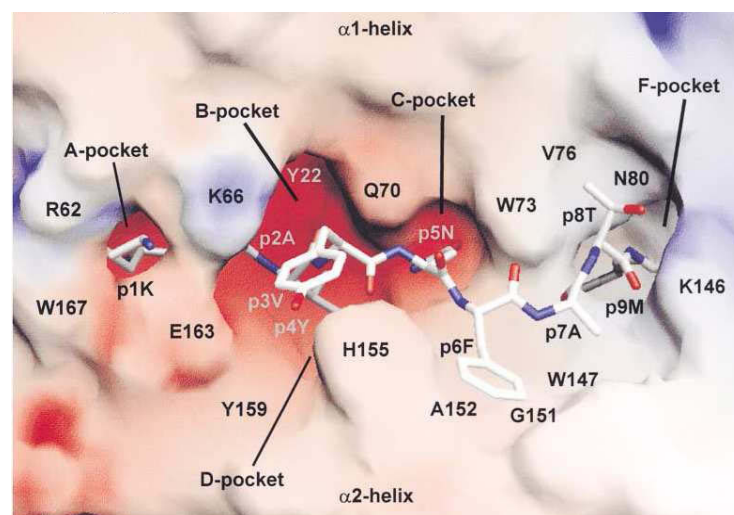
Figure 5 The LCMV derived immunodominant peptide gp33 is bound to the H-2D^b MHC class I molecule. The peptide is represented as a stick model while the MHC class I molecule is represented by its surface (viewed from above). Negatively charged regions of the surface are colored in red and positively charged regions in blue. Peptide residue labels begin with p. Peptide residues p1K, p7A and particularly p4Y and p6F project up into the solvent likely to be the main TCR contacts.[20]

Single amino acid substitutions at position 4 of the immunodominant epitope gp33 affect TCR contacts when presented by H-2D^b and result in very different outcomes of T-cell activity. The natural occurring variant gp33-Y4F evades recognition by the TCR resulting in viral escape. The variant gp33-Y4F and the synthetic variant gp33-Y4S are classified as antagonists. Interestingly, the substitution with alanine at position 4 resulted in a weak agonist. (Table 1)[78]

Table 1 Modifications at position 4 are marked with bold and underlined letters. Position 4 is colored green in the structural presentation, where the crystallographic structure of gp33 was published; gp33-Y4A, gp33-Y4S and gp33-Y4F were modelled using the program PyMol.[6, 19, 20]

Name	Amino acid sequence	Recognition	Structure
gp33	KAVYNFATM	full agonist	
gp33-Y4A	KAV**A**NFATM	weak agonist	
gp33-Y4S	KAV**S**NFATM	antagonist	
gp33-Y4F	KAV**F**NFATM	antagonist	

1.2 *Streptococcus pneumoniae*

Independently isolated in 1881 by George Sternberg and Lois Pasteur *Streptococcus pneumoniae* is today known as one of the major human pathogens. The gram positive bacteria causes pneumonia and other types of pneumococcal infections including otitis media, bacteremia, meningitis and sepsis leaving more than 1.6 million dead per year. Especially children under the age of 5 and elderly people can suffer from serious forms of infections. (Figure 6)[48, 49]

Figure 6 Distribution of deadly ending pneumonia infections of children under the age of 5 years. One red dot equals 1000, leading to estimated 700,000 – 1 million deaths of children per year in total.[54]

Streptococcus pneumoniae are found in the nasopharynx of healthy people where they are attached via bacterial surface adhesins. Infections occur when bacteria are transported into the Eustachian tube or paranasal sinus causing ear infections (otitis media, sinusitis). Pathogen bacteria present in the bloodstream (bacteremia) can cause sepsis or can further be transported to the meninges, causing meningitis. The most common disease, pneumonia, is caused when *Streptococcus pneumoniae* are inhaled into the lungs. Beside of adherence to the host cell, the extracellular pathogen needs to escape clearance and phagocytosis by the host cell. Pneumococci belong to the group of encapsulated bacteria. Repeating oligosaccharides that make up the capsule are transported to the cell surface, where they bind tightly with the cell-wall polysaccharides. The capsule prevents fixation of *Streptococcus pneumoniae* by the complement cascade of the innate immune system and therefore limits the host-cell phagocytosis ability.

Several virulence factors are expressed by *Streptococcus pneumoniae* such as the polysaccharide capsule, hyaluronate lyase (Hyl), pneumolysin (Ply), neuraminidases (NanA and NanB), choline binding protein A (CbpA),

pneumococcal surface antigen A (PsaA), pneumococcal surface protein A (PspA) and the major autolysin (LytA).[2] Virulence factors enable the pathogen to adhere onto host cells, to evade and inhibit the immune response and to obtain nutrition from the host.

1.2.1 Autolysins

Bacteria possessing peptidoglycans have enzymes which can degrade the peptidoglycan backbone. Secreted into the cell wall, autolysins play a role in cell wall growth, cell wall turnover and cell separation.[2] Excessive amounts of autolysins will degrade the peptidoglycan matrix until cell lysis, leading to cell death as a consequence.[3] Autolysins are also involved in programmed cell death (PCD) mechanisms and pneumococcal fracticide.[2, 51] In *Streptococcus pneumoniae* two autolysins, LytA and LytC, are found.[1, 51]

The amidase LytA cleaves the peptidoglycan strand between L-alanine and N-acetylmuramic acid, whereas LytC, a lysozyme, cleaves between N-acetylmuramic acid and N-acetylglucosamin. López *et al.*[50] showed that the activity of the pneumococcal autolysins is depended on the presence of choline. Choline is part of teichoic (TA) and lipoteichoic acids (LTA) which are present in *Streptococcus pneumoniae* cell walls.

The peptidoglycan strand in gram positive bacteria consists of two repeating amino sugars, N-acetylglucosamin (GlcNAc) and N-acetylmuramic acid (MurNAc) both linked together via a β-(1,4) glycosidic bond. N-acetylmuramic acid has a pentapeptide bound, which is made of L-alanine, D-iso-glutamine, L-lysine and one or two D-alanines. To build the peptidoglycan matrix L-lysine from one GlcNAc-MurNAc-pentapeptide unit is linked to D-alanine from another unit via 5 glycine residues. (Figure 7)[51, 53]
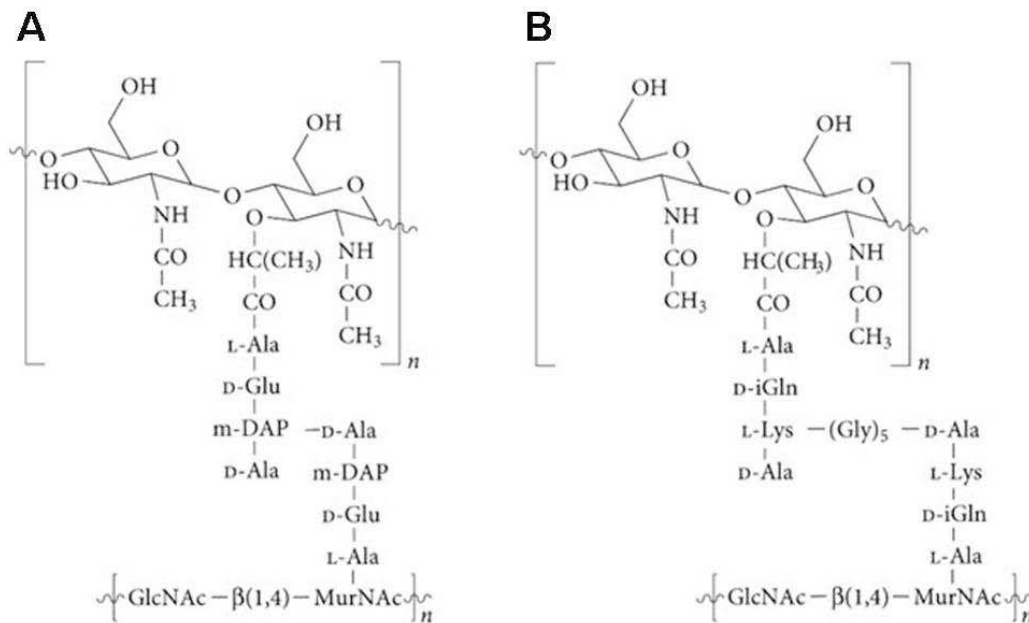
Figure 7 Peptidoglycan strands consist of repeating units of N-acetylglucosamin (GlcNAc) and N-acetylmuramic acid (MurNAc). **A** Peptides bound to MurNAc contain a meso-diaminopimelic acid (DAP) in gram negative bacteria linked to D-alanine from another peptidoglycan backbone (DAP)-type peptidoglycan. **B** Gram positive bacteria belong to Lys-type peptidoglycan linking peptidoglycan backbones between L-lys and D-ala via a five Gly residue linker.[53]

Two models for teichoic and lipo teichoic acid residues present in *Streptococcus pneumoniae* cell walls have been suggested.[52] LTA is composed of a lipid anchor and a pentameric repeating unit made of Glc, AATGal (2-amido-4-amino-2,4,6-trideoxy-D-galactose), two GalNAc residues and ribitol-phosphate residue and has modifications of either one or two GalNAc residues with phosphocholine. The polymer is formed by linking 2-8 repeating units (Figure 8, Model A). A second proposed model has a repeating unit made of AATGal, ribitol-phosphate and two GalNAc residues modified with phosphocholine. Repeating units (n = 2-8) are linked to a lipid anchor via a Glc residue. (Figure 8, Model B)
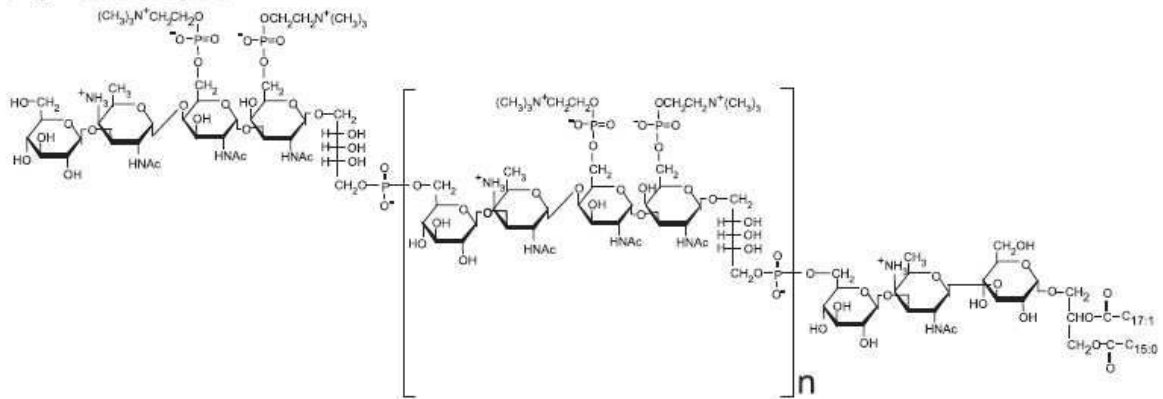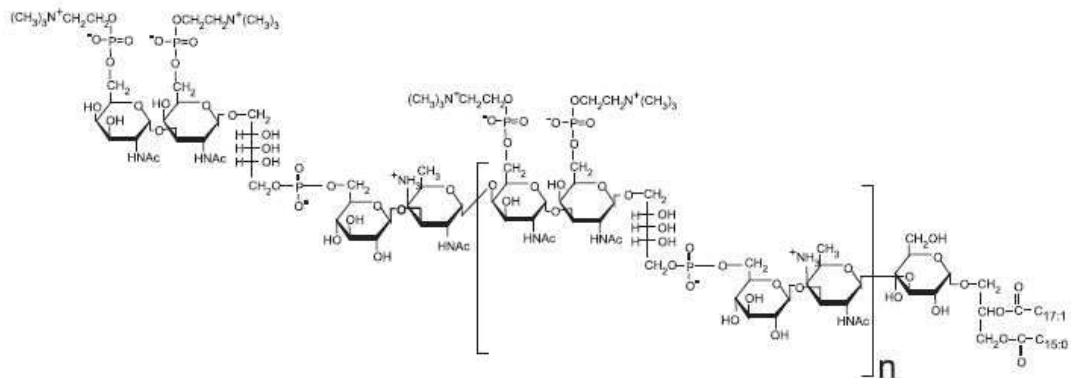
Figure 8 Two models of LTA from *Streptococcus pneumoniae*. **A** shows the model with a repeating unit consisting of Glc, AATGal, two residues of GalNAc and one ribitolphosphate. Both GalNAc residues have phosphocholine bound. Model **B** shows the repeating unit containing two GalNAc residues, ribitolphosphate, Glc and AATGal. Two phosphocholine residues are bound to the GalNAc residues. Both models show the lipid anchor to the right.[52]

Most bacterial autolysins share a two domain fold, consisting of a choline binding domain and a catalytic domain. The choline binding domain is responsible for the attachment of the autolysin to teichoic acids in the peptidoglycan. Upon non-covalently binding, the typical solenoid β hairpin fold is stabilized. It is discussed that choline binding is involved in the activation and dimerization of autolysins.[35, 41, 42, 51] The catalytic domain carries the enzymatic function to hydrolyze peptidoglycan strands.

Since autolysins are able to degrade the essential cell wall, leading to cell lysis and cell death, it is thought that the enzyme must be under strict control. The mechanisms how activity is controlled remain elusive.[2]

1.2.2 *Streptococcus pneumoniae* major autolysin LytA

LytA is a 318 amino acid residues long protein consisting of two modules: a C-terminal choline binding domain (residues 181-318) and a N-terminal catalytic domain (residues 1-180). The structure of the choline binding domain was solved in 2001 (pdb: 1HCX) and presents a left-handed ββ-3-solenoid spiral staircase and is made of six repetitive units, the choline binding regions 1-6 (ChBR). Each conserved ChBR forms two antiparallel β-strands connected via a short internal loop, resulting in a β hairpin structure. ChBR 6 is slightly different compared to the others and is involved in the formation of a dimer rather than choline binding. Between hairpins 1 and 2, hairpins 2 and 3, hairpins 3 and 4 and hairpins 4 and 5, four choline binding sites (ChBS 1-4) are found. The hydrophobic cavity is made of three aromatic amino acids (two tryptophanes, one tyrosine) and one hydrophobic residue (methionine or leucine). Monomeric LytA proteins show an over 90 % decrease in activity, compared to homodimeric formations. Hydrophobic coupling of two hairpins (ChBR6) at the C-terminal end of the choline binding domain leads to the formation of a homodimer. This boomerang shaped structure has an angle of 85° between the two 50 Å long choline binding domains. The catalytic domains are positioned at the end of the arms. Dimerization seems to be important for catalytic activity but the connection between dimerization and activation of the amidase is not known. (Figure 9)[41, 42]
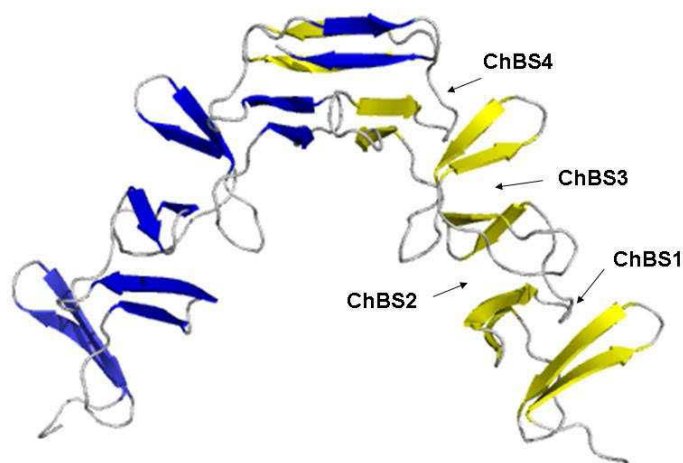
Figure 9 Cartoon representation of the choline binding domains A (yellow) and B (blue) forming a dimer. The locations of the four choline binding sites (ChBS1-4) are indicated with arrows.[41, 42]

The N-terminal located catalytic domain hydrolyzes the link between N-acetylmuramoyl residues and L-alanine residues of lysine type peptidoglycan strands. Substrate binding, the detailed mechanism of catalysis and the activation

of LytA is not known. In comparison to LytC, LytA does not have a signal peptide, but is nonetheless secreted into the cell wall.

Functionally LytA is also involved in penicillin induced lysis. Penicillin and its derivates are the primary antibiotics of choice to treat pneumococcal infections. Penicillin binds and blocks the penicillin-binding proteins (PBPs) which constitute the cell wall synthesis machinery. This mode of action should theoretically be bacteriostatic (i.e. growth inhibiting) and should not cause bacteriocidal (lytic) effects. However, in addition to blocking cell wall synthesis, penicillin somehow causes activation of LytA which leads to cell lysis and cell death. In LytA deficient pneumococcal strains penicillin works in a bacteriostatic manner and clinical penicillin tolerant strains have often a decreased expression of LytA.[69, 70]

Strains of drug-resistant *Streptococcus pneumoniae* have become increasingly common in the United States and in other parts of the world. Up to 35 % of pneumococcal isolates have been reported to be resistant to penicillin. Many penicillin-resistant pneumococci are also resistant to other antimicrobial drugs (e.g. erythromycin, cephalosporins). The development of future vaccines which act differently against pneumococcal diseases is therefore extremely important.

1.3 X-ray Crystallography

Since the 19th century, crystallography played an important role in the determination of molecular structures. Nowadays there are more than 60,000 different structures solved by crystallography (RCSB Protein Data Bank, http://www.rcsb.org/pdb/home/home.do; March 2011). The introduction of X-ray diffraction analysis in the late 1930s and the recombinant DNA technology in the 1980s were two major breakthroughs. Therefore macromolecular structures at atomic level of proteins, nucleic acids, viruses and other large molecular complexes could be determined. Still, the determination of new protein structures is an important field in crystallography but also the raising field of genetic engineering of proteins requires the knowledge of three-dimensional structures. Since the beginning of crystallography the key element are the crystals. New methodologies such as synchrotron radiation and cryocrystallography allow the use of small size crystals, but having good quality crystals is still the main objective in crystallography.

The physico-chemical knowledge of how self-organization of macromolecules is achieved is limited. Therefore at the present time no predictions on crystal formation can be made, leaving crystal growth within a black box. Systematically searching and optimizing of parameters which influence crystal formation is the effort crystallographers are doing.[23, 25, 26]

1.3.1 Protein crystals

At specific chemical conditions many macromolecular molecules can form solid crystals. Crystals have a repetitive three-dimensional array of molecules, where the smallest repetitive unit is called unit cell. If the content of the unit cell is known, the content of the whole crystal is known. To specify the exact location of atoms within the unit cell, cartesian coordinates for each atom are given in x, y, and z and are measured in Ångström. Seven crystal lattices are possible (cubic, tetragonal, hexagonal, orthorhombic, rhombohedral, monoclinic and triclinic). Four lattice centerings are known, primitive (P), body (I) face (F) and base-centered (A, B or C). In combination with the seven crystal systems, 14 Bravais lattices are known. (Figure 10)[23]
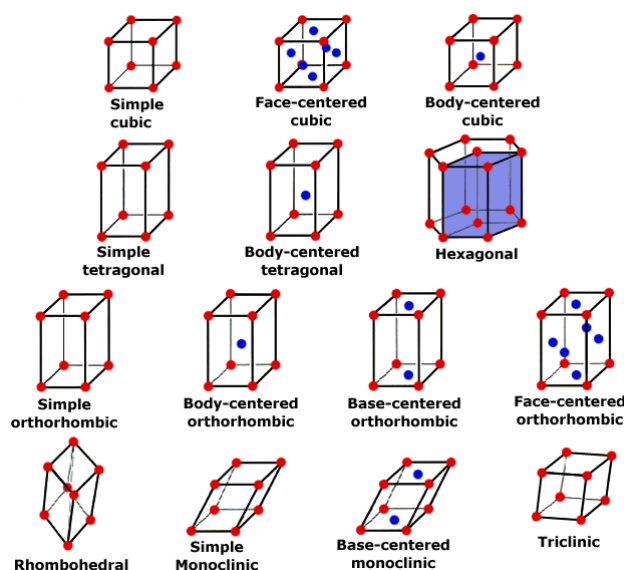
Figure 10 The 14 Bravais lattices with red colored lattice points and blue colored P, I, F or A/B/C centerings.[59]

To further describe the unit cell, the content of the unit cell has to be specified. Usually the unit cell contains several identical molecules in a symmetrical arrangement (translation, rotation and reflection). The smallest element without symmetry is called asymmetric unit. The combination of the lattice type and the internal symmetry of the unit cell is called space group. In total there are 230 unique space groups describing all possible crystal symmetries, which are listed in the International Tables for Crystallography.[44] For proteins mirror symmetry is prohibited, that is why only 65 space groups are possible for macromolecular crystals.

As said before, the growth of protein crystals remains a not well understood process. Though, basic principles can be followed to obtain protein crystals. A purified protein sample is dissolved in an aqueous buffer containing a precipitant (e.g. polyethylene glycol) at a concentration which is below that to precipitate the protein. Water from the aqueous buffer solution is evaporated which leads to an increase of protein and precipitant concentrations. When the solution is saturated with protein, the first step in protein crystal growth, nucleation can occur. Nucleation is the formation of ordered molecular clusters and is a requirement for the second step, crystal growth. (Figure 11)[23]
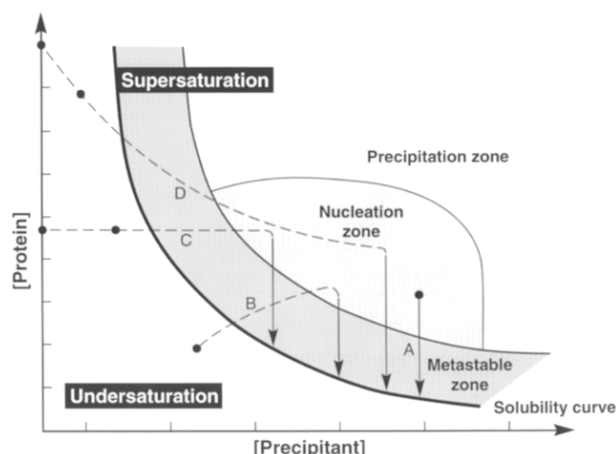
Figure 11 Phase diagram for crystallization using variable parameters of protein and precipitant concentrations. The undersaturated and supersaturated state is shown as well as the nucleation zone. Different techniques and ways to reach the nucleation zone are shown. (A) Batch crystallization (B) Vapor diffusion (C) Dialysis (D) Free-interface diffusion.[45]

Beside protein and precipitant concentrations, several parameters influence crystal growth which are listed in Table 2. Altering these parameters may affect the nucleation probability, crystal growth, crystal size and the overall quality of the crystal. Finding the best combination of all involved parameters is referred to as the "art of crystallography".[26]

Table 2 Factors effecting crystallization[26, 25]

Physical	Chemical	Biochemical
Temperature	pH	Purity of the macromolecule
Surfaces	Precipitant type	Ligands, inhibitors, effectors
Approach to equilibrium	Precipitant concentration	Aggregation state of macromolecule
Gravity	Ionic strength	Post-translational modifications
Pressure	Specific ions	Source of macromolecule
Time	Degree of supersaturation	Proteolysis/hydrolysis
Vibrations	Reductive/oxidative environment	Chemical modifications
Electrostatic/magnetic fields	Concentration of macromolecules	Genetic modifications
Dielectric properties of the medium	Metal ions	Symmetry of macromolecule
Viscosity	Crosslinkers/polyions	Stability of macromolecule
Rate of equilibration	Detergents	Isoelectric point
Homo/heterogeneous nucleants	Non-macromolecular impurities	History of the sample

1.3.2 Collecting X-ray diffraction data

An incident X-ray beam is shot onto a single crystal where it can diffract into several discrete beams when colliding with electrons. Having constructive interference of reflected X-ray beams, their position and intensity can be recorded with a detector. Constructive interference of scattered waves is achieved when successive parallel crystallographic planes are present. This condition is described in Bragg´s Law. (Figure 13)[46]

$$2d_{hkl} \sin \theta = n\lambda$$

Fictitious crystallographic planes which can be drawn through lattice points can be thought as the source of diffraction. They are designated by the lattice indices (Miller indices), *hkl.* The index *h* gives the number of planes per unit cell in the x direction, *k* and *l* specify the number of planes in the y and z direction. All parallel lattice planes are called a set of planes and are written as *(hkl)*. (Figure 12) Bragg showed that a set of parallel planes *(hkl)* with interplanar spacing d_{hkl} produces a constructive interference for the diffracted beam when the incident and reflecting angle *θ* is the same and the additional distance travelled by waves going deeper into the crystal (2*d* sin *θ*) must be an integral multiple of λ.
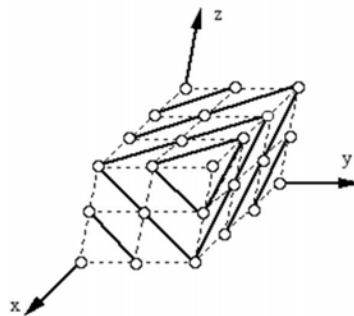
Figure 12 A set of planes with Miller index (111) is shown within a unit cell. The (111) planes occur one time in each direction of the unit cell.[60]
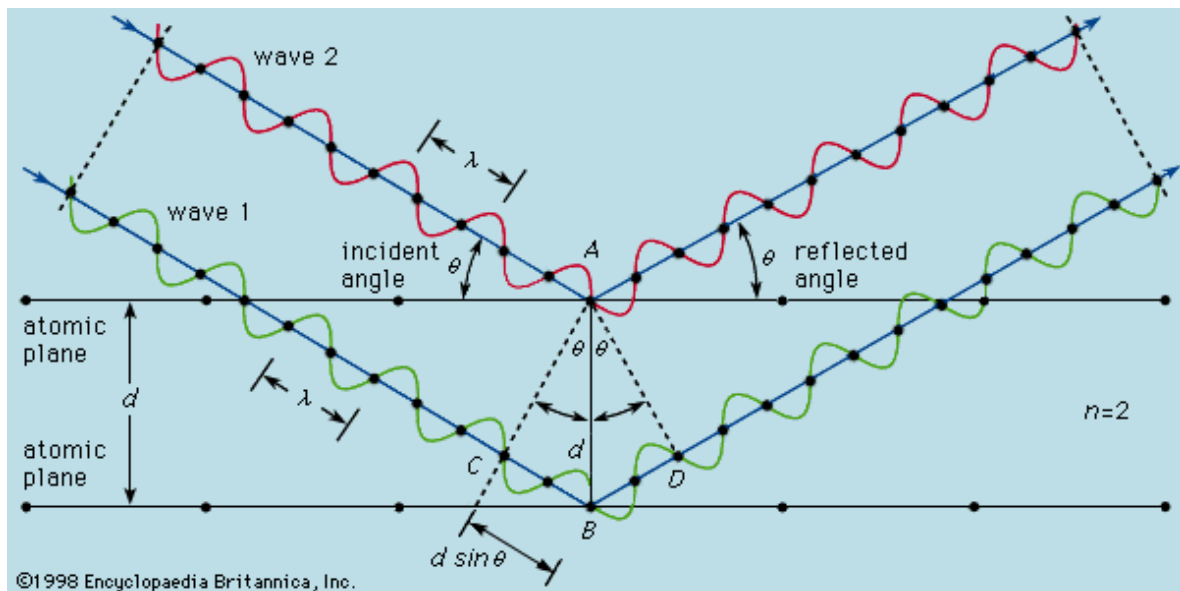
Figure 13 Conditions which produce constructive interference are given in Bragg´s Law and are schematically drawn. Incident X-ray beams (wave 1 and 2) must have the same incident and reflected angle *θ*. Constructive interference between wave 1 and wave 2 is achieved when the additional distance travelled by wave 1 (2*d* sin *θ*) is an integral multiple of λ. The distance between the two atomic planes is given in *d*, which is the interplanar spacing d_{hkl}.[61]

Each reflection on the diffraction pattern comes from a set of diffraction planes *(hkl)* within the crystal. Each reflection point can be assigned a position *hkl* in the

reciprocal space, which is the three-dimensional space occupied by reflections. The origin (000) lies in the direct path of the X-ray beam. The reciprocal lattice is a Bravais lattice, and the reciprocal of the reciprocal lattice is the original lattice. Different reflections (positions and intensities) are observed as the crystal is rotated during data collection, because positions of diffraction planes perpendicular to the X-ray beam change.[23, 44]

The result of X-ray data collection is a list of intensities corresponding to their position in the three-dimensional reciprocal space, I_{hkl}.

1.3.3 From diffraction data to electron density

The contribution of one individual atom (j) to reflection hkl on the diffraction pattern is given in the atomic structure factor, f_{hkl}.

$$f_{hkl} = f_j e^{2\pi i(hxj+kyj+lzj)}$$

The term f_j is the scattering factor for atom j and describes an atom as a sphere of its electron density. The exponential wave function combines the xj, yj and zj coordinates from atom j in the unit cell and the indices h, k, l of specific reflections in the reciprocal lattice. The sum of all atomic structure factors gives the structure factor for reflection F_{hkl}.

$$F_{hkl} = \sum_{j=1}^{n} f_j e^{2\pi i(hxj+hyj+lzj)}$$

The structure factor describes a reflection hkl as a Fourier sum in which the sum of diffractive contributions from all atoms in the unit cell are combined. The contribution of each atom j to F_{hkl} firstly depends on its element, which defines the amplitude of the contribution (f_j), and secondly the position in the unit cell (xj, yj, zj) which establishes the phase of its contribution.

Since the structure factor F_{hkl} describes a diffracted wave, three parameters must be determined: amplitude, frequency and phase. The amplitude of the structure factor F_{hkl} is given as it is proportional to the square root of measured intensities $(I_{hkl})^{1/2}$ of each reflection. The frequency of the structure factor is equal to $1/d_{hkl}$, making the wavelength the same as the distance between planes responsible for diffraction. The loss of information regarding the phase must be compensated by

means of another measurement in order to calculate the electron density using following equation, where V is the volume of the unit cell.[23]

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

1.3.4 Obtaining phases

Several methods are used to obtain phases (isomorphous replacement, anomalous scattering and molecular replacement).

Isomorphous replacement is the addition of heavy atoms into the unit cells by various means. With the knowledge that every atom in the unit cell contributes to every reflection one can modify atoms at identical sites in all unit cells and measure the differences in the diffraction pattern. Elements of higher atomic numbers (heavy atoms) contribute stronger to the diffractions. Measured differences in the diffraction pattern can be used to obtain estimates of phases. When heavy atoms like mercury or gold are added, crystal packing and the conformation of the protein must not be disturbed.

Anomalous scattering makes also use of heavy atom derivatives. Heavy atoms absorb X-rays at a specific wavelength which leads to an inequality of symmetry related reflections.

Molecular replacement is the most common way to overcome the phase problem but is only possible if very similar three-dimensional structures are available. When placing the model of a known protein (phasing model) into the unit cell of the new protein an initial phase can be calculated.[23]

1.3.5 Molecular model

Having the electron density map, the main polypeptide chain and side chains of amino acids can be modelled into it. The correct three-dimensional positions of the side chains can be seen, ligands and water molecules can be determined. Several parameters are defined to describe the quality of a model.

Resolution, given in Ångström, is a measure of the level of detail present in the diffraction pattern and in the resulting electron density map. Having a perfectly ordered crystal, all X-rays will diffract in the same way showing fine details of the

crystal. If molecules are not exactly the same in the whole crystal, some fine structural information is therefore lost. Ramachandran plots plot the dihedral angel φ versus ψ of all amino acid residues in the polypeptide chain. Ramachandran plots show areas of preferred angles of secondary structure elements (helices, sheets). Not allowed φ/ψ combinations are therefore easily seen.[47] The R-factor tells the difference between the experimental diffraction data and the crystallographic model. It is used as a quality index during refinement process. If it increases, the correlation between the model and the experimental data is worsened. A perfect model would yield in an R-factor of 0.[23] A less biased parameter is the R_{free} value. Before refinement 5-10 % of the data are removed and refinement is done with the 90 % left. The R_{free} value is then calculated to see how well the model fit the electron density.

2 Aims of the thesis

During previous experiments very different outcomes of T-cell recognition were seen, although only very small changes at the presented peptides were made. To further investigate these altered peptides affinity studies and thermodynamical studies are going to be performed. Together with already available structural data explanations for those huge differences in T-cell recognition should be made. Therefore soluble amounts of refolded pMHCs and TCR p14 are going to be produced to calculate binding affinities and thermodynamical signatures of those interactions. All those results should further give a more detailed insight of how T-cell recognition between agonist and antagonist is done on molecular level.

Streptococcus pneumoniae major autolysin LytA is an immunological interesting protein, since it destroys the peptidoglycan network of bacteria. The catalytic domain and the full length protein are going to be crystallized. Having the crystalstructure of this protein, detailed mechanisms of ligand binding and cutting the peptidoglycan strands should become clearer. The necessity of forming homodimers and how the activation of the protein is done on a molecular level might be answered too.

3 Materials and Methods

3.1 Production of soluble β$_2$-microglobulin and of the heavy chain of H-2 class I histocompatibility complex, H-2D^b

3.1.1 Protein expression system

T7 Express *lysY* Competent *Escherichia coli* (High Efficiency) (New England Biolabs) and BL21(DE3)pLysS Chemically Competent Escherichia coli (Promega) were used as expression hosts to produce β$_2$m and H-2D^b.

The strains main characteristics are their T7 RNA polymerase and their T7 lysozyme (*lysY*). T7 polymerase gene is inserted into the *lac* operon and is expressed under the control of the *lac* promoter. T7 lysozyme (*lysY*) expresses the variant K128Y which lacks amidase activity but is still able to inhibit T7 RNA polymerase. Induction of the *lac* operon is done by the lactose analogue IPTG (Isopropyl-1-thio-β-D-galactopyranoside) due to binding and inactivating the *lac* repressor. Both strains are resistant to chloramphenicol.[9]

3.1.2 Plasmid vectors

The expression hosts were transformed with pET-8c (mβ$_2$m) and pET-3a (H-2D^b) vectors, respectively (Novagen). pET system vectors contain a T7 promoter, a cloning site (*Nde* I or *Nco* I and *Bam*H I) where the coding sequences of mβ$_2$m, H-2D^b and H-2D^b-bsp are cloned into, a T7 terminator and the β-lactamase gene *bla* leading to ampicillin resistance of the host *Escherichia coli*.[4] Protein attributes of mouse β$_2$-microglobulin and mouse H-2 class I histocompatibility antigen are shown in Table 3.

Table 3 Protein attributes of mouse β_2-microglobulin and mouse H-2 class I histocompatibility antigen. Red marked areas belong to full length sequences but were not present in the used cloning vectors.

Name	beta-2-microglobulin, mβ_2m
Organism	*Mus musculus* (Mouse)
Uniprot ID	P01887
Mass (Da)	13823
Extinction coefficient (ε)	19940
Sequence length	119

Aminoacid sequence

```
          10         20         30         40         50         60
MARSVTLVFL VLVSLTGLYA IQKTPQIQVY SRHPPENGKP NILNCYVTQF HPPHIEIQML
          70         80         90        100        110
KNGKKIPKVE MSDMSFSKDW SFYILAHTEF TPTETDTYAC RVKHDSMAEP KTVYWDRDM
```

Name	H-2 class I histocompatibility antigen, D-B alpha chain; H-2D(B)
Organism	*Mus musculus* (Mouse)
Uniprot ID	P01899
Mass (Da)	40836 (32947)
Extinction coefficient (ε)	84340
Sequence length	362 (276)

Aminoacid sequence

```
          10         20         30         40         50         60
MGAMAPRTLL LLLAAALAPT QTRAGPHSMR YFETAVSRPG LEEPRYISVG YVDNKEFVRF
          70         80         90        100        110        120
DSDAENPRYE PRAPWMEQEG PEYWERETQK AKGQEQWFRV SLRNLLGYYN QSAGGSHTLQ
         130        140        150        160        170        180
QMSGCDLGSD WRLLRGYLQF AYEGRDYIAL NEDLKTWTAA DMAAQITRRK WEQSGAAEHY
         190        200        210        220        230        240
KAYLEGECVE WLHRYLKNGN ATLLRTDSPK AHVTHHPRSK GEVTLRCWAL GFYPADITLT
         250        260        270        280        290        300
WQLNGEELTQ DMELVETRPA GDGTFQKWAS VVVPLGKEQN YTCRVYHEGL PEPLTLRWEP
         310        320        330        340        350        360
PPSTDSYMVI VAVLGVLGAM AIIGAVVAFV MKRRRNTGGK GGDYALAPGS QSSEMSLRDC

KA
```

3.1.3 Protein expression and isolation

Induction of 1 litre expression cultures having an OD_{600} amount between 0.6 – 0.7 (Novaspec III, GE Healthcare) was done with IPTG to a final concentration of 1 mM. Bacterial growth at 37° C in lysogeny broth (LB) medium containing 100 mg mL^{-1} ampicillin, was harvested after 4 hours and centrifuged at 8000 g for 20 minutes. The cell pellet was resuspended in 40 mL STET Lysis buffer (50 mM Tris HCl p*H* = 8.0, 25 % sucrose (w/v), 1 mM EDTA p*H* = 8.0), PMSF (final concentration 1 mM), DTT (final concentration 2 mM) and frozen in liquid nitrogen to rupture the cell walls for at least two times. Lysis of the cells was done with lysozyme (final concentration 20 µg mL^{-1}) having an incubation time of 30 minutes. DNA degradation was done using $MgCl_2$ (final concentration 10 mM), $MnCl_2$ (final concentration 1 mM) and Dnase (final concentration 10 µg mL^{-1}). To remove any left big particles the solution was sonicated 3 x 1 minute (Soniprep 150, MSE). After precipitating the inclusion bodies (8000 g, 20 min) (Multifuge 1 S-R, Heraeus), they were resuspended in 20 mL DET buffer (0.2 M sodiumchloride, 1 % deoxycholic salt (w/v), 1 % Nonidet P40 (v/v), 20 mM Tris HCl p*H* = 7.5, 2 mM

EDTA pH = 8.0), DTT (final concentration 2 mM) and PMSF (final concentration 1 mM). Inclusion bodies were removed (8000 g, 20 min) leaving fusion proteins in the cell pellet. Proteins were washed repetitively with 24 mL Triton X-100 buffer (0.5 % Triton X-100, 50 mM Tris HCl pH = 8.0, 100 mM sodiumchloride, 1 mM EDTA pH = 8.0) and DTT (final concentration 2 mM) until a tight pellet was obtained. Triton X-100 solubilises the membrane fraction completely, leaving the fusion protein in the pellet. The isolated proteins were resuspended in 5-10 mL of 8 M Urea, 50 mM MES, pH = 6.5, 0.1 mM EDTA and 0.1 mM DTT.[6, 7, 8, 10] Determination of the protein concentration was done spectrophotometrically (NanoDrop 2000c, Thermo Scientific). To ensure the correct expression and purity of the desired protein, molecular weights were examined by comparing with a molecular weight marker (PageRuler Plus Prestained Protein Ladder, Fermentas) on a SDS - PAGE (Sodium Dodecyl Sulfate – Poly Acrylamid Gel Electrophoreses) under denaturing conditions.[9]

Background - SDS-PAGE

Gel electrophoresis is used to separate proteins according to their electrophoretic mobility, e.g. molecular weight, when an electrical field is applied to a gel matrix. The gel consists of polymerized acrylamid and forms a solid and porous matrix. The dissociating agent sodium dodecyl sulphate denatures secondary structures of the protein sample and covers the polypeptide backbone when heated to 100 ℃. As a result the protein shows a high density of negative charges on its surface at a constant weight ratio of 1.4 g SDS per gram polypeptide. With uniform negative charges, the intrinsic charges of the proteins have just a minor impact, separating proteins exclusively in respect to their mass. Negative charged particles migrate towards the positive charged anode exclusively depending on their mass.[12, 13] The stacking part of the gel allows SDS coated proteins to concentrate to a thin starting zone before reaching the resolving part of the gel.[12]

After preparing gels (Table 4) the gel, electrophoresis module and power supply (Mini-Protean 3 cell, Bio-Rad) were assembled.[14] A 10 µL protein solution (resuspended in 8 M urea) was mixed with 20 µL of DTT and heated for about 10 minutes at 80-90° C. Usually 2-5 µL of these sample s were loaded onto the gel. DTT reduces disulfide linkages and degenerates tertiary protein structures. An

electrical field of 190 V was usually applied to the gel for 60 minutes, allowing the proteins to migrate towards the anode and separate according to their mass. Staining of the gels was performed using Coomassie Blue staining and destaining solutions or by Page Blue Protein Staining Solution (Fermentas). A molecular weight marker (PageRuler Plus Prestained Protein Ladder, Fermentas) was used as a reference.

Table 4 Chemical composition of SDS-gels

Resolving buffer		Stacking buffer
7.1	dH$_2$O [mL]	7.3
5.0	1.5 M Tris pH 8.8 [mL]	-
-	1.0 M Tris pH 6.8 [mL]	1.25
7.5	40 % PAA [mL]	1.25
0.2	10 % SDS [mL]	0.1
0.2	10 % APS [mL]	0.1
16	TEMED [µL]	8

3.1.4 Refolding of soluble peptide MHC complexes

All three units (H-2D^b, mβ_2m, peptide) are necessary for a successful refolding. Refolding was performed by rapid dilution at 4° C using a molar ratio for H-2D^b:mβ_2m:peptide at 1:2:10 (final concentrations 30 mg L^{-1}, 24 mg L^{-1} and 10 mg L^{-1} respectively).[8, 9] A 1 litre refolding buffer comprised 0.4 M L-argininechloride, 20 mM Tris–HCl pH 7.5, 1 mM EDTA, 5 mM GSH (oxidized glutathione), 0.5 mM GSSG (reduced glutathione) 100 mM PMSF and 0.01 % (v/v) sodiumazide to prevent bacterial growth. After addition of the peptide, the stable and small subunit mβ_2m was added which can refold independently. It functions also as a preliminary template for the H-2D^b heavy chain subunit and increases therefore the yield of refolded pMHC complexes.[5] Furthermore, the refolding solution was pulsed with H-2D^b to increase the yield of refolded pMHC.

After stirring for 48 hours at 4° C, concentration of pMHC complexes to a volume of approximately 50 mL was done with a tangential flow device (Ultrasette TM, Pall) with a 10 kDa MWCO. Further concentration to a volume of 5 mL was performed using Centriprep centrifugal filter units with a 10 kDa MWCO (Millipore). Refolded pMHC complexes were purified by Fast Performance Liquid Chromatography (FPLC) using a Superdex 75prep grade 16/60 gel-filtration column with a column volume of 125 mL (ÄKTApurifier, GE Healthcare) equilibrated with 20 mM Tris-HCl, pH 7.5.

> **Background - Size Exclusion Chromatography (SEC)**
>
> Superdex 75 gel-filtration columns have a spherical composite of porous agarose particles to which dextran has been covalently bonded. The average particle size of agarose particles is 13 µm. The separation of proteins is due to differences in their sizes. While a continuously flow of buffer flows through the column, the molecules diffuse in and out of the pores of the stationary phase. Low molecular weight molecules diffuse further into the matrix and stay longer on the column compared to large molecules. Larger molecules are eluted prior to smaller molecules.[11]

5 mL of concentrated pMHC solution was applied onto the column (Superdex 75 10/300 GL) and isocratically eluted with a flow rate of 0.8 mL min^{-1}. UV absorbance at 280 nm showed purified pMHC complexes. At 280 nm proteins show an absorbance peak due to light absorption by aromatic amino acids (tyrosine, tryptophan and phenylalanine).

Fractions consisting of pMHC complexes were pooled together and were concentrated to a final concentration higher than 4 mg mL^{-1} using Microsep centrifugal devices having a 10 kDa MWCO (Pall Corporation). For affinity studies and crystallization assays pMHC complexes need to have a concentration of approximately 90 µM or 4 mg mL^{-1} at least.

3.2 Production of soluble TCR p14 monomers (p14α, p14β)

3.2.1 Protein expression system

Rosetta(DE3)pLysS (Novagen) were used for expression of α and β subunits with a chromosomal copy of the T7 RNA polymerase gene (DE3). T7 RNA polymerase is under the control of the *lacUV5* promotor. Addition of IPTG will induce expression of the target protein. T7 lysozyme gene suppresses basal expression of T7 RNA polymerase prior to induction. tRNA genes which are rarely used in *Escherichia coli* (pLysSRARE) are also present on an additional plasmid. The cell line carries a chloramphenicol resistance gene (Cam^R).

3.2.2 Plasmid vector

p14α and p14β were each cloned into the bacterial expression vector pGMT7 and transformed into Rosetta(DE3)pLysS Competent Cells.[21] Protein attributes and amino acid sequences of TCR p14 subunits α and β are listed in Table 5.

Table 5 Protein attributes and amino acid sequences of TCR p14 subunits α and β.

Name	p14-alpha
Mass (Da)	23181
Extinction coefficient (ε)	26930 [M-1 cm-1]
Sequence length	206
Aminoacid sequence	10 20 30 40 50 60
	MQQKEKHDQQ QVRQSPQSLT VWEGGTTVLT CSYEDSTFNY FPWYQQFPGE GPALLISILS
	70 80 90 100 110 120
	VSDKKEDGRF TTFFNKREKK LSLHIIDSQP GDSATYFCAA LYGNEKITFG AGTKLTIKPN
	130 140 150 160 170 180
	IQNPEPAVYQ LKDPRSQDST LCLFTDFDSQ INVPKTMESG TFITDKCVLD MKAMDSKSNG
	190 200
	AIAWSNQTSF TCQDIFKETN ATYPSS

Name	p14-beta
Mass (Da)	26641
Extinction coefficient (ε)	49390
Sequence length	238
Aminoacid sequence	10 20 30 40 50 60
	MAVTQSPRSK VAVTGGKVTL SCHQTNNHDY MYWYRQDTGH GLRLIHYSYV ADSTEKGDIP
	70 80 90 100 110 120
	DGYKASRPSQ ENFSLILELA SLSQTAVYFC ASSDAGGRNT LYFGAGTRLS VLEDLRNVTP
	130 140 150 160 170 180
	PKVSLFEPSK AEIANKQKAT LVCLARGFFP DHVELSWWVN GKEVHSGVCT DPQAYKESNY
	190 200 210 220 230
	SYSLSSRLRV SATFWHNPRN HFRCQVQFHG LSEEDKWPEG SPKPVTQNIS AEAWGRAD

3.2.3 Protein expression and isolation

Successfully transformed *Escherichia coli* cells, seen as single cell colonies on ampicillin (100 mg mL^{-1}) and chloramphenicol (34 mg mL^{-1}) containing plates (LA, p*H* = 7.2), were then used to inoculate 20 ml of TYP media (1 litre consists of: 16 g tryptone, 16 g yeast extract, 5 g sodiumchloride and 2.5 g di-potassium hydrogenphosphate) containing ampicillin (50 µg) and chloramphenicol (34 µg) overnight. Then 1 litre of TYP media was inoculated with 20 mL from the overnight culture. At OD$_{600}$ = 0.5 (measured spectrophotometrically with Novaspec III, GE Healthcare) expression of insoluble inclusion bodies was induced with IPTG (final concentration 0.2 mM) and incubated for 3 to 4 hours. After sedimentation (2500 g, 20 min) the supernatant was removed and the pellet was dissolved in 10 mL lysis buffer (10 mM magnesiumchloride, 2 mM DTT).

Lysozyme (final concentration: 10 µg mL^{-1}) and Dnase (final concentration: 10 µg mL^{-1}) were added and incubated for 30 minutes on ice and 30 minutes at 37° C. Freezing at -20° C for 2 hours and thawing a t 37° C for 20 minutes in a

water bath was carried out two times as well as one time freezing at -20 ℃ overnight to fully rupture the cell membranes. Triton wash buffer (1 % Triton X-100, 50 mM Tris pH = 8.1, 100 mM sodiumchloride, 10 mM buffered EDTA and 2 mM DTT) was used to separate p14α or p14β inclusion bodies from the solution using at least two washing cycles. Each cycle consists of the addition of 24 mL Triton wash buffer with an incubation time of 30 minutes, followed by centrifugation (2500 g, 20 min). The last washing step was performed with 24 mL resuspension buffer (50 mM Tris pH = 8.1, 100 mM sodiumchloride, 10 mM buffered EDTA and 2 mM DTT). The obtained pellet (2500 g, 20 min) was dissolved and denatured in 5 mL Urea buffer (8 M Urea, 50 mM MES pH = 6.5, 0.1 mM EDTA and 0.1 mM DTT). The protein concentration was spectrophotometrically determined at a wavelength of 280 nm (NanoDrop 2000c, Thermo Scientific). The purity was controlled using SDS-PAGE.

3.1.4 Refolding of the soluble TCR p14

Refolding of the TCR p14 was performed by rapid dilution at 4° C while stirring, using a 1:1 molar ratio of α and β subunits (final concentration 30 mg 1000 mL^{-1} for each subunit) for 48 hours. The refolding buffer consisted of 100 mM Tris pH = 8.1, 400 mM L-argininechloride and 2 mM buffered EDTA, 0.77 g reduced glutathione and 0.155 g oxidised glutathione. The refolding was concentrated to 50 mL using a tangential flow device (Ultrasette, Pall) with 10 kDa MWCO. The refolding was extensively dialyzed (5 x 8 hours against 5 L of 10 mM Tris pH = 8.3) with standard regenerated cellulose dialysis tubings with 6-8 kDa MWCO (Spectra/Por 1, Spectrum Laboratories).

The first purification step was performed using ion exchange chromatography on a 5 mL Q FF column (HiTrap Q FF, GE Healthcare) which is a strong anion exchange column. The second purification step was performed using size exclusion chromatography.

Background - Ion Exchange Chromatography (IEC)

Ion exchange chromatography is a form of adsorption chromatography which separates proteins according to their relative net charge. Positively charged functional groups of the stationary phase can interact with negatively charged molecules present in the mobile phase via ionic interactions (anion exchange chromatography). First, the stationary phase is equilibrated with exchangeable counterions. During sample application negatively charged target molecules interact with the positively charged functional groups depending on their affinities and leading to equilibrium between counterions and target molecules. Target molecules are eluted from the ion exchanger due to changes in buffer conditions. Mostly, the ionic strength of the buffer is increased to do so. A regeneration step removes all bound molecules from the stationary phase.

The p14 refolding solution was loaded onto the HiTrap Q FF column, equilibrated with 10 mM Tris pH = 8.3 and was eluted with an increasing salt gradient (10 mM Tris pH = 8.3, 1 M sodiumchloride) from 0-100 % in 30 minutes. The second purification was performed by size exclusion chromatography on a 23 mL Superdex 75 10/300 GL column, equilibrated with 10 mM HEPES, 0.15 M NaCl and 0.005 % (v/v) Tween20. The TCR p14 was further concentrated using Centriprep centrifugal filter units with a 10 kDa MWCO (Millipore) to a final concentration of approximately 30 µg mL^{-1} which was spectrophotometrically measured. SDS-PAGE under native and denaturing conditions were used to verify purity and molecular weights of soluble TCR p14 after IEC and SEC.

3.3 Surface Plasmon Resonance binding analysis

**Background – Surface Plasmon Resonance (SPR)**

The phenomenon of surface plasmon resonance can be used to study affinities and kinetics of biomolecular reactions. Using this technique the association and dissociation of biomolecular complexes can be monitored and binding constants can be calculated.

On a prepared surface of a sensor chip, covalently bound ligands are prepared. A solution with the analyte is applied to the ligand where binding can occur. Due to the change in mass concentration after binding on the surface, a response in resonance units (RU) is detected, linked to mass, where 1000 RU equals 1 ng mm^{-2}. A sensorgram shows the continuous response of resonance units over time.

Surface plasmons, also called surface electromagnetic waves, can be stimulated through polarized photons (e.g. Biacore2000) or electrons. Total internal reflection must be achieved, which is done using a prism for the incident beam. At the interface between a high refractive index media (glass of the sensor surface) and a low refractive media (buffer) surface plasmons are excited when striking an electrically conducted gold layer. The plasmons are prolongating along the gold surface until they reach free electron clouds (from bound ligand with or without analyte) where they are absorbed. This absorbance of energy leads to the emission of light and to a reduction of intensity of the reflected light beam. Since many light beams with different angles are refracted, the knowledge of the one light beam with the minimum intensity is crucial to measure the differences in the angles of the refracted light beams at status one (bound ligand without ligand) and status two (bound ligand with analyte). The monitored change in angle of the reflected light is caused by binding or dissociation of molecules bound on the sensor surface and is proportional to the mass bound.
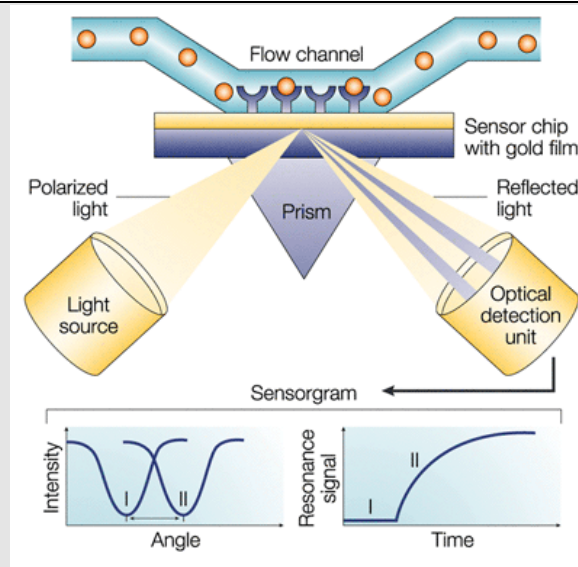
Sensorgrams show the resonance signal in RU over time. Typical sensorgrams from kinetic studies show TCR binding, association of TCR/pMHC complexes, steady state equilibrium, dissociation of TCR/pMHC complexes and regeneration of the sensor surface.[15, 16, 17, 18]
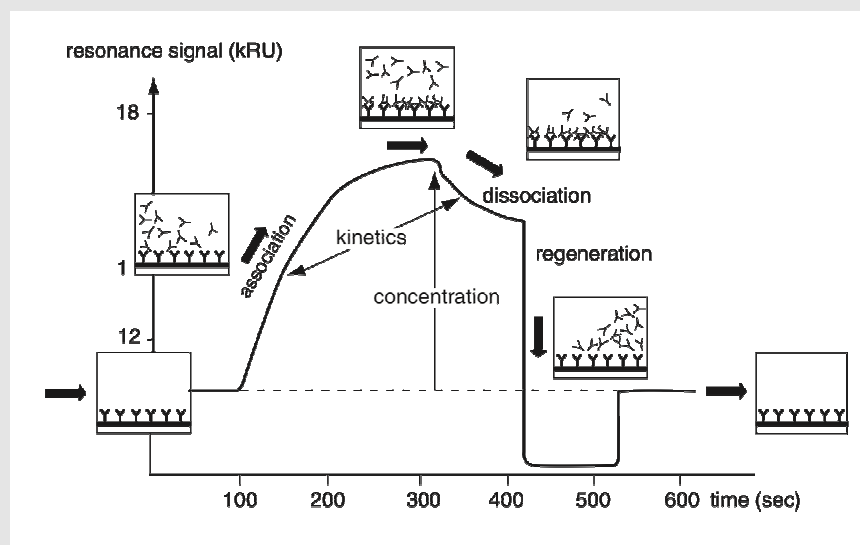
3.3.1 Immobilization and activation of CM5 sensor chips

All measurements were performed on a Biacore 2000 (GE Healthcare, USA) using the running buffer HBS (10 mM HEPES, 150 mM NaCl, 0.005 % (v/v) Tween20).

CM5 sensor chips (GE Healthcare, USA) with immobilized anti-C_β antibodies were used for kinetic studies and affinity determinations. Anti-C_β antibodies couple non-covalently the constant part of the beta chain of the TCR p14. pMHC complexes are non-covalently bound to the TCRs variable α and β chains. CM5 sensor chips

are glass slides coated with a thin gold film and have a carboxymethylated dextran matrix covalently bound. Carboxyl groups allow covalent attachments of the ligand (anti-C_β antibody) through free amino groups. The surface was activated with 0.05 M N-hydroxysuccinimide (NHS) and 0.1 M 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC), both provided by GE Healthcare. Anti-C_β antibody (clone H57-597, eBioscience)[22] was dissolved in 10 mM sodium acetate pH = 4.5 to a final concentration of 20 µg mL^{-1}. H57-597 was injected to the activated chip surface until approximately 8000 RU were obtained. Through injection of 0.1 M ethanolamine-HCl pH = 8.5 the surface was deactivated. A deactivated control surface (usually flow cell 1) was also prepared using the same protocol, except that no antibody was injected.

Immobilization and activation of the sensor chip can be done manually or using according programmes of the Biacore 2000 software.

3.3.2 Analysis of binding affinity

Concentration series of pMHC complexes were injected over the chip using the same injection program for all measurements at 25° C (Table 6). Concentration series consist of the highest achieved concentration of refolded pMHC complex, and of 80 µM, 40 µM, 20 µM, 10 µM, 5 µM, 2.5 µM, 1.25 µM pMHC complexes. The chip surface was regenerated with regeneration buffer (0.1 M Glycine-HCl, 0.5 M NaCl pH = 2.5) to remove anti-C_β antibody H57-597. Unspecific binding was corrected for by subtracting responses from the reference flow cell.

Table 6 The standard program used for analysing TCR/pMHC binding affinity on Biacore2000.

Operation	Value	Unit
flow rate	5	[µL min^{-1}]
waiting time	30	[s]
injection TCR p14	10	[µL]
waiting time	180	[s]
injection pMHC	10	[µL]
dissociation time	120	[s]
waiting time	240	[s]
flow rate	30	[µL min^{-1}]
injection regeneration buffer	10	[µL]
waiting time	600	[s]

BIAevaluation 2000 software (BIAcore AB, Uppsala, Sweden) was used for data analyses. Binding affinities (K_D values) were obtained from steady-state fitting of equilibrium binding curves to following equation, where k_{on} is the association of TCR/pMHC complexes and k_{off} their dissociation.

$$K_D = k_{off}/k_{on}$$

3.3.3 Thermodynamic analysis

For thermodynamic analysis K_D values were measured, as described before, over a temperature range from 5 to 30 degrees in 5° C in tervals. The overall free binding energy ΔG was derived from each measured K_D value using the Gibbs free energy equation. Thermodynamic parameters enthalpy (ΔH), entropy (ΔS) and heat capacity (ΔC_p) were calculated by fitting the data to a non-linear van 't Hoff equation using GraphPad Prism 5 (GraphPad Software, La Jolla, USA).

$$\Delta G = RT \ln K_D$$

$$\Delta G = \Delta H_{T0} - T\Delta S_{T0} + \Delta C_p (T - T_0) - T\Delta C_p \ln(T - T_0)$$

The non-linear equation can also be written as

Y = A – X * (B - C + C * ln (X / 298.15)) – C * 298.15

where Y= ΔG, X = T, A = ΔH_{T0} (at 298.15 K), B = ΔS_{T0} (at 298.15 K), C = ΔC_p and T_0 = 298.15 K.

3.4 X-ray Crystallography

3.4.1 Crystallization assays

Usually cysteine residues are hardly found in gram positive bacteria, making the two present cysteines possible active site residues. Since previous crystallization trials with the wildtype form of LytA and N-LytA were unsuccessful, the two cysteine residues were mutated to alanines. In creating an inactive enzyme a more stable conformation of the molecule could be achieved and therefore make it more likely to form nucleation sites and well ordered crystals.

Soluble protein samples of wildtype LytA, LytA-C60/136A, wildtype N-LytA, N-LytA-C60/136A and phage LytA were kindly provided by Dr Peter Mellroth (Karolinska Institutet, Sweden). The LytA protein was produced in *Escherichia coli* Rosetta2 (DE3) cells and purified from lysates. The protein was soluble in the lysate and did thus not form any inclusion bodies. It was purified on DEAE sepharose by affinity binding in high salt buffer (50 mM phosphate buffer pH = 7.5, 500 mM NaCl) and eluted with 150 mM choline chloride in the same phosphate buffer. Final protein concentrations ranged between 10-20 mg mL^{-1}.

Several 96 well sitting drop crystallization microplates (Corning) were automatically set up and incubated at either 4 or 24° C. Screening solutions (INDEX, Crystal Screen 2, SaltRX) were purchased from Hampton Research and Quiagen (JCSG plus, PACT, PEGs, pHClear, pHClear II). Conditions showing crystals, microcrystals, polycrystals, nucleation sites, spherulites or other interesting formations were repeated using 24 well hanging drop crystallization plates having slightly variable conditions. The amino acid sequence of LytA is seen in Table 7.

Table 7 The amino acid sequence of the major autolysin LytA from *Streptococcus pneumoniae.*

Name	Autolysin LytA; N-acetylmuramoyl-L-alanine amidase
Organism	*Streptococcus pneumoniae*
Uniprot ID	P06653
Mass (Da)	36544
EC number	3.5.1.28
Sequence length	318

Aminoacid sequence

```
          10         20         30         40         50         60
MEINVSKLRT DLPQVGVQPY RQVHAHSTGN PHSTVQNEAD YHWRKDPELG FFSHIVGNGC
          70         80         90        100        110        120
IMQVGPVDNG AWDVGGGWNA ETYAAVELIE SHSTKEEFMT DYRLYIELLR NLADEAGLPK
         130        140        150        160        170        180
TLDTGSLAGI KTHEYCTNNQ PNNHSDHVDP YPYLAKWGIS REQFKHDIEN GLTIETGWQK
         190        200        210        220        230        240
NDTGYWYVHS DGSYPKDKFE KINGTWYYFD SSGYMLADRW RKHTDGNWYW FDNSGEMATG
         250        260        270        280        290        300
WKKIADKWYY FNEEGAMKTG WVKYKDTWYY LDAKEGAMVS NAFIQSADGT GWYYLKPDGT
         310
LADKPEFTVE PDGLITVK
```

Background – Sitting drop/hanging drop crystallization assays and seeding techniques

The underlining thermodynamically technique for both methods is the vapor diffusion. Purified protein is mixed with an equal amount of buffer and precipitant solution and either placed onto the desired sitting drop place or on a glass slide which is then used to seal the reaction chamber. In this closed system vapor diffusion (evaporation and condensation) will result in a net transfer of water from the protein solution to the higher concentrated reservoir (buffer and precipitant) solution until equilibrium is reached. At this stage concentrations of protein and precipitant are increased in the protein containing droplet, which can lead to the formation of nucleations. Nucleation, the first step of crystal growth, is the formation of molecular clusters. As a consequence of nucleation and crystal growth, the protein concentration decreases leading to optimal growth conditions when looking at a phase diagram for crystallization.
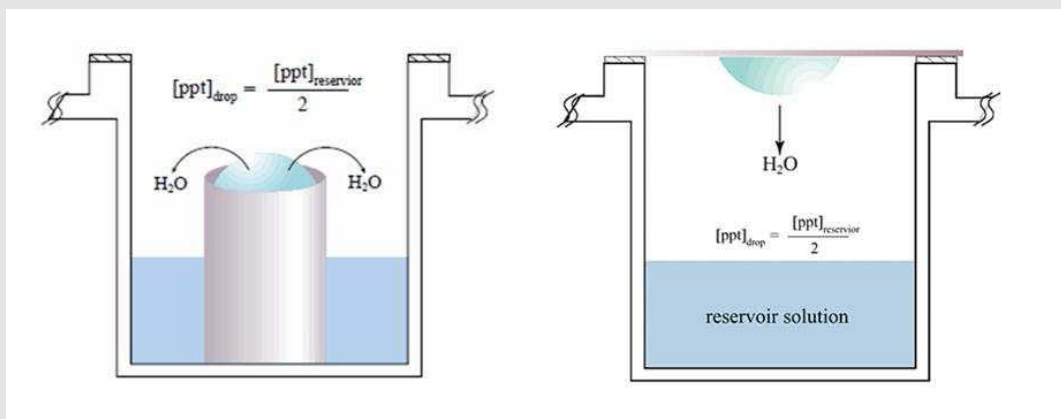
Since it may be difficult to obtain spontaneous nucleation, or diffraction properties of crystals do not meet our needs, one can support the formation of nucleation sites using seeding techniques. Also, the crystal quality and size is often improved, when seed crystals are used. Seed crystals are small crystals, parts of crystals or polycrystals providing a ready made nucleus or uncouple nucleation from crystal growth. Streak seeding and solution microseeding techniques are mostly used. When streak seeding is used, a stiff fiber is crushing parts of the crystal into small seeds. They are then transferred into the crystallization droplets. Microseeding solutions have the advantage to better control the number of applied seeds through dilution of the microseeding stock solution.[23, 25, 26]

Microseeding stock solution of LytA C60/136A was prepared using Seed Bead (Hampton Research). Manually crushed crystals were vortexed for 90 seconds in the PTFE (Polytetrafluorethylen) containing seed bead tube with 50 µL of the reservoir solution where the polycrystals were observed. 1 µL of the microseeding stock solution was usually added to the crystallization droplets of LytA C60/136A.[24]

Crystals of LytA C60/136A, N-LytA C60/136A and phage LytA were observed, cryomounted and stored in liquid nitrogen before performing X-ray diffraction analysis. The best successful crystallization conditions are listed in Table 8. Pictures of the obtained crystals can be found in chapter 4 Results and Discussion.

Mounting of crystals was done using appropriate sizes of cryoloops of the Crystalcap system (Hampton Research). To protect crystals from radiation damage they were suspended in a film of antifreeze solution (cryoprotectant solution) immediately after mounting into the cryoloop. The cryoprotectant solution contained the corresponding reservoir solution and 25 % glycerol. When the crystal is stored and handled in liquid nitrogen (100 K), water molecules will form vitrified water instead of crystalline ice.

Table 8 Aminoacid sequences and successful crystallization conditions are shown for the LytA C60/136A mutant, the N-LytA C60/136A mutant and the phage LytA. Bold and underlined letters display introduced mutations, red labelled letters display additional aminoacid residues originating from cloning procedures.

LytA C60/136A

Aminoacid sequence	
	` 10 20 30 40 50 60`
	`MEINVSKLRT DLPQVGVQPY RQVHAHSTGN PHSTVQNEAD YHWRKDPELG FFSHIVGNGA`
	` 70 80 90 100 110 120`
	`IMQVGPVDNG AWDVGGGWNA ETYAAVELIE SHSTKEEFMT DYRLYIELLR NLADEAGLPK`
	` 130 140 150 160 170 180`
	`TLDTGSLAGI KTHEYATNNQ PNNHSDHVDP YPYLAKWGIS REQFKHDIEN GLTIETGWQK`
	` 190 200 210 220 230 240`
	`NDTGYWYVHS DGSYPKDKFE KINGTWYYFD SSGYMLADRW RKHTDGNWYW FDNSGEMATG`
	` 250 260 270 280 290 300`
	`WKKIADKWYY FNEEGAMKTG WVKYKDTWYY LDAKEGAMVS NAFIQSADGT GWYYLKPDGT`
	` 310`
	`LADKPEFTVE PDGLITVK`

Method	hanging drop crystallization assay on 24 well plate
Droplet	1 μL protein solution, 1 μL reservoir solution, 0.5 μL microseeding stock solution
Buffer	0.1 M Sodiumacetate p*H* 4.3
Salt	0.2 M Zincacetate
Precipitant	PEG 3000 8 %
Incubation temperature	24°C

N-LytA C60/136A

Aminoacid sequence	
	` 10 20 30 40 50 60`
	`ASMEINVSKLRT DLPQVGVQPY RQVHAHSTGN PHSTVQNEAD YHWRKDPELG FFSHIVGNGA`
	` 70 80 90 100 110 120`
	`IMQVGPVDNG AWDVGGGWNA ETYAAVELIE SHSTKEEFMT DYRLYIELLR NLADEAGLPK`
	` 130 140 150 160 170`
	`TLDTGSLAGI KTHEYATNNQ PNNHSDHVDP YPYLAKWGIS REQFKHDIEN GL`

Method	hanging drop crystallization assay on 24 well plate
Droplet	2 μL protein solution, 2 μL reservoir solution
Buffer	0.1 M HEPES p*H* 6.8
Salt	1.0 M Lithiumchloride
Precipitant	PEG 6000 9 %
Incubation temperature	24°C

phage LytA, gene nr SPH_0121 from *Streptococcus pneumoniae* strain Hu19A-6

Aminoacid sequence	
	` 10 20 30 40 50 60`
	`MDIDRNRLRT GLPQVGVQPY RQVHAHSTGN RNSTVQNEAD YHWRKDPELG FFSHVVGNGR`
	` 70 80 90 100 110 120`
	`IMQVGPVNNG SWDVGGGWNA ESYAAVELIE SHSTKEEFMA DYRLYIELLR NLADEAGLPK`
	` 130 140 150 160 170 180`
	`TLDTDDLAGI KTHEYCTNNQ PNNHSDHVDP YPYLASWGIS REQFKQDIEN GLSAATGWQK`
	` 190 200 210 220 230 240`
	`NGTGYWYVHS DGSYPKDKFE KINGTWYYFD GSGYMLSDRW KKHTDGNWYY FDQSGEMATG`
	` 250 260 270 280 290 300`
	`WKKIADKWYY FDVEGAMKTG WVKYKDTWYY LDAKEGAMVS NAFIQSADGT GWYYLKPDGT`
	` 310`
	`LADKAEFTVE PDGLITVK`

Method	sitting drop crystallization assay on 96 well plate
Droplet	2 μL protein solution, 2 μL reservoir solution
Condition 1	
Buffer	0.1 M BIS-TRIS propane p*H* 7.0
Salt	0.7 M Sodiumcitrate tribasic dihydrate
Condition 2	
Buffer	0.1 M Tris p*H* 8.5
Salt	1.2 M Sodiumcitrate tribasic dihydrate
Condition 3	
Buffer	0.1 M BIS-TRIS propane p*H* 7.0
Salt	0.7 M Magnesium formate dihydrate
Incubation temperature	24°C

3.4.2 X-ray diffraction

The beamline ID29 at the ESRF (European Synchrotron Radiation Facility, Grenoble, France) was used for performing X-ray analysis of N-LytA C60/136A crystals while Pilatus 6M was used as a detector. An X-ray fluorescence scan was performed at the same beamline.

Data sets of the native crystal were collected. The first dataset gave the amplitudes of each reflection (reflection intensities, I_{hkl}). The second collected dataset, at wavelength where zinc ions show maximum anomalous scattering, was used to calculate phases, α_{hkl}. In the next step information of intensities and phases were put together to calculate the Fourier transform of all structure factors, which resulted in the electron density $\rho(x, y, z)$.

$$\rho(x, y, z) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} F_{hkl} e^{i\alpha_{hkl}} e^{-2\pi i(hx+ky+lz)}$$

Background - X-ray Fluorescence (XRF)

XRF is used for elemental analysis of samples. When high energy X-rays hit a material, ionization of atoms can occur. Due to the high energy of X-rays, strongly held electrons from the inner orbitals are expelled. The energetically unstable atom replaces the expelled electron with another electron from one of its outer orbitals.

As every element has electronic orbitals of characteristic energy, the replacement of an expelled electron results in a characteristic release of energy in form of a photon. Using the Planck-Einstein equation (λ = c * h / E), the wavelength of the released photon can be calculated and linked to their specific element. (Zinc transition takes place from shell L to K called $K\alpha_1$ transition releasing a wavelength of 0.1435 nm).[23, 55]
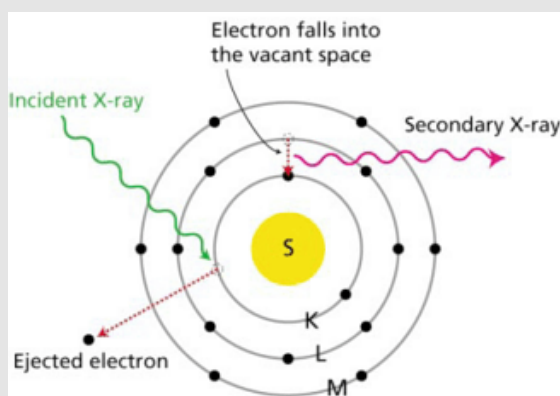
3.4.3 Data refinement and crystal structure

Data refinement was done using the program REFMAC from the CCP4 suite version 6.1.13.[40]

Having the electron density map, N-LytA´s main and side chains can automatically be placed into it and can manually be controlled and adjusted. Water molecules and other ligands can also be defined. The used program was the Crystallographic Object-Oriented Toolkit (Coot) version 0.6.1.[39]

3.5 Circular Dichroism

CD spectra of N-LytA and N-LytA C60/136A were recorded using a spectropolarimeter (Jasco J-810). Samples were diluted to a concentration of 1.1 and 0.9 mg mL^{-1} respectively. Salt containing buffer (20 mM Tris p*H* = 7.5) was exchanged to a 200 mM phosphatebuffer p*H* = 7.5 using NAP-5 columns (GE Healthcare). Spectra were recorded between a wavelength of 190 and 250 nm in 0.2 nm intervals using a 0.1 mm cuvette. Values of ellipticity (θ), given in millidegree (mdeg), were baseline corrected. To calculate secondary structure elements using the program K2D2[27], molar CD values (Δε) in dL mol^{-1} cm^{-1} were first calculated.

Background - Circular Dichroism (CD)

UV circular dichroism measures the different absorbance of left and right circularly polarized light of optically active chiral molecules. As a result the circular polarization changes to an elliptical polarization. The ratio of the minor to the major axis of the ellipse is by definition the tangent of angle θ, which is called ellipticity.

$$\tan \theta = (E_R - E_L) / (E_R + E_L)$$

Together with absorbance of the incident light beam, refraction occurs too, resulting in an optical rotation, α.
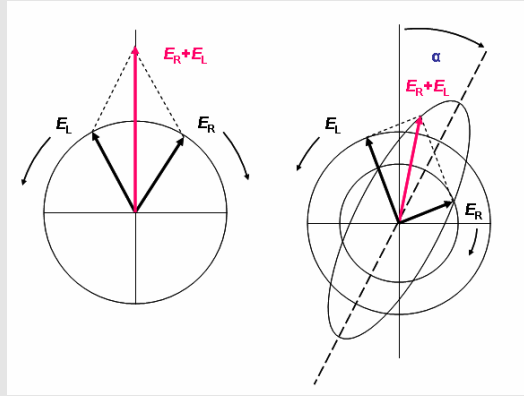

To remove dependence of cuvette path and solute concentration, molar ellipticity can be calculated using following equation,

$$[\theta] = (100 / \theta) / (c * d)$$

where c is the concentration in mol L^{-1} and d the cuvette path length in cm. Molar CD values ($\Delta\varepsilon$), which were used for secondary structure prediction, are linked to the molar ellipticity by the following conversion factor given in L mol^{-1} cm^{-1} units.

$$\Delta\varepsilon = [\theta] / 3298$$

The main secondary structure elements (helices, sheets, coils) show characteristic CD spectra. Algorithms have been developed to predict the percentage of secondary structure elements of structurally unknown protein samples. CD can not state where the secondary structures are located within a molecule, but it can easily monitor changes in the conformation, due to denaturation, amino acid replacement or ligand binding.[28, 29, 56, 57]
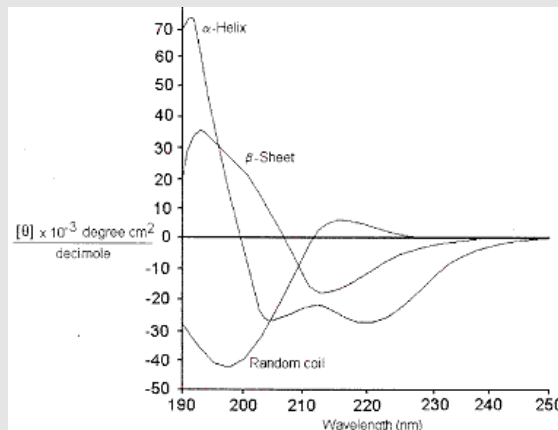
4 Results and Discussion

4.1 Structural and biochemical studies between the T-cell receptor p14 and MHC class I complexes loaded with the LCMV derived antigen gp33

4.1.1 Refolding of H-2D^b/gp33/mβ$_2$m and H-2D^b/gp33-Y4A/mβ$_2$m MHC complexes

Molecular weights of the MHC subunits mβ$_2$m (MW = 11687 Da) and H-2D^b (MW = 32047 Da) were controlled using gel-electrophoresis. Before refolding H-2D^b/gp33/mβ$_2$m and H-2D^b/gp33-Y4A/mβ$_2$m a Coomassie stained SDS gel was performed displaying typical sizes of the subunits. (Figure 14)
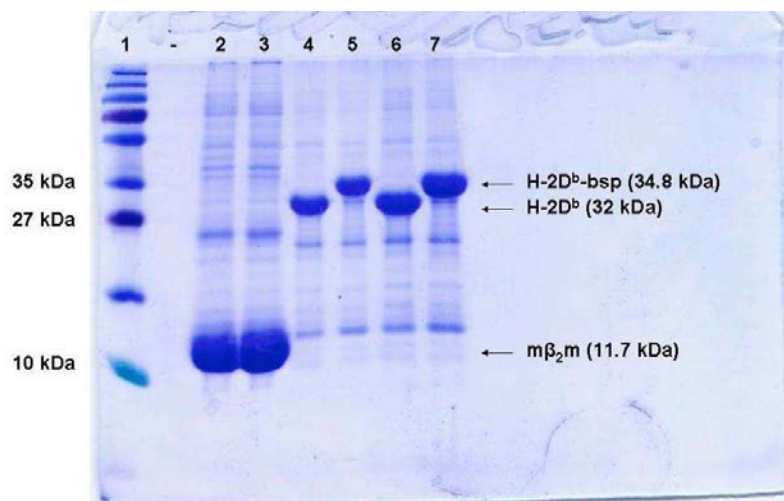
Figure 14 SDS Coomassie stained electrophoresis gel of MHC I subunits mβ$_2$m (MW = 11687 Da), H-2D^b (MW = 32047 Da) and H-2D^b-bsp (MW = 34843 Da). Lane number **1** shows 5 µL of PageRuler Plus Prestained Protein Ladder (10-250 kDa) (Fermentas), lane number **2** consists of 3 µL mβ$_2$m (6.6 mg mL^{-1}), lane number **3** consists of 3 µL mβ$_2$m (8.0 mg mL^{-1}), lane number **4 and 6** are 1.3 mg mL^{-1} H-2D^b containing 3 and 6 µL respectively, lane number **5 and 7** show the subunit H-2D^b-bsp (1.7 mg mL^{-1}) with 3 and 6 µL respectively. Samples in lanes 2 to 7 were mixed with Coomassie Brilliant Blue containing DTT in a ratio of 1:1 prior to heating at 80-90 ℃ for 10 minutes.

Following refolding and concentration pMHC complexes were purified using size-exclusion chromatography. Aggregations (incorrect folded pMHC complexes and unbound mβ$_2$m subunits) were separated from stable pMHC complexes which were then used for binding affinity assays. Two typical chromatograms are shown in figures 15 and 16. Blue lines show the UV absorption at 280 nm scaled on the Y-axis in mAU. At 280 nm proteins show an absorbance peak due to light absorption by aromatic amino acids, like tyrosine, tryptophan and phenylalanine. Red lines show the UV adsorption at 254 nm indicating DNA impurities. The X-axis

is scaled according to the amounts of eluted millilitre. The allocated fractions are also indicated. In order to separate the pMHC peak totally from aggregation peaks, size exclusion chromatography was performed a second time when necessary.
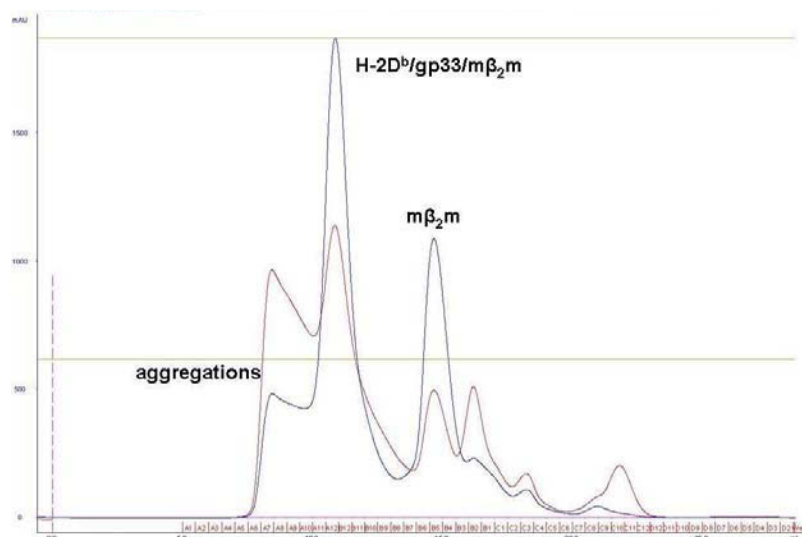

Figure 15 0.5 mL of a H-2D^b/gp33/mβ_2m complex was purified by size exclusion chromatography (Superdex 75 10/300 GL) using HEPES running buffer (10 mM HEPES, 0.15 M NaCl, 0.005 % (v/v) Tween20).
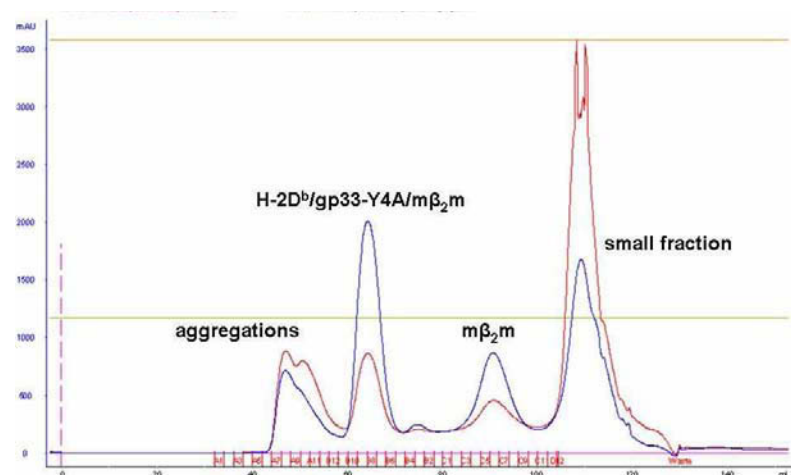
Figure 16 5 mL of a H-2D^b/gp33-Y4A/mβ_2m complex was purified by size exclusion chromatography (Superdex 75 prep grade 16/60 column) using HEPES running buffer (10 mM HEPES, 0.15 M NaCl, 0.005 % (v/v) Tween20).

4.1.2 Refolding of the TCR p14

The purity of both the p14α (MW = 22213 Da) and p14β (MW = 26848) chains of the TCR p14 were controlled using SDS Coomassie stained gel-electrophoresis. (Figure 17)
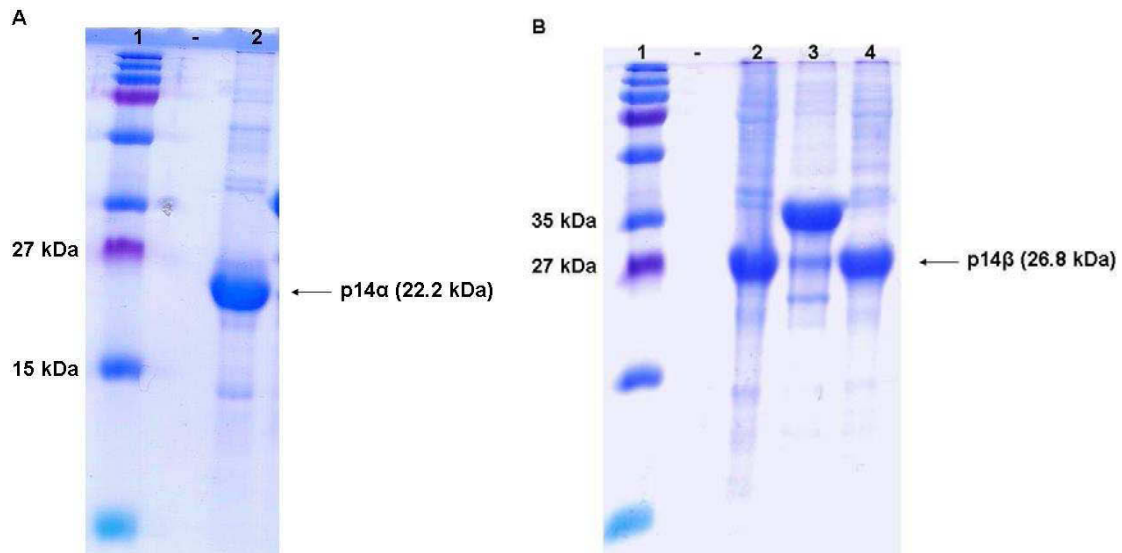
Figure 17 SDS Coomassie stained gels with purified p14 subunits. **A** The subunit p14α (MW = 22213 Da). Lane number **1** shows 5 µL of PageRuler Plus Prestained Protein Ladder (10 - 250 kDa) (Fermentas), lane number **2** shows 4 µL of a 30 mg batch of p14α. The sample in lane 2 was mixed with Coomassie Brilliant Blue containing DTT in a ratio of 1:1 prior to heating at 80-90 ℃ for 10 min. **B** The subunit p14β (MW = 26848 Da). Lane number **1** shows 5 µL of PageRuler Plus Prestained Protein Ladder (10 - 250 kDa) (Fermentas), lane number **2** shows 4 µL of a 14 mg batch, in lane **3** the majority of protein is not at the molecular size according to p14β and was therefore not used for TCR refolding, lane **4** shows 4 µL of a 11 mg batch of p14β. Samples in lanes 2 to 4 were mixed with Coomassie Brilliant Blue containing DTT in a ratio of 1:1 prior to heating at 80-90 ℃ for 10 minutes.

After refolding and concentration, the TCR p14 was purified in two steps. The first purification step was performed using an ion-exchange column (5 mL HiTrap Q FF). Fractions were eluted with a linear increasing salt gradient from 0–100 % in 30 minutes (10 mM Tris p*H* = 8.3, 1.0 M sodiumchloride). A typical chromatogram is shown in figure 18, where blue lines show the UV absorption at 280 nm scaled on the Y-axis in mAU, red lines show the UV adsorption at 254 nm indicating DNA impurities. The green line indicates the linear increasing concentration of the elution buffer. The X-axis is scaled according to the amount of eluted millilitre, allocated fractions are also indicated.

As an additional quality control, SDS gel electrophoresis of the fractions from the first peak was performed. Successfully refolded TCR p14 has a MW of 49822 Da, (Figure 19). A second purification was performed by size exclusion chromatography using a Superdex 75 10/300 GL column equilibrated with HEPES buffer (10 mM HEPES, 0.15 M NaCl, 0.005 % (v/v) Tween20). (Figure 20)
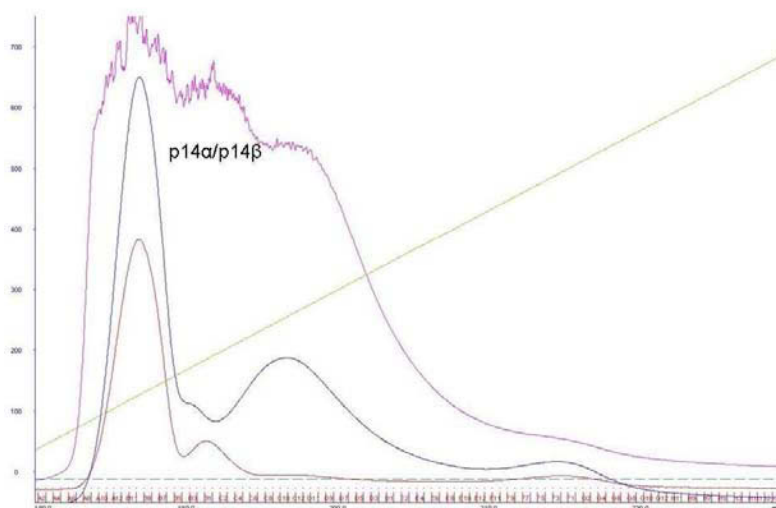
Figure 18 Ion exchange chromatogram of a 60 mL p14α/p14β refolding sample. The used running buffer was HEPES (10 mM HEPES, 0.15 M NaCl, 0.005 % Tween20), the linear increasing elution buffer was 10 mM Tris p*H* = 8.3, 1.0 M sodiumchloride (green). Fractions under the first peak were collected (A11, A12, B12, B11, B10, B9, B8, B7, B6) and their molecular weight and purity controlled.
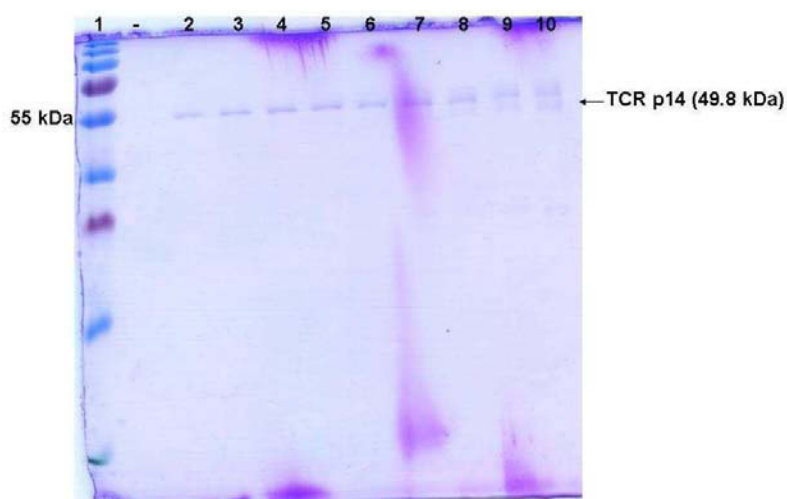
Figure 19 An SDS Coomassie stained electrophoresis gel of refolded p14 following the first purification step. Lane number **1** shows 5 µL of PageRuler Plus Prestained Protein Ladder (10 - 250 kDa) (Fermentas), lane number **2 – 10** show each a 10 µL sample of fractions A11, A12, B12, B11, B10, B9, B8, B7 and B6. Samples in lanes 2 to 10 were mixed with Coomassie Brilliant Blue in a ratio of 1:1. An electrical current of 180 V was applied for 60 minutes. Samples in lane 2 – 6 show pure p14, whereas samples in lane 7 – 10 needed a second purification on a Superdex 75 10/300 GL column.
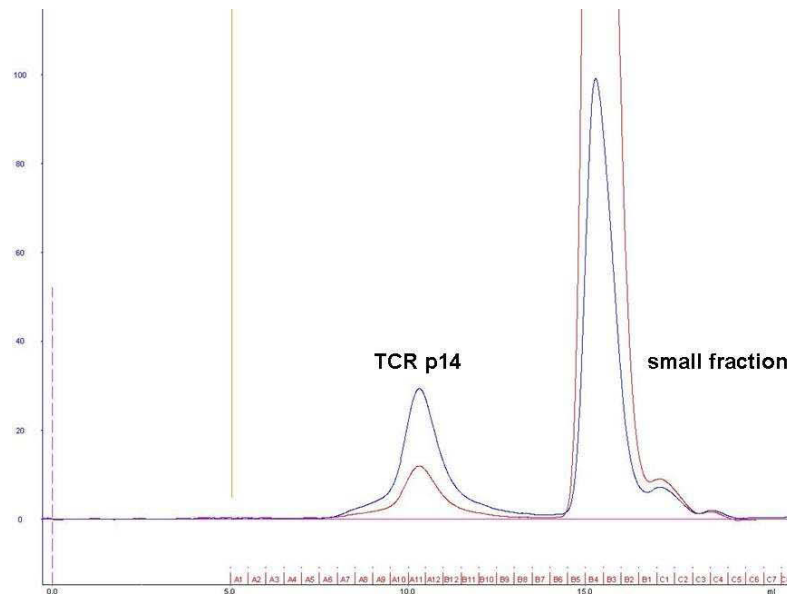
Figure 20 Second purification of refolded p14 was performed using size exclusion chromatography with a Superdex 75 10/300 GL column equilibrated with HEPES buffer (10 mM HEPES, 0.15 M NaCl, 0.005 % (v/v) Tween20).

4.1.3 Analysis of binding affinity

The binding affinities of the TCR p14 to the H-2D^b/gp33 and H-2D^b/Y4A complexes were measured at 25° C using SPR as described in th e chapter Materials and Methods. The binding affinity of p14 to H-2D^b/gp33 was measured to 8.9 µM and is within the frame of K_D-values previously reported for the same interaction.[21, 72, 73] The binding affinity of p14 to H-2D^b/Y4A was determined to 54.5 µM. No binding was observed between p14 and H-2D^b/Y4S and H-2D^b/Y4F as described by Allerbring et al.[79] In figures 21 and 22 typical Biacore sensorgrams are displayed, showing the association, the steady-state equilibrium and the dissociation of the TCR p14 binding to the pMHC complexes. Steady-state fitting of the curves as a function of the pMHC concentrations is also seen in figures 21 and 22 which were used to calculate the K_D-values. Functional comparative analyses of recognition of H-2D^b in complex with gp33, Y4A, Y4S or Y4F by p14 were previously carried out. It was demonstrated that gp33 and Y4A were able to proliferate CD8$^+$ p14 T-cells, while Y4S and Y4F did not induce proliferation.[78]
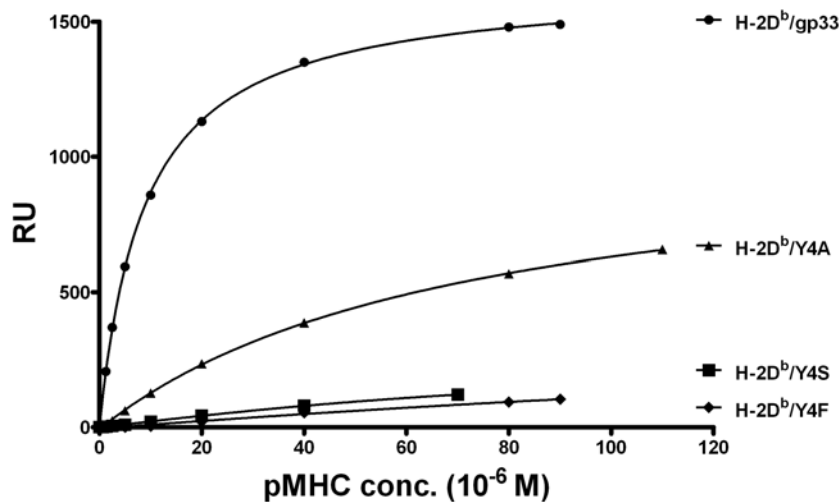
When plotting equilibrium binding responses of each pMHC to the TCR p14 as a function of pMHC concentration in one diagram, the peptide hierarchy, gp33>Y4A>Y4S=Y4F, is in agreement with the functional data. (Figure 23)
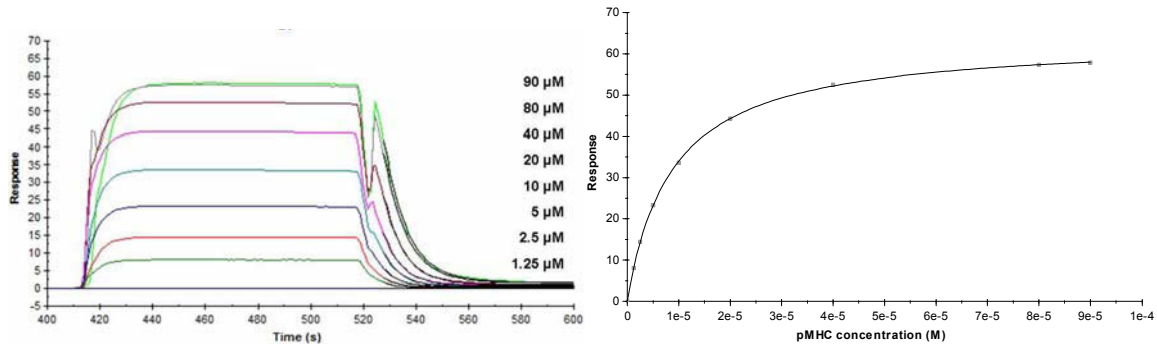
Figure 21 Left A typical sensorgram of the association, the steady-state equilibrium and the dissociation of the TCR p14 binding to H-2D^b/gp33 is displayed. The used concentrations of the pMHC molecules are indicated, the injection points were set to 0 seconds and data were subtracted with reference data having 0 µM. **Right** The K_D-value was calculated from steady state fitting of curves using the BIAevaluation software. In this example, TCR p14 binding to H-2D^b/gp33, a K_D value of 8.85 µM was derived.
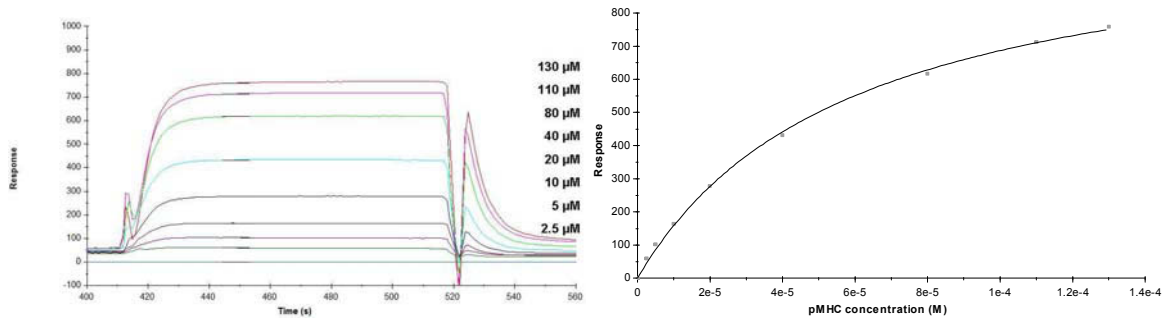
Figure 22 Left A typical sensorgram of the association, the steady-state equilibrium and the dissociation of the TCR p14 binding to H-2D^b/Y4A is displayed. The used concentrations of the pMHC molecules are indicated, the injection points were set to 0 seconds and data were subtracted with reference data having 0 µM. **Right** The K_D-value was calculated from steady state fitting of curves using the BIAevaluation software. In this example, TCR p14 binding to H-2D^b/Y4A, a K_D value of 58.6 µM was derived.

Figure 23 Equilibrium binding response of each pMHC to p14 as a function of pMHC concentration. Solid lines represent non-linear square fit to a one-site specific binding using GraphPad Prism 5. K_D-values were calculated from steady state fitting of curves using the BIAevaluation software, following subtraction of reference responses. K_D-values for binding of p14 to H-2D^b/gp33 and H-2D^b/Y4A were measured to 8.9 and 54.5 µM, respectively. The TCR p14 did not bind to H-2D^b/Y4S and H-2D^b/Y4F.[78]

Kinetic binding parameters were also calculated for the interaction of p14 with H-2D^b/gp33 and H-2D^b/Y4A. (Table 9) The dissociation rates of p14 from H-2D^b/gp33 and H-2D^b/Y4A can be considered as similar with k_{off} values of 0.5 s^{-1} and 0.7 s^{-1}, respectively. The association rates were derived from the equation $k_{on}= k_{off}/K_D$ to 0.56×10^5 M^{-1}s^{-1} and 0.13×10^5 M^{-1}s^{-1} for the interaction of p14 with H-2D^b/gp33 and H-2D^b/Y4A, respectively. These values indicate a slower association rate of p14 with H-2D^b/Y4A when compared to H-2D^b/gp33.

Table 9 Kinetic parameters from p14 binding to gp33 and Y4A in complex with H-2D^b measured with Biacore2000 at 25°C.

Peptide	Sequence	K_D [a] [µM]	K_{off} [b] [s^{-1}]	K_{on} [c] [10^5 M^{-1} s^{-1}]	$t_{1/2}$ [d] [s]	ΔG [e] [kcal mol^{-1}]
gp33	KAVYNFATM	8.9+0.4	0.5+0.003	0.60	1.4+0.03	-6.8
gp33-Y4A	KAVANFATM	54.5+0.3	0.7+0.004	0.13	1.0+0.08	-5.7
gp33-Y4F	KAVFNFATM	N.D.	N.D.	N.D.	N.D.	N.D.
gp33-Y4S	KAVSNFATM	N.D.	N.D.	N.D.	N.D.	N.D.

[a] Determined from steady state SPR data with the BIAevaluation software. The value is a mean +SD of two independent experiments.
[b] Determined from kinetic SPR data with the BIAevaluation software. The value is a mean +SD of two independent experiments.
[c] Calculated from $K_{off}/K_D = K_{on}$
[d] Calculated from ln2/K_{off}
[e] Calculated from $\Delta G = RT$ ln K_D
N.D. = Non-determinable

4.1.4 Analysis of thermodynamical data

Binding affinity values were calculated from steady state fitting of curves using the BIAevaluation software for binding of the TCR p14 to H-2D^b/gp33 and H-2D^b/Y4A in a temperature range between 5° and 30°C. (Table 10)

Table 10 Binding affinity (K_D) values between 5° and 30°C for the TCR p14 to H-2D^b/gp33 and H-2D^b/Y4A interaction were calculated from steady state fitting of curves using the BIAevaluation software and are listed according to the applied temperatures in Kelvin during binding experiments. Underlined values are those where the author was directly involved in carrying out the experiment.

H-2D^b/gp33	278 K	283 K	287 K	293 K	298 K	303 K
K_D1	3.88	5.18	4.81	8.48	8.70	18.4
K_D2	4.15	5.02	6.12	6.54	9.18	18.2
K_D3		4.79				
H-2D^b/Y4A						
K_D1	43.3	37.8	49.1	55.8	54.3	61.2
K_D2	54.7	45.2	50.6	62.0	54.7	52.4
K_D3		41.6	61.0			77.6

Thermodynamic parameters enthalpy (ΔH), entropy (ΔS) and heat capacity (ΔC_p) were calculated by fitting SPR data to a non-linear van 't Hoff equation as described in chapter 3 Materials and Methods. (Table 11, Figure 24). The thermodynamic signature of recognition used by the TCR p14 was different for the two pMHC complexes H-2D^b/gp33 and H-2D^b/Y4A. While the recognition of H-2D^b/gp33 by p14 was enthalpy-driven, the recognition of the weak agonist H-2D^b/Y4A was only entropy-driven. (Figure 25)
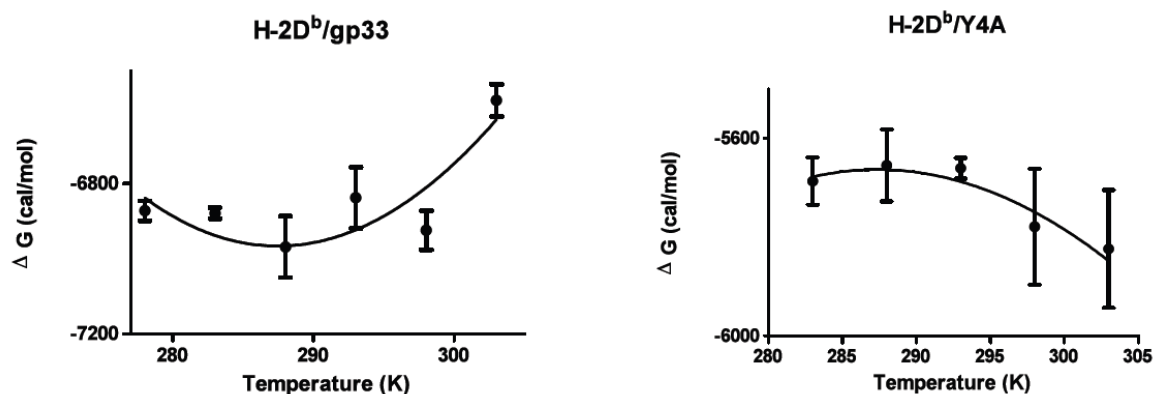

Figure 24 K_D-values were measured over a temperature range between 5 to 30°C. The free Gibbs energy values ΔG were calculated for each temperatures using equation $\Delta G = RT \ln K_D$. Each datapoint is the mean of at least two independent experiments.

Table 11 Thermodynamic parameters from p14 binding to H-2D^b in complex with gp33 and Y4A at 25°C.

Peptide	Sequence	ΔH [a] [kcal mol^{-1}]	$T\Delta S$ [a] [kcal mol^{-1}]	ΔG [b] [kcal mol^{-1}]
gp33	KAVYNFATM	-15.7	-8.9	-6.8
gp33-Y4A	KAVANFATM	-0.5	5.2	-5.7

[a] Calculated from fitting experimental values to a non-linear van 't Hoff equation using GraphPad Prism 5
[b] Calculated from $\Delta G = RT \ln K_D$
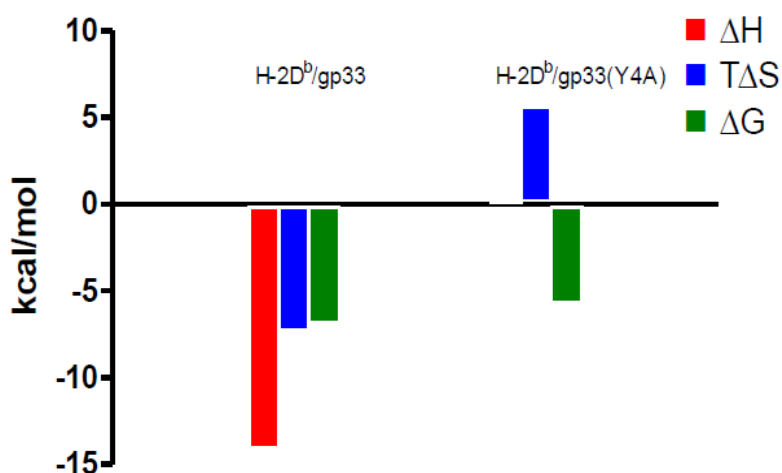
Figure 25 Thermodynamic analysis revealed that while recognition of H-2D^b/gp33 by p14 was enthalpy-driven, recognition of H-2D^b/Y4A by the same TCR was entropy-driven. The energetic profiles of the interactions are described by the overall binding energy (ΔG) as well as its enthalpy (ΔH) and entropy ($T\Delta S$) components.

4.1.5 Structural explanation of diametrically opposed thermodynamic signatures

Crystal structures of H-2D^b/gp33, H-2D^b/Y4A, H-2D^b/Y4S and H-2D^b/Y4F were previously solved.[78] With a predicted model of the TCR p14, p14/H-2D^b/gp33 and p14/H-2D^b/Y4A complexes could be modelled. Comparative analysis of pMHC crystal structures revealed, that the peptides are bound to H-2D^b in the same way. Peptide residues p5N and p9M act as main anchor positions, residue p3V act as a secondary anchor. The main T-cell recognition residues at position 4 protrude towards the TCR. Subtle differences in the arrangement of p1K are seen for H-2D^b/Y4A. (Figure 26)
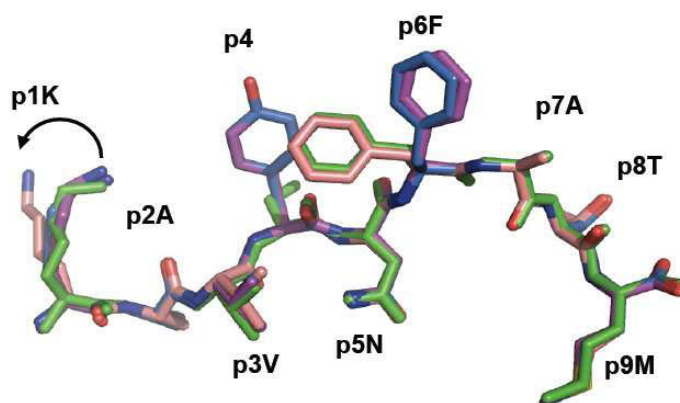

Figure 26 Superposition of LCMV derived gp33 (blue) and APL gp33-Y4A (pink), gp33-Y4F (violet) and gp33-Y4S (green). The anchoring residues to H-2D^b at positions 5, 9 and 3 are clearly pointing downwards. The TCR recognition residues at positions 1, 4 and 6 are pointing upwards. Subtle changes in the arrangement of p1K in the case of H-2D^b/Y4A are indicated with an arrow.

The favourable enthalpic part of the recognition of H-2D^b/gp33 by the TCR p14 can be explained through the formation of several hydrogen bonds between residues Y36(Vα) and G102(Vβ) of the TCR to the main T-cell recognition residue p4Y. Another hydrogen bond between p4Y and residue H155 from the MHC main chain is formed. (Figure 27, A) The unfavourable entropic term likely reflects an inherent loss of flexibility, consistent with the potential ordering of p14 CDR loops upon binding to H-2D^b/gp33.

Partial recognition of H-2D^b/Y4A by the same TCR is exclusively entropy-driven. The reduction of the enthalpic term may be explained by the loss of several hydrogen bonds upon mutating the main TCR contact from tyrosine to alanine at position 4. In comparison to H-2D^b/gp33 the side chain from alanine is too far away to form a hydrogen bond interaction with residue Y36(Vα). Due to the subtle changes in the arrangement of lysine at position 1 in H-2D^b/Y4A and hinge-bending movements of CDR1α, two new hydrogen bonds can be formed. One hydrogen bond interaction is formed between p1K and residue T33(Vα), another one between T33(Vα) and residue R62 from MHC α$_1$ helix, which is possibly important in the recognition process by the TCR. (Figure 27, B) The favourable entropy could be related to the capacity to release water molecules bound to H-2D^b/Y4A before the TCR p14 binds to the pMHC. That an exclusion of water molecules from the peptide binding groove can result in a favourable entropic term was recently demonstrated by Ely.[74]
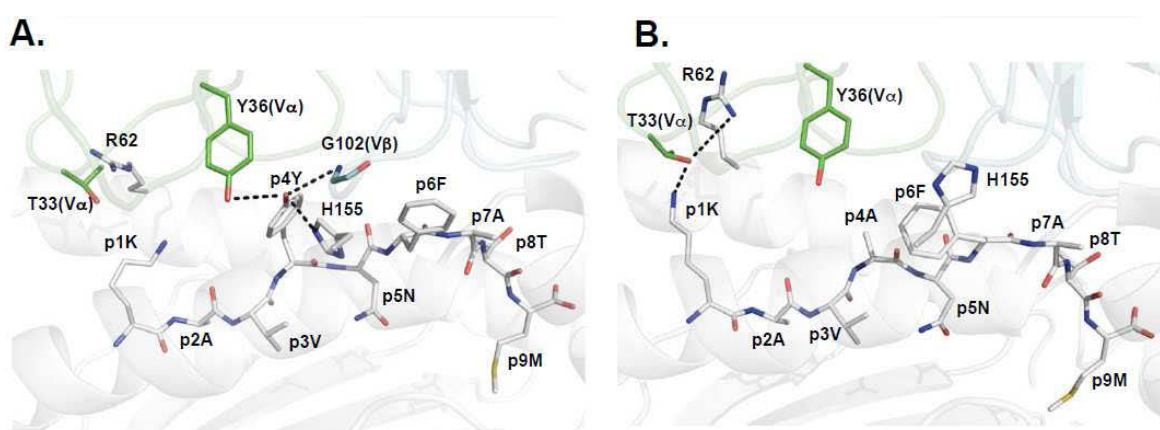
Figure 27 The pMHC interface during recognition by the TCR p14. **A** H-2D^b/gp33 is recognized by p14 via its main TCR residue at position 4. p4Y forms hydrogen bonds with Y36(Va) and G102(V) from the TCR and with residue H155 from MHC main chain. **B** The distance between TCR residue Y36(V) and p4A from H-2D^b/Y4A is too far in order to form a hydrogen bond. Alternatively, the partial recognition of Y4A may be achieved via the formation of two hydrogen bonds at the secondary hot spot around p1K.

4.1.6 Concluding remarks

TCR p14 can recognize H-2D^b/gp33 and H-2D^b/Y4A while APL with even smaller changes compared to Y4A at position 4 (phenylalanine in Y4F, serine in Y4S) are not recognized by the TCR.

Removal of the hydroxyl group in the antagonist Y4F enhances considerably the hydrophobicity of this peptide residue and abolishes several hydrogen bond interactions formed between p4Y and p14 explaining the total loss of affinity and functional recognition of p14 to H-2D^b/Y4F. The distance between the hydroxyl group of p4S in H-2D^b/Y4S and the TCR may be too large in order to build a hydrogen bond network around these residues explaining the non-recognition of H-2D^b/Y4S by the TCR p14. The recognition of H-2D^b/Y4A can be structurally explained, and results in a different thermodynamic signature compared to H-2D^b/gp33. The structural explanation includes the loss of several hydrogen bonds at p4 and the formation of a secondary hot spot around p1 when recognizing the weak-agonist H-2D^b/Y4A. Thermodynamic data further suggest changes in the entropic term of the reaction, which can be explained by the expulsion of several water molecules. The slower association rates for Y4A compared to gp33 may reflect the additional time needed for structural rearrangements of the CDR1 loop, especially around residue T33(Vα), when recognizing the altered peptide ligand. Those structural rearrangements can also contribute positively to the entropic term. Why the TCR p14 does not make use of a secondary hotspot at position p1 when peptides gp33-Y4F or gp33-Y4S are presented is not known.

Although binding of pMHCs to TCRs has been described as an enthalpically driven process, several studies demonstrated entropically driven recognitions too.[68] Diametrically opposed thermodynamic recognitions of APLs by TCRs were also demonstrated by Krogsgaard[71], Iversen[75] and Persaud[76]. Thus, at the present time, no clear-cut conclusion can be drawn regarding the strategy used by TCRs in order to recognize their cognate ligands.

4.2 The structure and function of the catalytic domain of *Streptococcus pneumoniae* major autolysin LytA

4.2.1 Circular Dichroism measurements of N-LytA, N-LytA C60/136A, LytA and LytA C60/136A

CD measurements revealed that the introduction of two alanine residues at position 60 and 136 did not change the overall secondary structure conformation of N-LytA and LytA. Therefore it can be assumed that the solved N-LytA C60/136A crystal structure has got the same conformation as the wildtype N-LytA.

Figures 28 and 29 illustrate the overlaid CD spectra of N-LytA with N-LytA C60/136A and LytA with LytA C60/136A. Protein secondary structure prediction was performed using the deconvolution programme K2D2 after normalization of the CD spectra data.[27] (Table 12)
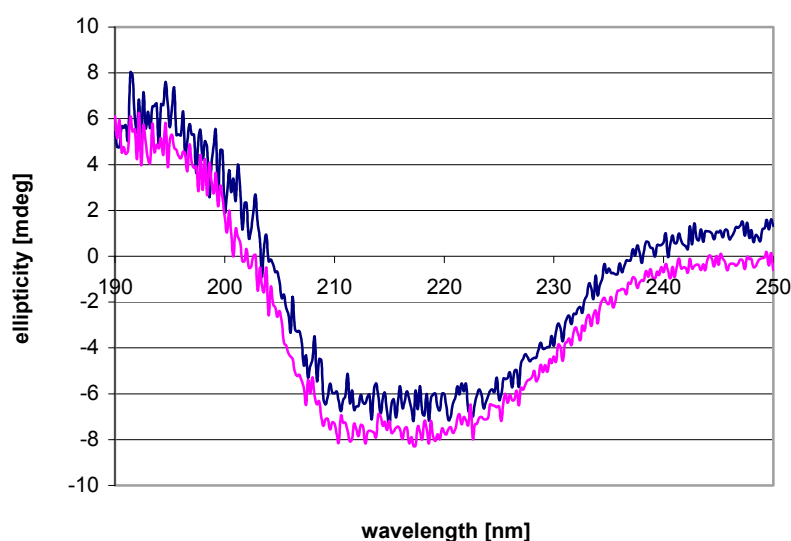
Figure 28 Overlaid CD spectra of N-LytA (blue) and N-LytA C60/136A (pink) reveal that there is no change in the overall conformation when positions 60 and 136 are mutated to cysteines. Spectra were recorded between a wavelength of 190 and 250 nm in 0.2 nm intervals using a 0.1 mm cuvette on a Jasco J-810 spectropolarimeter Concentrations of 1.1 mg ml^{-1} for N-LytA and 0.9 mg mL^{-1} for N-LytA C60/136A were used.
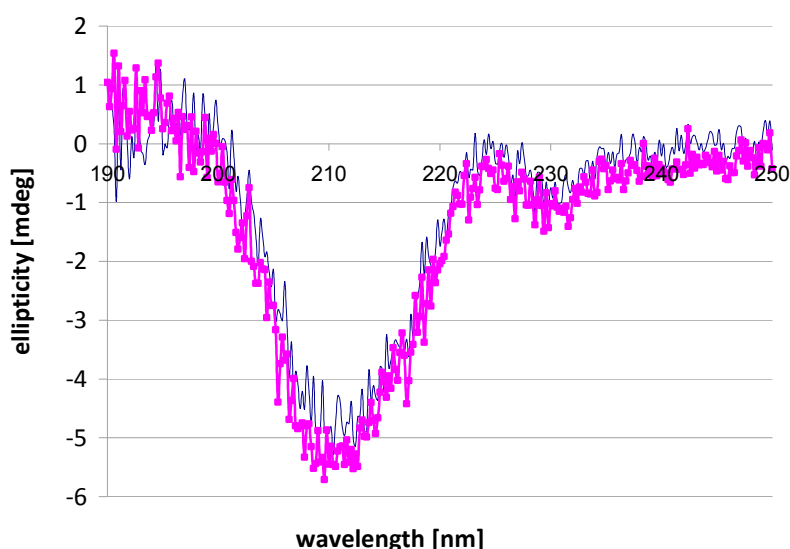
Figure 29 Overlaid CD spectra of LytA (blue) and LytA C60/136A (pink) reveal that there is no change in the overall conformation when positions 60 and 136 are mutated to cysteines. Spectra were recorded between a wavelength of 190 and 250 nm in 0.2 nm intervals using a 0.1 mm cuvette on a Jasco J-810 spectropolarimeter Concentrations of 1.0 mg ml^{-1} for LytA and 1.2 mg mL^{-1} for LytA C60/136A were used.

Table 12 Predicted secondary structure elements using the program K2D2 (http://www.ogic.ca/projects/k2d2/). 50 datapoints, in the range of 190 – 240 nm, in $\Delta\varepsilon$ values (dL mol^{-1} cm^{-1}) were included for prediction.[27]

	α helix [%]	β sheet [%]
N-LytA	32	15
N-LytA C60/136A	38	9
LytA	11	35
LytA C60/136A	11	35

4.2.2 Crystallization of the LytA C60/136A and the N-LytA C60/136A mutant and phage LytA

Crystallization trials were carried out using the full length LytA and N-LytA with and without mutations. However, only for the mutated form of full length LytA (LytA C60/136A) and the mutated form of N-LytA (N-LytA C60/136A) crystallization conditions were found. Phage LytA had no mutations, which makes the observed crystallization condition interesting.
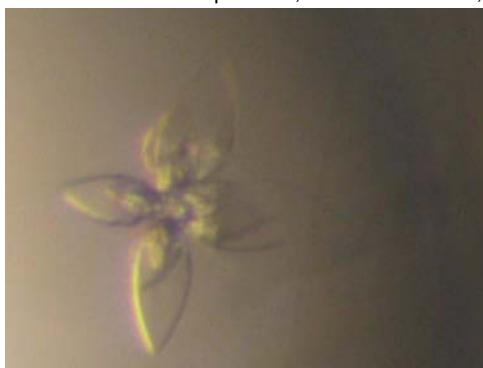
4.2.2.1 LytA C60/136A

Two crystallization conditions were established for LytA C60/136A. First, a flower-rosetta like crystal with thin leaves pointing outwards and rod like crystals were observed in 0.1 M Sodiumacetate pH = 4.3, 0.2 M Zincacetate and PEG 3000 8 % used as the precipitant. None of the crystals diffracted. Two crystal plates with sharp edges formed in 0.2 M Potassiumsulfate and PEG 3350 20% as a precipitant. One had a more elongated shape when compared to the other. Diffraction trials of crystals from the second crystallization condition were not performed until present. Repetitions of this condition were unsuccessful until present.

LytA C60/136A

0.1 M Sodiumacetate pH = 4.3, 0.2 M Zincacetate, PEG 3000 8 %

0.2 M Potassiumsulfate, PEG 3350 20%
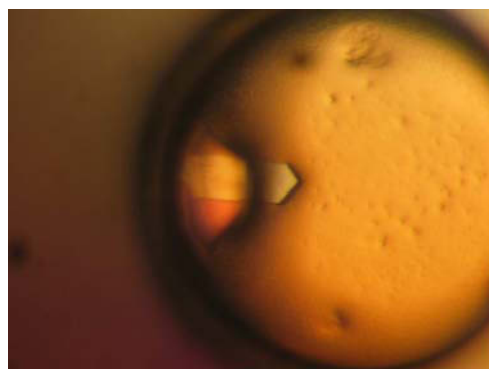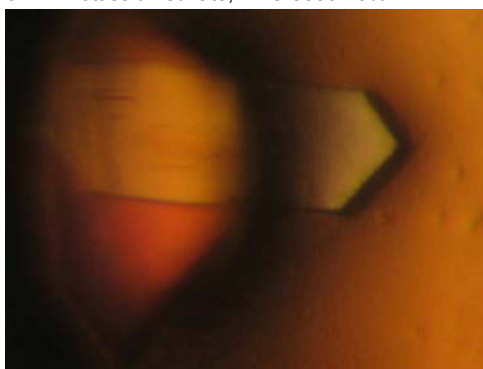
4.2.2.1 LytA C60/136A

4.2.2.2 N-LytA C60/136A

Rod like crystals were formed from N-LytA C60/136A after one to two days that grew to approximately 0.2 x 0.2 x 0.6 mm. The precipitant consisted of 0.1 M HEPES p*H* = 6.8 – 7.1, 1.0 M Lithiumchloride and PEG 6000 9 – 11 %. Crystals diffracted to 1.0 Ångström resolution.

N-LytA C60/136A

0.1 M HEPES p*H* = 6.8 – 7.1, 1.0 M Lithiumchloride and PEG 6000 9 – 11 %

4.2.2.3 Phage LytA

Phage LytA formed rod like plates in three different crystallization conditions; condition 1 (0.1 M BIS-Tris propane p*H* = 7.0, 0.7 M Sodiumcitrate dihydrate), condition 2 (0.1 M Tris p*H* = 8.5, Sodiumcitrate tribasis dihydrate) and condition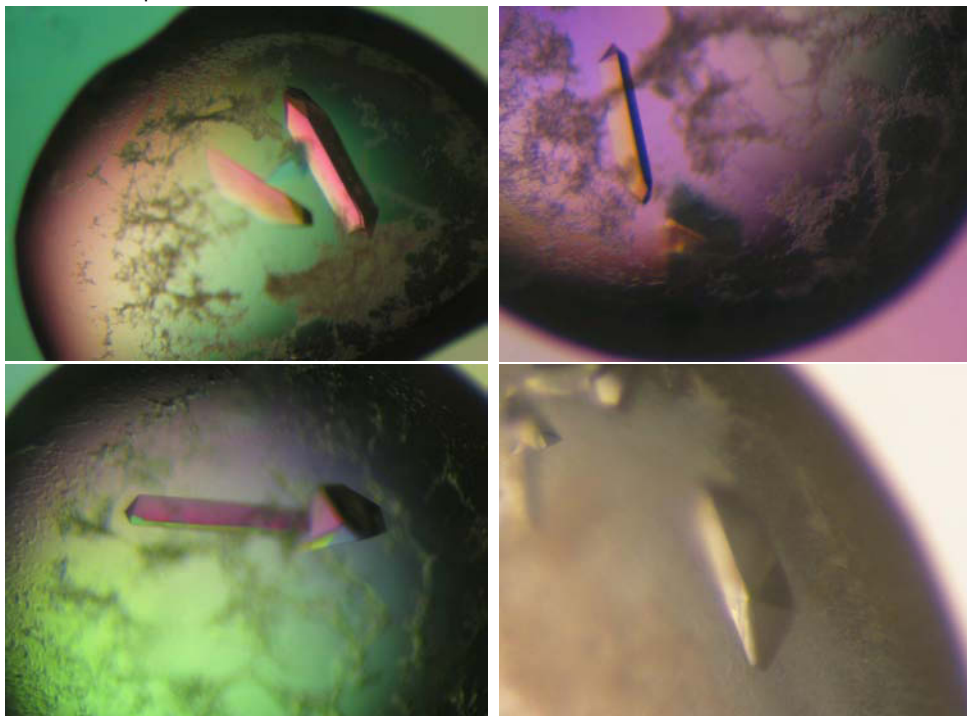 3 (0.1 M BIS-Tris propane p*H* = 7.0, 0.7 M Magnesiumformate dihydrate). All different phage LytA crystals have not yet been analysed regarding their X-ray diffraction potential.
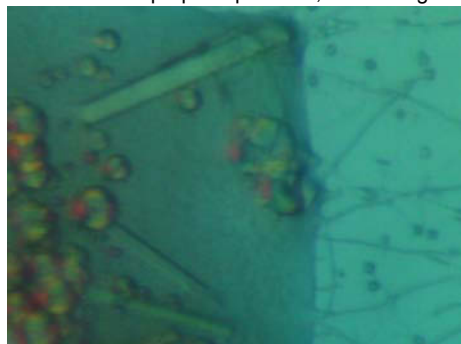
phage LytA

0.1 M BIS-Tris propane p*H* = 7.0, 0.7 M Sodiumcitrate dihydrate


0.1 M Tris p*H* = 8.5, Sodiumcitrate tribasis dihydrate

0.1 M BIS-Tris propane p*H* = 7.0, 0.7 M Magnesiumformate dihydrate

4.2.3 Determination of the crystal structure of the catalytic domain of *Streptococcus pneumoniae* major autolysin LytA

The presence of zinc atoms was detected by fluorescence X-ray diffraction. In figure 30 the absorbance spectrum at multiple wavelengths is displayed, indicating the zinc peak. One exemplarily diffraction pattern of N-LytA X-ray diffraction is displayed in figure 30. Data refinement was done using the program REFMAC from the CCP4 suite[40], the data collection statistics are shown in table 13.
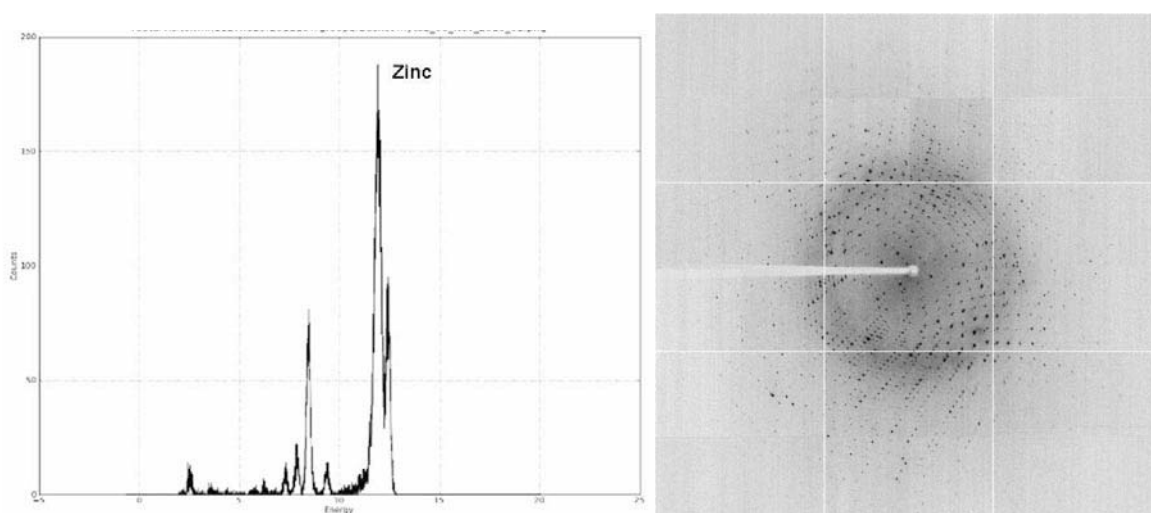
Figure 30 The fluorescence X-ray absorbance spectrum is displayed on the left, indicating the zinc absorbance peak. On the right, one diffraction pattern of N-LytA is displayed.

Table 13 Data collection and refinement statistic

Data collection	N-LytA
Resolution [Å]	18.71-1.11 (3.32-1.05)
Cell dimensions [Å] *a, b, c*	50.44, 50.44, 72.61
alpha, beta, gamma [°]	90.0, 90.0, 120.0
Space group	P 32
Number of reflections	
observed	253,820
unique	95,581
I/σ	19.2 (1.7)
Completeness [%]	73.9 (100.0)
R_{merge} [%]	3.9 (53.2)
Refinement statistics	
R_{cryst} [%]	20.1
R_{free} [%]	21.1
Number of atoms	
Total	1640
Solvent	240
B-factor [Å^2]	15.34
Ramachandran plot [%]	
Residues in most favoured regions	93.2
Residues in additional allowed regions	6.8
Residues in generously allowed regions	0
Residues in disallowed regions	0

Figures in parentheses correspond to the highest resolution shell

One example of the atomic resolution at 1.0 Å is displayed in figure 31, illustrating the electron density for residues Thr82, Tyr83 and Ala84. Dihedral angles of all amino acid residues were plotted with angles φ versus ψ revealing that most φ/ψ combinations belong to the secondary structures of β-sheets (top-left area in the Ramachandran plot) and right handed α-helices (middle-left). To some extent left handed α-helices (right) but no outliers were found. (Figure 32)
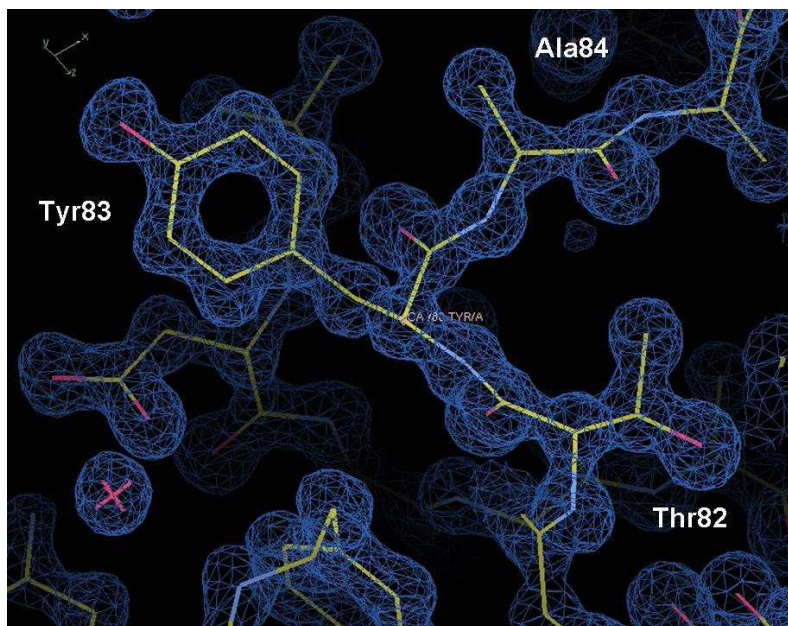
Figure 31 The electron density map with residues Thr82, Tyr83 and Ala84 at 1.0 Å resolution. The figure was prepared with the program Coot. [39]
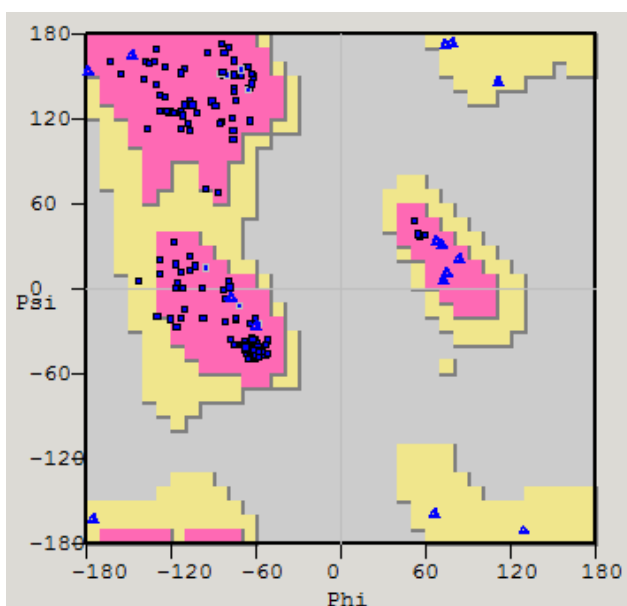
Figure 32 Ramachandran plot analysis demonstrates that 98.8 % of N-LytA sidechains are in a favourable conformation and that 1.2 % are in an allowed conformation. No outliers were found. The program Coot was used to calculate the Ramachandran plot.[39]

4.2.4 Structural and functional analysis of *Streptococcus pneumoniae* major autolysin LytA

4.2.4.1 Homologue sequences to N-LytA

Homologue sequences to N-LytA were found using the BLASTp suite (http://blast.ncbi.nlm.nih.gov/). Sequence alignment was performed using the multiple sequence alignment program ClustalW2 (http://www.ebi.ac.uk/).[30, 31]

All identified homologue proteins are derived from gram positive bacteria or *Streptococcus* phages and belong to the superfamily of peptidoglycan recognition proteins (PGRPs). PGRPs are pattern recognition molecules found in vertebrates and invertebrates belonging to the innate immune system. Peptidoglycan hydrolyzing PGRPs, found in mammals and insects, have a similar three dimensional fold. Similar to type 2 bacteriophage amidases, peripheral α-helices and several central β-sheets form the protein. The front face has a cleft which forms a peptidoglycan groove. At the back of the molecule there is a diverse PGRP-specific segment which is discussed to be a second interaction site for non-peptidoglycan ligands or other effector molecules involved in signalling pathways. The involvement in the enzymatic activation of amidase active PGRPs is also discussed. Zinc is a requirement for peptidoglycan hydrolyzing amidase activity. Therefore inactive PGRPs without zinc can only bind but not cleave peptidoglycans. Almost all PGRPs have two closely spaced conserved cysteine residues in the middle of the PGRP domain that form a disulfide bond, which is needed for the activity of PGRPs. Most vertebrate and some invertebrate PGRPs have two additional conserved cysteines that form a second disulfide bond and many mammalian PGRPs form a third disulfide bond too. Cysteine residues are also involved in the formation of disulfide-linked heterodimers in mammal PGRPs whereas insect PGRPs have not been shown to form disulfide-linked dimers. However, binding to ligands may induce dimerization.

Insect PGRPs activate the Toll or the immune deficiency signal transduction pathway, or induce proteolytic cascades, phagocytosis or hydrolyze peptidoglycans. Mammals have four families of secreted PGRPs (PGRP-S, PGRP-L, PGRP-Iα, PGRP-Iβ). Only PGRP-L is an active amidase, hydrolyzing bacterial peptidoglycans, the other three have bactericidal effects.[77] Structures of homologue PGRP domains to N-LytA are know from *Drosophila melanogaster*

(PGRP-LB and PGRP-SA), and *Homo sapiens* (PGRP-IαC and PGRP-S) but only one of them, PGRP-LB, is an active amidase.

The overall sequence identity varies between 40 and 90 %. Sequence alignment of N-LytA and homologue proteins show a high degree of conserved residues in the peptidoglycan binding groove (Figure 33 and 34). Conserved residues on the PGRP-specific element of N-LytA might be involved in the (unknown) process of activation of LytA or are involved in binding events of ligands (peptidoglycan, teichoic acid, activating proteins).
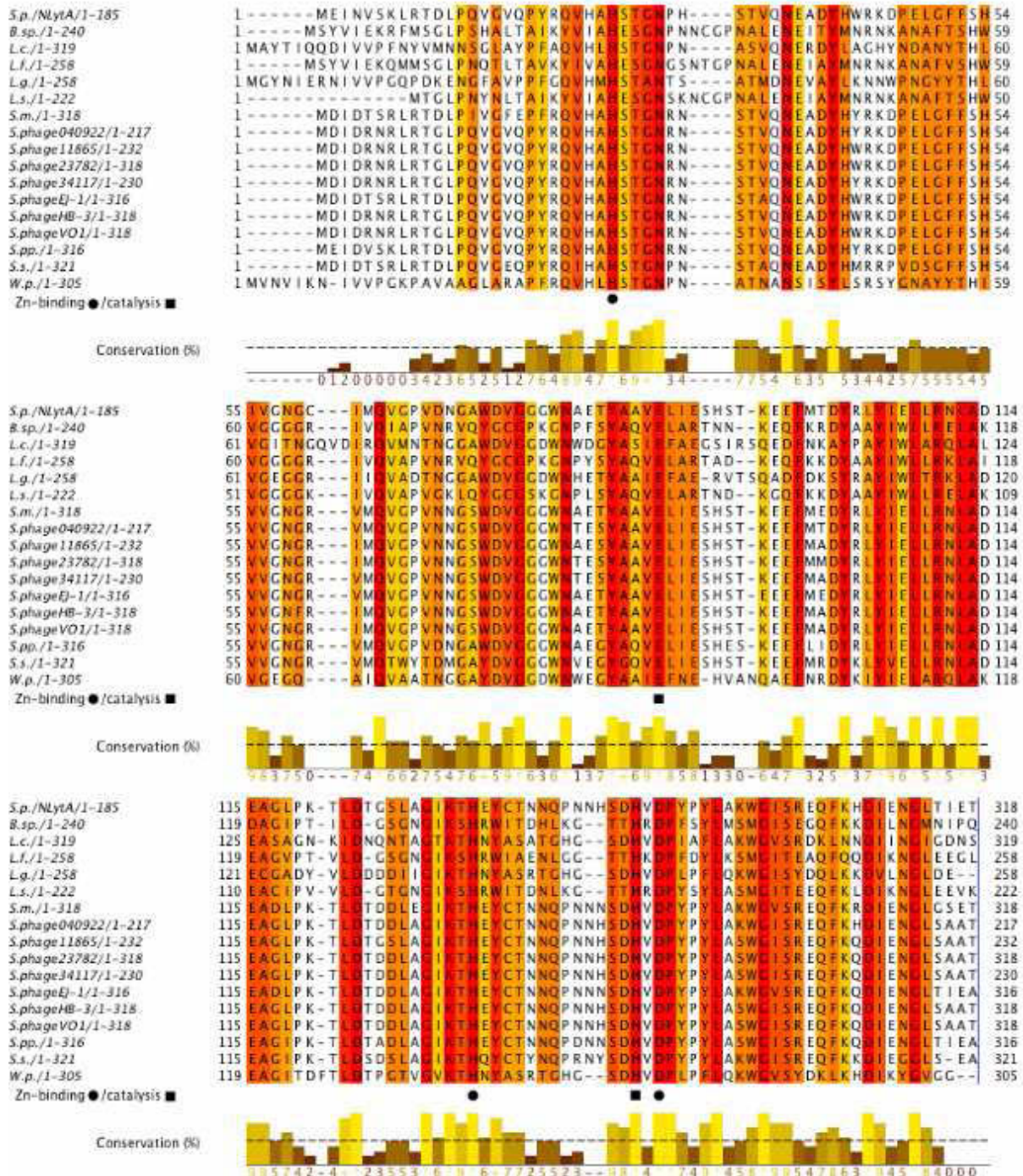
Figure 33 Homologue sequences were aligned using the program ClustalW.[30] The percentage of conservation is visually (50-100%, yellow-red) and numerically (1-10) shown. Residues involved in zinc binding and catalysis are indicated with a black circle and a black rectangle, respectively.
Used abbreviations: **S.p.** *Streptococcus pneumoniae*; **B.sp**. *Bacillus sp. B14905*; **L.c.** *Leuconostoc citroneum KM20*; **L.f.** *Lysinibacillus fusiformis ZC1*; **L.g.** *Leuconostoc gasicomitatum LMG 18811*; **L.s.** *Lysinibacillus sphaericus C3-C41*; **S.m.** *Streptococcus mitis*; **S.phage040922** *Streptococcus phage 040922*; **S.phage11865** *Streptococcus phage 11865*; **S.phage23782** *Streptococcus phage 32782*; **S.phage34117** *Streptococcus phage 34117*; **S.phageEJ-1** *Streptococcus phage EJ-1*; **S.phageVO1** *Streptococcus phage VO1*; **S.pp.** *Streptococcus pseudopneumoniae ATCC BAA-960*; **S.s.** *Streptococcus sanguinis ATCC 49269*; **W.p.** *Weissella paramesenteroides ATCC 33313*. The figure was made using Jalview (http://www.jalview.org/)[33]
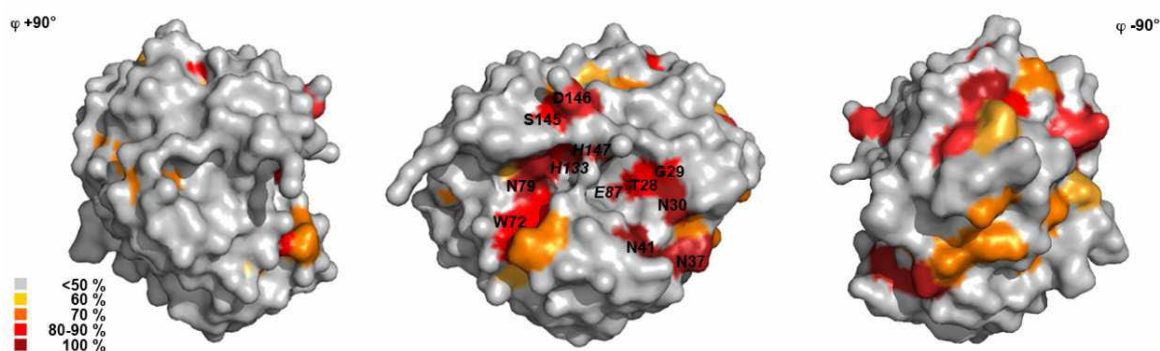
Figure 2 Surface presentation of the peptidoglycan binding groove of N-LytA seen from three different angles. Sequence-conserved residues are colored according to their degree of conservation. The same color-coding is used as for the sequence alignment. Aminoacids which are part of zinc-binding (H133) and catalysis (E87, H147) are shown in italic letters. The figure was created using the program PyMol.[19]

4.2.4.2 The overall fold of the catalytic domain of LytA

The catalytic domain of *Streptococcus pneumoniae* LytA comprises amino acid residues 1 – 180. The final molecular model comprises residues 1 to 172. The last 8 residues and uncleaved His-tag were not included into the model, since the quality of the electron density was too poor for this part of the molecule.

The catalytic domain adopts a typical α/β hydrolase fold with six mixed β-sheets (β3-β2-β4-β1-β6-β5) flanked by seven α-helices. The zinc ion is clearly seen in the electron density map, positioned between the β-sheet and the peptidoglycan binding cleft. (Figure 35) Cysteine residues at positions 60 and 136 were mutated to alanines which did not affect the catalytic activity of N-LytA and LytA. Furthermore, cysteines are not part of the active site and are unlikely to form a disulphide bond between each other, since they are located at opposing sites of the N-LytA fold. (Figure 36) A disulphide bond linked formation of homo- or heterodimers can not be ruled out at the present time. The involvement of one or both cysteine residues in the process of activating the autolysin is another suggestion for their presence. The catalytic domain is linked to the choline binding domain via a short loop. This connection may allow the catalytic domain to be flexible in order to connect to the peptidoglycan strands.
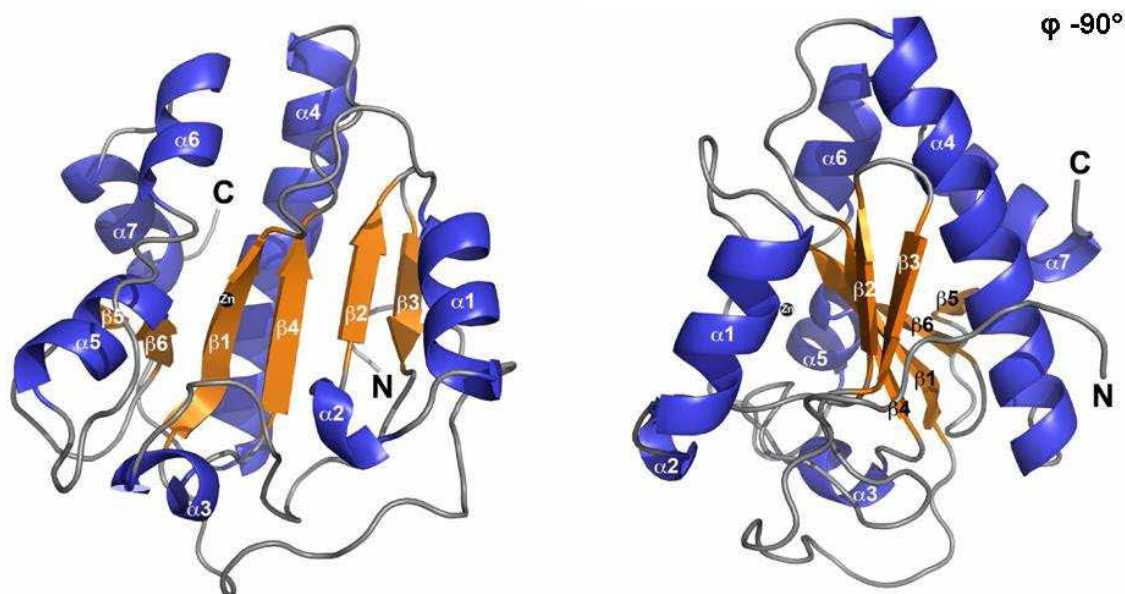
Figure 35 Cartoon representations of the catalytic domain of LytA (residues 1 – 172) seen from two different angles. Helices and β-sheets are colored in blue and orange respectively. The fold of the catalytic site of LytA corresponds to a typical α/β hydrolase with six mixed β-sheets (β3-β2-β4-β1-β6-β5) flanked by seven α-helices. A zinc ion, which is bound in the active site is displayed as a black sphere. The figures were created using the program PyMol.[19]
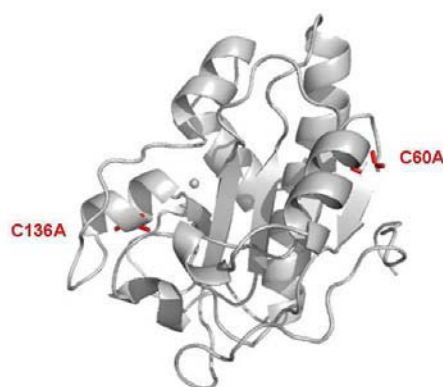
Figure 36 Cartoon representation of N-LytA indicating the positions of C60A and C136A. The figure was made using the program PyMol.[19]

Structural homologues to N-LytA were identified by DALI analysis (http://ekhidna.biocenter.helsinki.fi). The three-dimensional electron density map of N-LytA was thereby used as a template versus all entries of the protein data bank (PDB). Similarity of homologue structures to the template is specified with rmsd-values (root mean square deviation), which is the measurement of the average distances between the atoms (usually the backbone atoms) of superimposed proteins.[34]

DALI results revealed that *Staphylococcus epidermis* bifunctional autolysin AmiE (amidase *Staphylococcus epidermis*) is the most similar molecule to N-LytA (Z-score = 21, rmsd = 1.8 Å). Prophage endolysin PlyL encoded by *Bacillus anthracis* (Z-score = 18.5, rmsd = 2.2 Å) and AmiE require zinc for being a catalytically active protein. Structural homologue proteins without the presence of zinc are peptidoglycan recognition proteins of *Drosophilia melanogaster* PGRP-SD (Z-score = 12.2, rmsd = 2.7 Å) and *Homo sapiens* PGRP-S (Z-score = 12.2, rmsd = 2.7 Å) and PGRP-IαC (Z-score = 12.0, rmsd = 2.8 Å).

Although only a sequence identity of 23 % is given, AmiE showed the highest structural similarity to the catalytic domain of LytA when compared with all available structures in the RCSB Protein Databank. The 561-amino acid long AmiE is the catalytical N-terminal subunit of the major staphylococcal autolysin AtlE (autolysin *Staphylococcus epidermis*). AtlE plays a role in the degradation of bacterial cell wall through its N-terminal based N-acetylmuramoyl-L-alanine amidase and its C-terminal based Mannosyl-glycoprotein endo-β-N-acetylglucosaminidase. A mechanism of catalysis was proposed in the structure of AmiE (pdb: 3LAT) that is very similar to the one of LytA.[35] As in LytA the zinc ion is bound within two histidines residues (H60, H165) and one aspartic acid residue (D179). However, the zinc atom is not coordinated via three water molecules. Similar to the LytA active site, it is flanked by a histidine (H177) and a glutamic acid (E119) residue having an important role during catalysis. Mutations of zinc binding residues H60A and D179A lead to an inactive form of the protein. Mutating H177, which is involved in catalysis, to an alanine, lead to an inactive form or at least to a protein with decelerated reaction kinetics.[35] These findings are in accordance with results obtained for muations of LytA active site residues H26A (no activity), H133A (no activity) and H147A (strongly reduced activity) demonstrating that the correct coordination of the zinc atom is crucial for the activity of cleaving peptidoglycan strands and that residue H147 is actively involved during the process of cleaving. How strong the involvement of residue E87 is, must be investigated in future experiments. (see also figure 40, Proposed mechanism for catalysis within the catalytic domain of the major autolysin from *Streptococcus pneumoniae*)

The λ prophage Ba02 endolysin PlyL (pdb: 1YB0) was also found to have a similar fold compared to N-LytA. The prophage endolysin comprises an N-terminal catalytic domain and a C-terminal cell wall binding domain. The active site of the N-acetylmuramoyl-L-alanine amidase contains a zinc coordination site including two histidine residues (H29, H129) and one cysteine residue (C137). Residues glutamic acid (G90) and lysine (K135) govern catalysis. Interestingly, Low et al. showed that binding of C-terminal cell wall binding domain of PlyL is not a prerequisite for its lytic activity as it is for *Streptococcus pneumoniae* autolysin LytA.[36]

Drosophila melanogaster PGRP-SD (pdb: 2RKQ) has a clearly lower similarity and does not contain zinc in the active site and lacks two of three zinc coordinating residues. Therefore no amidase activity is seen. DAP-type but no lysine-type peptidoglycan can bind to the classical L-shaped binding cleft which is a difference to LytA. PGRP-SD recognises infections of gram positive bacteria and responses over the Toll signalling pathway with the synthesis of antimicrobial peptides.[37]

The crystal structure of the human peptidoglycan recognition protein PGRP-S was solved in 2005 (pdb: 1YCK). Having an overall similar fold to N-LytA, it contains also some specific attributes. N-terminal binding domain does not contain a zinc ion and can only bind but not hydrolyze peptidoglycans. Present cysteine residues form disulfide bonds indicating to be important for structural integrity of the PGRP domain. PGRP-S forms three disulfide bond interactions between cysteine residues 9 and 133, 25 and 70, 46 and 52. An N-terminal specific PGRP segment (residues 1 to 30) is tethered by the first two disulfide bonds. The third disulfide bond (C46 – C52) links the loop between α1 and β3.[38] Human PGRP-IαC (pdb: 2APH) can bind but not hydrolize peptidoglycan. No zinc ion is present. Three disulfide bonds at similar locations as in PGRP-S were found.[43]

Superpositions of structural homologue proteins (AmiE, PlyL, PGRP-SD, PGRP-S and PGRP-IαC) to N-LytA are displayed in figure 37 (Left), indicating the similar overall fold. Superpositions of catalytically active zinc binding sites (N-LytA AmiE, PlyL) are shown in figure 37 (Right) demonstrating the conserved conformations of the zinc binding and the active site residues.
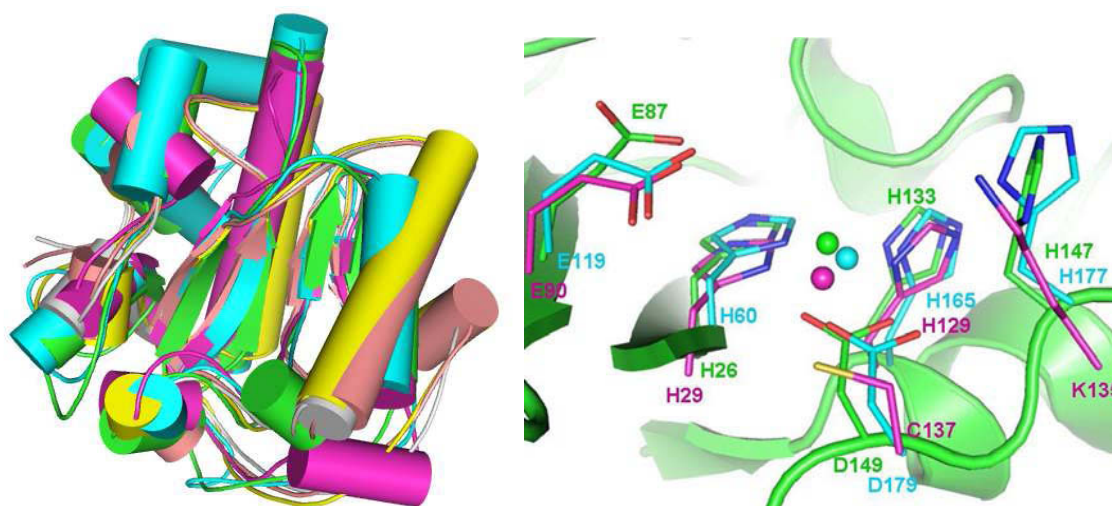
Figure 37 Homologue structures superposed to N-LytA (green). **Left** The most homologue structure found by DALI server analysis[34] is the staphylococcal autolysin AmiE (Z-score = 21, rmsd = 1.8 Å) shown in light blue, followed by purple colored prophage endolysin PlyL (Z-score = 18.5, rmsd = 2.2 Å), yellow colored PGRP-SD (Z-score = 12.2, rmsd = 2.7 Å), *Homo sapiens* PGRP-S (raspberry) (Z-score = 12.2, rmsd = 2.7 Å) and grey colored PGRP-IαC (Z-score = 12.0, rmsd = 2.8 Å). **Right** Zinc binding residues of catalytically active N-LytA, AmiE and PlyL are shown. The same color coding is used as on the left figure. The main chain of N-LytA is shown in cartoon presentation (green). Alignments were done using the program Coot. The figures were prepared using PyMol.[19, 39]

4.2.4.3 The peptidoglycan binding site of LytA

A model of N-LytA with a hypothetical orientation of a bound peptidoglycan fragment is displayed in figure 38. The peptidoglycan fragment is a N-acetomuramyl-L-alanyl-D-iso-glutamyl-L-lysine (MTP) compound (the atomic coordinates of MTP were kindly provided by Dr Sebastian Zoll, University of Tübingen, Germany). Putative interactions between N-LytA and MTP were predicted using the docking program ZDock (http://zdock.bu.edu/)[32] (Figure 38). The nitrogen atom ND2 of residue Asp79 might therefore bind to oxygen atoms of D-Ala and OE1 of D-i-Gln. D-i-Gln possible interacts further between its oxygen atom O and the nitrogen atom N of residue Gly75 and between its nitrogen atom N2 and oxygen atom OG of residue Ser145. Another interaction might occur between the oxygen atom O7 of MurNAc and the nitrogen atom NZ of Lys45.

Aspartic acid at position 79 is conserved among proteins sharing the amidase fold. It is likely that this residue is very important in binding of the peptidoglycan. To verify these proposed interactions, the crystal structure of N-LytA with bound peptidoglycan or peptidoglycan fragment is still required. Future mutations of residues N79 and S145 should also allow to tell more about the importance of these residues.
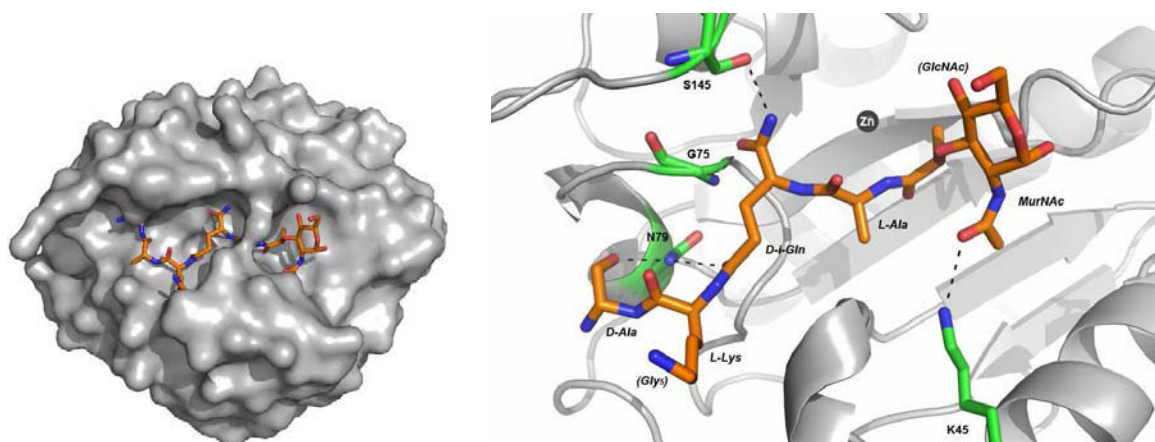
Figure 38 Possible binding mechanisms of bound peptidoglycan fragment MTP (N-acetylmuramyl-L-alanyl-D-iso-glutamyl-L-lysine-D-alanine) within the binding cleft of N-LytA. The figure to the left shows the N-LytA subunit in its surface presentation (grey) and the peptidoglycan fragment MTP as sticks (orange). The figure to the right shows residues N79, G75, S145 and K45 which are likely to be involved in the binding of peptidoglycan. Chemical compounds written in brackets refer to the theoretical positions of GlcNAc and the linking molecules of Gly within the peptidoglycan network. The figures were prepared using PyMol.[19]

4.2.4.4 The active site of the catalytic domain of LytA

The zinc atom present within the active site is bound to six ligands. Nitrogen atoms of two histidine residues (His26 and His133) form hydrogen bond interactions. Furthermore three water molecules are coordinated to the zinc atom. Water molecules are also connected to each other by hydrogen bonds. Additionally every water molecule forms hydrogen bonds to at least one protein atom. The water molecule W141 forms a hydrogen bond with the oxygen atom OD2 of Asp149, the water molecule W89 forms a hydrogen bond with main chain oxygen O73 and water molecule W2 forms a hydrogen bond with the main chain oxygen O27 as well as with the oxygen atom OE2 of Glu87. Residue Glu87 is thought to act as a proton donor during catalysis, residue His147 stabilizes the transition state during catalysis. (Figure 39)
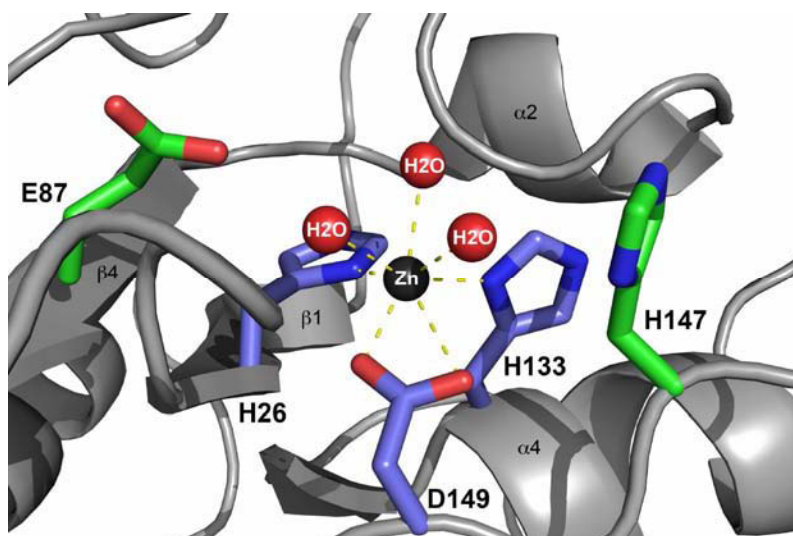
Figure 39 The active site of the catalytic domain of LytA shows a zinc ion (black) bound to side chain residues H26, H133 and D149 (blue) as well as to three water molecules (red). Residues E87 and H147 (green) are thought to be important in catalysing the cleavage of peptidoglycan strands. The figure was prepared using PyMol.[19]

A hypothesis of the hydrolyzation of peptidoglycan strands is graphically shown in figure 40. Upon docking of peptidoglycan or a peptidoglycan fragment, water molecules W89 and W141 are repelled out from the zinc binding cleft (2). One electron pair from the carbonyl oxygen of L-alanine is accepted by the electrophilic zinc. This results in a pentacoordinated zinc ion and a displacement of the water molecule W2 towards residue E87. The positively charged zinc atom and the negatively charged residue E87 result in a nucleophilic attack of the water oxygen to the carbonyl carbon. Also, a proton from the water molecule is transferred to the nitrogen of the peptide bond of L-alanine. The formed transition state has a tetrahedral carbonyl carbon. The negative charged carbon oxygen is stabilized by the protonated sidechain of H147. A second proton from E87 is transferred to the peptide stem which promotes cleavage of L-alanine´s peptide stem (3). After the release of L-alanine´s peptide stem, the N-acetylmuramoyl part is still attached to the zinc ion (4). Three incoming water molecules replace the N-acetylmuramoyl part and reconstitute the initial zinc coordination (1).[35]
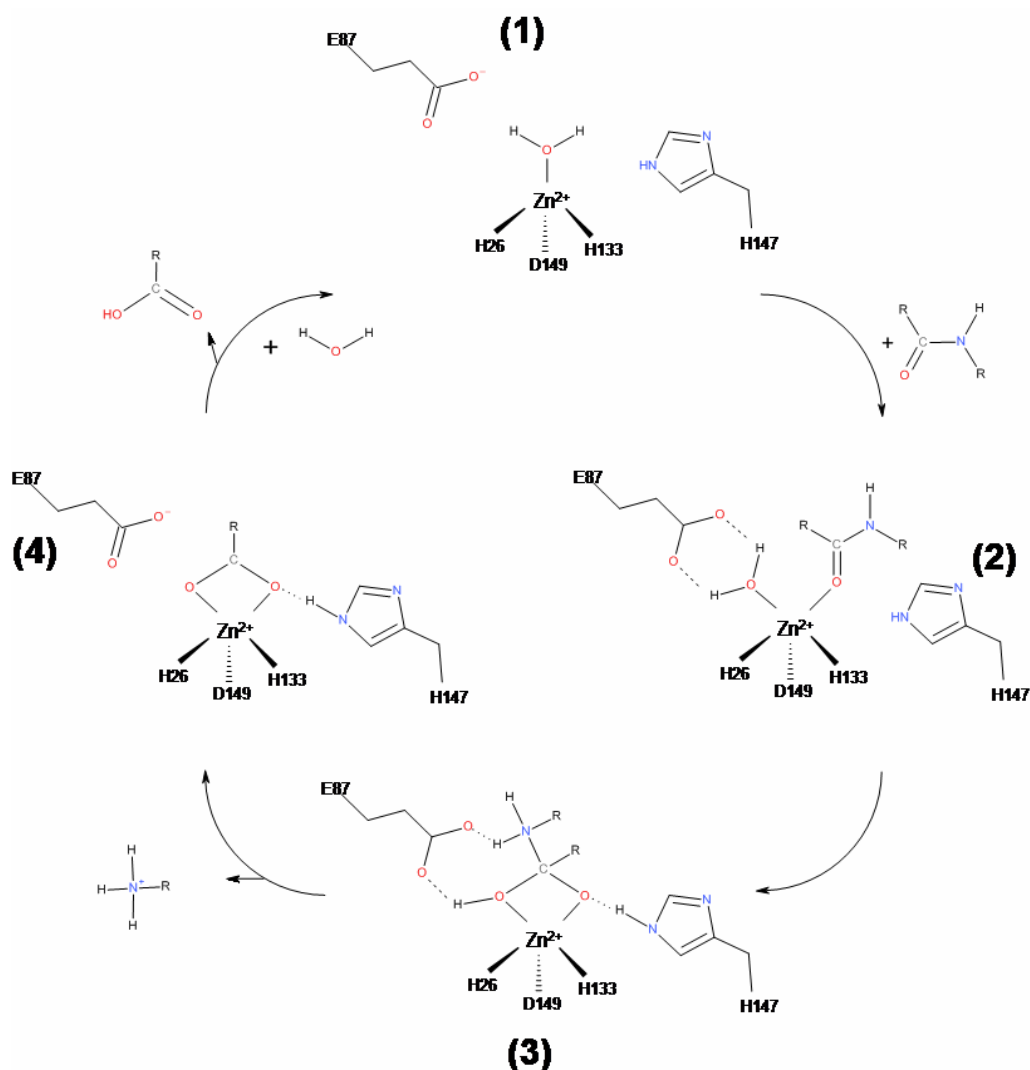
Figure 40 Proposed mechanism for catalysis within the catalytic domain of the major autolysin LytA from *Streptococcus pneumoniae*. The software programme SymyxDraw 4.0 was used to create the figure.

4.2.5 A molecular model of the LytA dimer and its biological function

Fernandez-Tornero *et al.*[41, 42] demonstrated that the formation of a homodimer is necessary for a catalytically active LytA. The hydrophobic dimerization takes place between two hairpins 6 at the C-terminal part of the choline binding domain (pdb: 1HCX). The catalytic domain is located at the N-terminal end of the choline binding domain. (Figure 41) In the displayed model, choline containing molecules are present in the choline binding sites. A peptidoglycan fragment (MTP) is also modelled into the peptidoglycan binding site located at the surface from the catalytic domain, pointing outwards. Since the exact position of the catalytic domain in relation to the choline binding domain is not known, the catalytic domains could also point more to the outside and away from each other as well as they could be in a very close distance. The necessity of the dimer formation

suggests that LytA has a zipper function to break a strand into the peptidoglycan mesh. Teichoic or lipoteichoic acid strands, bound to the choline binding domain, could thereby act as a line where the zipper is moving into one direction.
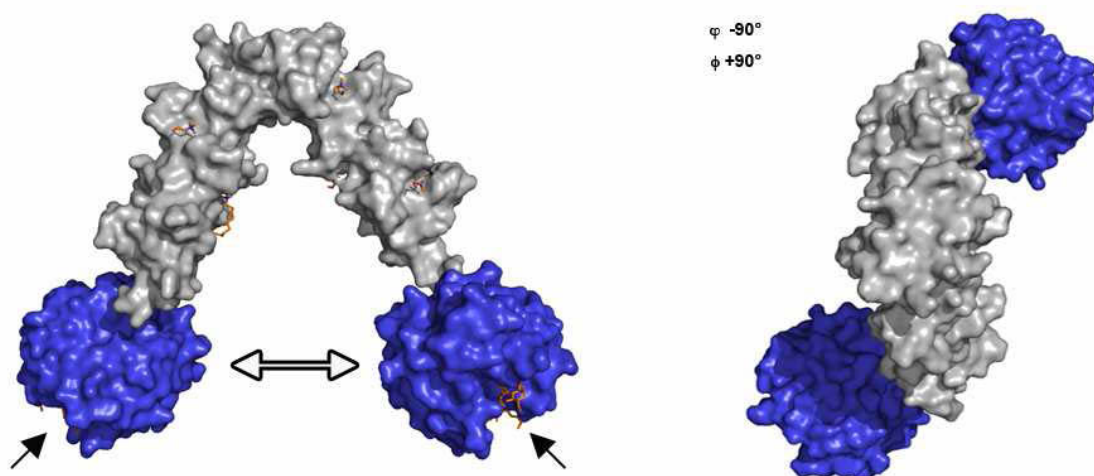
Figure 41 Models of the LytA dimer seen from two different angles show the C-terminal choline binding domain (grey) with bound choline molecules (orange) and the N-terminal catalytic domain (blue) with bound peptidoglycan fragments (orange, indicated with black arrows). The white arrow indicates that the exact position of the catalytic domain in relation to the choline binding domains is not known. The models were created using the programmes Coot and PyMol.[19, 39]

4.2.6 Concluding remarks

With the knowledge of the solved crystal structure of the catalytic domain of *Streptococcus pneumoniae* major autolysin N-LytA, one important step in the process of developing new antibacterial drugs was taken. Still, the process of activation of the autolysin remains elusive and needs to be further investigated. Especially the orientation of the catalytic domain in relation to the choline binding domain is one of the next questions to ask. Knowing these orientations, possible interactions between two catalytic domains or between the choline binding domain and the catalytic domain could be answered. Beside of the question regarding the activation of the enzyme, the detailed mechanisms in destroying the peptidoglycan matrix are waited to be answered. Crystal structures of bound peptidoglycan to the catalytic domain as well as bound teichoic acids to the choline binding domain are needed in order to get more insights into the biological function of LytA.

5 References

1. López R., García J.L., García E., Ronda C. & García P. (1992). Structural analysis and biological significance of the cell wall lytic enzymes of Streptococcus pneumoniae and its bacteriophage. *FEMS Microbiol Lett.* **79**, 439-47.
2. Lewis K. (2000). Programmed death in bacteria. *Microbiol Mol Biol Rev.* **64**, 503-14
3. Jedrzejas M.J. (2001). Pneumococcal virulence factors: structure and function. *Microbiol Mol Biol Rev.* **65**, 187-207
4. Novagen (2005). pET System Manual 11th Edition
5. Sandalova T., Michaelsson J., Harris R. A., Ljunggren H.G., Kärre K., Schneider G. & Achour A. (2005). Expression, refolding and crystallization of murine MHC class I H-2Db in complex with human beta2-microglobulin. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **61**, 1090-3.
6. Achour A., Harris R.A., Persson K., Sundbäck J., Sentman C.L., Schneider G., Lindqvist Y. & Kärre K. (1999). Murine class I major histocompatibility complex H-2Dd: expression, refolding and crystallization. *Acta Crystallogr D Biol Crystallogr.* **55**, 260-2
7. Achour A., Michaelsson J., Harris R.A., Ljunggren H.G., Kärre K., Schneider G. & Sandalova T. (2006). Structural basis of the differential stability and receptor specificity of H-2Db in complex with murine versus human beta2-microglobulin. *J Mol Biol* **356**, 382-96.
8. Garboczi D.N., Hung D.T. & Wiley D.C. (1992). HLA-A2-peptide complexes: refolding and crystallization of molecules expressed in Escherichia coli and complexed with single antigenic peptides. *Proc Natl Acad Sci U.S.A.* **89**, 3429-33
9. Alberts B., Johnson A., Lewis J., Raff M., Roberts K. & Walter P. (2007). Molecular Biology of the Cell Fifth Edition. *Garland Science, Taylor & Francis Group*. New York
10. Reid S.W., Smith K.J., Jakobsen B.K., O'Callaghan C.A., Reyburn H., Harlos K., Stuart D.I., McMichael A.J., Bell J.I. & Jones E.Y. (1996). Production and crystallization of MHC class I B allele single peptide complexes. *FEBS Lett.* **383**, 119-23
11. Gel Filtration, Principles and Methods (2007). Handbook. GE Healthcare
12. Rehm H. & Letzl T. (2002). Der Experimentator: Proteinbiochemie/Proteomics 4. Auflage. *Spektrum Akademischer Verlag*. Heidelberg
13. Laemmli U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685
14. Mini-PROTEAN 3 cell. Instruction Manual. Bio-Rad Laboratories Inc. CA
15. Biacore 3000 Instrument Handbook BR-1003-81 Edition AG (2007). *GE Healthcare*. Sweden
16. Schuck P., Boyd L.F. and Andersen P.S. (2004). Measuring Protein Interactions by Optical Biosensors. *Current Protocols in Cell Biology*, Unit: **17.6**., 1-22
17. Gunnarsson K. (1999). Affinity-based Biosensors for Interaction Analysis. *Current Protocols in Immunology,* Unit: **18.6**., 1-10
18. Cooper M.A. (2002). Optical biosensors in drug discovery. *Nature Reviews Drug Discovery*. 1, 515-528
19. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.
20. Achour, A., Michaelsson, J., Harris, R. A., Odeberg, J., Grufman, P., Sandberg, J. K., Levitsky, V., Karre, K., Sandalova, T. & Schneider, G. (2002). A structural basis for LCMV immune evasion: subversion of H-2D(b) and H-2K(b) presentation of gp33 revealed by comparative crystal structure analyses. *Immunity* **17**, 757-68.
21. Boulter J.M., Schmitz N., Sewell A.K., Godkin A.J., Bachmann M.F., Gallimore A.M. (2007). Potent T cell agonism mediated by a very rapid TCR/pMHC interaction. *Eur J Immunol* **37**, 798-806.
22. Kubo R.T., Born W., Kappler J.W., Marrack P. & Pigeon M. (1989) Characterization of a monoclonal antibody which detects all murine alpha beta T cell receptors. *J Immunol* **142**, 2736-42.
23. Rhodes G. (2006). Crystallography made crystal clear Third Edition. *Academic Press*, USA and UK
24. Research Tools Vol. 18. Hampton research, CA
25. Bergfors T.M. (Ed.) (2009). Protein Crystallization Second Edition. *International University Line*, USA
26. McPherson A. (2004) Introduction to protein crystallization. *Methods.* **34**, 254-65.
27. Perez-Iratxeta C. & Andrade-Navarro M.A. (2007). K2D2: estimate of protein secondary structure from circular dichroism spectra. *BMC Struct Biol*, 8-25.

28. Kelly S.M., Jess T.J. & Price N.C. (2005). How to study proteins by circular dichroism. *Biochim Biophys Acta* **1751**, 119-39

29. Pain R. (2005). Determining the CD spectrum of a protein. *Curr Protoc Protein Sci.,* Chapter 7: Unit 7.6.

30. Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. & Higgins D.G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.

31. McGinnis S. & Madden T.L. (2004) Blast: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* **32**, 20-5

32. Chen R., Li L. & Weng Z. (2003). ZDOCK: an initial-stage protein-docking algorithm. *Proteins* **52**, 80-7

33. Waterhouse A.M., Procter J.B., Martin D.M.A, Clamp M. & Barton, G. J. (2009). Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191

34. Holm L. & Rosenström P. (2010). Dali server: conservation mapping in 3D. *Nucl. Acids Res.* **38**, 545-549.

35. Zoll S., Pätzold B., Schlag M., Götz F., Kalbacher H. & Stehle T. (2010). Structural basis of cell wall cleavage by a staphylococcal autolysin. *PLoS Pathog* **6**, e1000807

36. Low L.Y., Yang C., Perego M., Osterman A. & Liddington R.C. (2005). Structure and lytic activity of a Bacillus anthracis prophage endolysin. *J Biol Chem.* **280**, 35433-9.

37. Leone P., Bischoff V., Kellenberger C., Hetru C., Royet J. & Roussel A. (2008). Crystal structure of Drosophila PGRP-SD suggests binding to DAP-type but not lysine-type peptidoglycan. *Mol Immunol.* **45**, 2521-30.

38. Guan R., Wang Q., Sundberg E.J. & Mariuzza R.A. (2005). Crystal structure of human peptidoglycan recognition protein S (PGRP-S) at 1.70 Å resolution. *J Mol Biol.* **347**, 683-91.

39. Emsley,P. & Cowtan K. (2004). Coot: model-building tools for molecular graphics *Acta Cryst DBiol Crystallogr* **60**, 2126-32.

40. Collaborative Computational Project Number 4. (1994). The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst. D* **50**, 760-763.

41. Fernández-Tornero C., López R., García E., Giménez-Gallego G. & Romero A. (2001). A novel solenoid fold in the cell wall anchoring domain of the pneumococcal virulence factor Lyt*A. Nat Struct Biol.* **8**, 1020-4.

42. Fernández-Tornero C., García E., López R., Giménez-Gallego G. & Romero A. (2002). Two new crystal forms of the choline-binding domain of the major pneumococcal autolysin: insights into the dynamics of the active homodimer. *J Mol Biol.* **321**, 163-73.

43. Guan R., Malchiodi E.L., Wang Q., Schuck P. & Mariuzza R.A. (2004). Crystal structure of the C-terminal peptidoglycan-binding domain of human peptidoglycan recognition protein Ialpha. *J Biol Chem.* **279**, 31873-82.

44. Hain T. (Ed.) (1992). International Tables for Crystallography. *Kluwer Academic Publishers*, Dordrecht-Boston-London

45. Chayen N.E. (1998). Comparative Studies of Protein Crystallization by Vapour-Diffusion and Microbatch Techniques. *Acta Crystallogr D Biol Crystallogr.* **54**, 8-15.

46. Bragg W.L. (1913). The Diffraction of Short Electromagnetic Waves by a Crystal. *Proceedings of the Cambridge Philosophical Society*, **17**, 43–57.

47. Ramachandran G.N., Ramakrishnan C. & Sasisekharan V. (1963). Stereochemistry of polypeptide chain configurations. *J Mol Biol.* **7**, 95-9.

48. O'Brien K.L., Wolfson L.J., Watt J.P., Henkle E., Deloria-Knoll M., McCall N., Lee E., Mulholland K., Levine O.S., Cherian T.; Hib and Pneumococcal Global Burden of Disease Study Team. (2009). Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. *Lancet* **374**, 893-902.

49. PneumoACTION of IVAC at Johns Hopkins Bloomberg School of Public Health. URL: <http://preventpneumo.org>; 25th March 2011

50. López R., García E., García P., García J.L. (1997). The pneumococcal cell wall degrading enzymes: a modular design to create new lysins? *Microb Drug Resist.* **3**, 199-211.

51. Pérez-Dorado I., González A., Morales M., Sanles R., Striker W., Vollmer W., Mobashery S., García J.L., Martínez-Ripoll M., García P. & Hermoso J.A. (2010). Insights into pneumococcal fratricide from the crystal structures of the modular killing factor LytC. *Nat Struct Mol Biol.* **17**, 576-81.

51. Meroueh S.O., Bencze K.Z., Hesek D., Lee M., Fisher J.F., Stemmler T.L. & Mobashery S. (2006). Three-dimensional structure of the bacterial cell wall peptidoglycan. *Proc Natl Acad Sci U S A.* **103**, 4404-9.

52. Seo H.S., Cartee R.T., Pritchard D.G., Nahm M.H. (2008). New Model of Pneumococcal Lipoteichoic Acid Structure Resolves Biochemical, Biosynthetic, and Serologic Inconsistencies of the Current Model. *J Bacteriol.* **190**, 2379-87.

53. Royet J.& Dziarski R. (2007). Peptidoglycan recognition proteins: pleiotropic sensors and effectors of antimicrobial defences. *Nature Reviews Microbiology* **5**, 264–277.

54. Weekly epidemiological record (2007). Pneumococcal conjugate vaccine for childhood immunization. *WHO position paper.* **12**, 93-104

55. Figure adapted from URL: <http://www.oxford-instruments.com>; 25th March 2011

56. Figure adapted from URL: <http://www.ruppweb.org/cd/cdtutorial.htm>; 25th March 2011

57. Figure adapted from URL: <http://www.mscwu.wur.nl/UK/Equipment/Circular+dichroism/>; 25th March 2011

58. Parham P. (2005). The immune system Second Edition. *Garland Science, Taylor & Francis Group*, USA

59. Figure adapted from URL: <http://frank.mtsu.edu/~njsmith/modernii/lec10I.html>; 25th March 2011

60. Figure adapted from URL: < http://jaeger.earthsci.unimelb.edu.au/ >; 25th March 2011

61. Figure adapted from URL: <http://www.britannica.com/EBchecked/media/17859/Bragg-diffraction>; 25th March 2011

62. Achour A. (2001). Major Histocompatibility Complex: Interaction with Peptides. *Encyclopedia of Life Science*. John Wiley & Sons, Ltd

63. Bjorkman P.J., Saper M.A., Samraoui B., Bennett W.S., Strominger J.L. & Wiley D.C. (1987). The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature.* **329**, 512-8.

64. Clements C.S., Dunstone M.A., Macdonald W.A., McCluskey J. & Rossjohn J. (2006). Specificity on a knife-edge: the alpha/beta T cell receptor. *Curr Opin Struct Biol.* **16**, 787-95.

65. Hennecke J., Wiley D.C. (2001). T cell receptor-MHC interactions up close. *Cell.* **104**, 1-4.

66. Rudolph M.G., Luz J.G. and Wilson I.A. (2002). Structural and Thermodynamic Correlates of T Cell Signaling. *Annu. Rev. Biophys. Biomol. Struct.* **31**, 121-49

67. Carreño L.J., González P.A. & Kalergis A.M. (2006). Modulation of T cell function by TCR/pMHC binding kinetics. *Immunobiology* **211**, 47-64.

68. Armstrong K.M., Insaidoo F.K. & Baker B.M. (2008). Thermodynamics of T-cell receptor-peptide/MHC interactions: progress and opportunities. *J Mol Recognit.* **21**, 275-87.

69. López R., Ronda C. & García E. (1990). Autolysins are direct involved in the bactericidal effect caused by penicillin in wild type and in tolerant pneumococci. *FEMS Microbiol Lett.* **54**, 317-22.

70. Moscoso M., Domenech M. & García E. (2010). Vancomycin tolerance in clinical and laboratory Streptococcus pneumoniae isolates depends on reduced enzyme activity of the major LytA autolysin or cooperation between CiaH histidine kinase and capsular polysaccharide. *Mol Microbiol.*

71. Krogsgaard M., Prado N., Adams E. J., He X. L., Chow D. C., Wilson D. B., Garcia K. C. & Davis M. M. (2003). Evidence that structural rearrangements and/or flexibility during TCR binding can contribute to T cell activation. *Mol Cell* **12**, 1367-78.

72. Tissot A.C., Pecorari F. & Plückthun A. (2000). Characterizing the functionality of recombinant T-cell receptors in vitro: a pMHC tetramer based approach. *J Immunol Methods.* **236**, 147-65.

73. Tissot A.C., Ciatto C., Mittl P.R., Grütter M.G. & Plückthun A. (2000). Viral escape at the molecular level explained by quantitative T-cell receptor/peptide/MHC interactions and the crystal structure of a peptide/MHC complex. *J Mol Biol.* **302**, 873-85.

74. Ely L.K., Beddoe T., Clements C.S., Matthews J.M., Purcell A.W., Kjer-Nielsen L., McCluskey J. & Rossjohn J. (2006).Disparate thermodynamics governing T cell receptor-MHC-I interactions implicate extrinsic factors in guiding MHC restriction. *Proc Natl Acad Sci U S A.* **103**, 6641-6.

75. Iversen A.K., Stewart-Jones G., Learn G.H., Christie N., Sylvester-Hviid C., Armitage A.E., Kaul R., Beattie T., Lee J.K., Li Y., Chotiyarnwong P., Dong T., Xu X., Luscher M.A., MacDonald K., Ullum H., Klarlund-Pedersen B., Skinhøj P., Fugger L., Buus S., Mullins J.I., Jones E.Y., van der Merwe P.A. & McMichael A.J. (2006). Conflicting selective forces affect T cell receptor contacts in an immunodominant human immunodeficiency virus epitope. *Nat Immunol.* **7**, 179-89.

76.   Persaud S.P., Donermeyer D.L., Weber K.S., Kranz D.M. & Allen P.M. (2010). High-affinity T cell receptor differentiates cognate peptide-MHC and altered peptide ligands with distinct kinetics and thermodynamics. *Mol Immunol*. **47**,1793-801.

77.   Dziarski R. & Gupta D. (2006). The peptidoglycan recognition proteins (PGRP). *Genome Biology* **7**, 232

78.   Duru A.D. *et al.* (2010) Paper submitted

79.   Allerbring E. *et al*. (2011) Paper submitted

6 Curriculum vitae

Personal Data

Name	Markus Bernhard TOMEK
Date of birth	18th November 1983
Place of birth	Vienna
Nationality	Austrian

University

University of Natural Resources and Applied Life Sciences, Vienna

Master program: Biotechnology
2009/10 - present

Bachelor program: Food Science and Biotechnology
2004/05 – 2008/09

Bachelor Thesis:
Isolation of Lactic Acid Bacteria (LAB) found on fermented fish (Plara) in Thailand

KTH – Royal Institute of Technology, Stockholm
spring 2008

Education

Federal College of Viticulture and Pomology
Wiener Straße 74; 3400 Klosterneuburg
graduated with excellent success; 17th June 2003
1998/99 – 2002/03

Gymnasium GRG19
Billrothstraße 26 – 30; 1190 Vienna
1994/95 – 1997/98

Elementary School
Cottagegasse 17; 1180 Vienna
1990/91 – 1993/94

Professional career

Medical University Vienna
Spitalgasse 23; 1090 Vienna
MTA

Pibulsongram Rajabhat University, Thailand
Muang District, Phitsanuloke 65000
Microbiology traineeship

Umweltbundesamt GmbH
Spittelauer Lände 5, 1090 Vienna
Analytical Chemistry traineeship