



Universiteit
Leiden
The Netherlands

The application of X-ray crystallography and site-directed mutagenesis to the study of protein structures

Thomassen, Ellen Anna Johannes

Citation

Thomassen, E. A. J. (2005, April 28). *The application of X-ray crystallography and site-directed mutagenesis to the study of protein structures*. Retrieved from <https://hdl.handle.net/1887/834>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/834>

Note: To cite this publication please use the final published version (if applicable).

The application of X-ray crystallography and site-directed mutagenesis to the study of protein structures

**The application of X-ray crystallography and site-directed
mutagenesis to the study of protein structures**

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus Dr. D.D. Breimer,
hoogleraar in de faculteit der Wiskunde en
Natuurwetenschappen en die der Geneeskunde,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 28 april 2005
klokke 14.15 uur

door

Ellen Anna Johannes Thomassen

geboren te Meerssen in 1974

Promotiecommissie

Promotores: Prof. Dr. J.P. Abrahams
Prof. Dr. C.J.M. Melief

Co-promotor: Dr. F. Koning

Referent: Prof. Dr. H.P. Spaink

Overige leden: Prof. Dr. J. Brouwer
Dr. M.J. van Raaij (Universiteit Santiago de Compostela, Spanje)
Dr. N.S. Pannu

De totstandkoming van dit proefschrift werd mede mogelijk gemaakt door een bijdrage van het Leids Universiteits Fonds/Van Trigt.

Cover and chapter art: Prof. Dr. Mathieu H.M. Noteborn.

This thesis was printed by Offsetdrukkerij Ridderprint B.V., Ridderkerk.

Kristallen

Mineralen ontmoeten elkaar

Groeien uit tot kristallen

Vormen vertellen ontstaan

Kleuren benadrukken aanwezigheid

Eiwitten in heldere oplossing

Fundamenten van ons leven

Stapelen in stilte tot kristallen

Analyses vinden biologische structuren

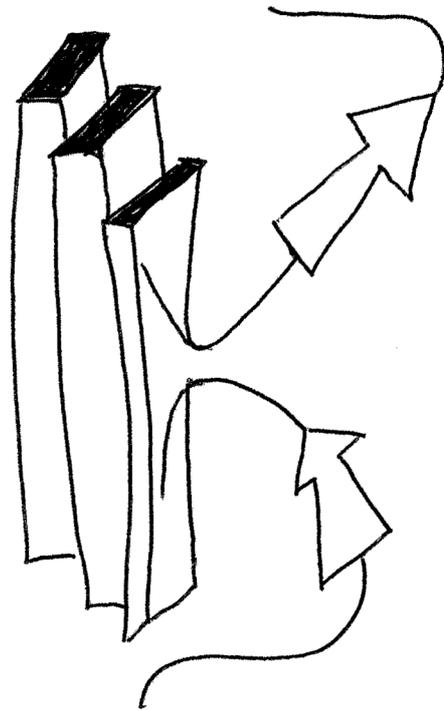
Fundamenten van het leven

Gevonden in fragiele kristallen

Contents

1 Introduction	9
1.1 Outline	11
1.2 Human T cell receptor - CD3 complex	13
1.3 Bacteriophage T4	23
1.4 Scope and aim of this thesis	35
2 Human T Cell Receptor - CD3 complex	43
2.1 The impact of single amino acid substitutions in CD3 γ on the CD3 $\epsilon\gamma$ interaction and T cell receptor - CD3 complex formation	45
2.2 Analysis of intracellular CD3 δ and CD3 ϵ synthetic peptides	65
3 Bacteriophage T4	69
3.1 The structure of the receptor-binding domain of the bacteriophage T4 short tail fibre reveals a knitted trimeric metal-binding fold	71
3.2 Crystallisation and preliminary crystallographic studies of a new crystal form of the bacteriophage T4 short tail fibre	97
4 Potato Serine Protease Inhibitor	105
4.1 Crystallisation and preliminary X-ray crystallographic studies on a serine protease inhibitor from the Kunitz family	107
5 Recombinant Human Lactoferrin	113
5.1 The protein structures of recombinant human lactoferrin produced in the milk of transgenic cows and human milk-derived lactoferrin are identical	115
6 Summary and Conclusions	127
Nederlandse samenvatting	135
List of abbreviations	138
List of publications	140
Curriculum Vitae	141
Nawoord	143
Appendix A colour pictures	145

Introduction



1.1

Outline

In 1912 Laue [1] was the first to suggest the use of a crystal to act as a lattice for the diffraction of X-rays. He showed that if a beam of X-rays passed through a crystal, diffraction would take place and a pattern would be formed on a photographic plate placed at a right angle to the direction of the rays. The pattern would define the symmetrical arrangements of the atoms in the crystal. Nowadays, X-ray crystallography has become a general tool for the determination of the three-dimensional structure of proteins.

In this thesis, X-ray crystallography was used to determine the three-dimensional structure of several proteins. **Chapter 1** is an introduction to two of the proteins investigated in this thesis (human T cell receptor - CD3 complex and bacteriophage T4).

The impact of single amino acid substitutions in CD3 γ on the CD3 $\epsilon\gamma$ interaction and T cell receptor - CD3 complex formation is described in **chapter 2.1**. The results indicate that several amino acids in CD3 γ are essential for an optimal association between CD3 γ and CD3 ϵ and the assembly of a cell surface expressed TCR- $\delta\epsilon\gamma\epsilon\zeta_2$ complex. In order to determine the three-dimensional structure of the intracellular domains of CD3 δ and CD3 ϵ , synthetic peptides corresponding to these domains were synthesised and used for crystallisation experiments and analysed by NMR. The results are described in **chapter 2.2**.

The crystal structure of the receptor-binding domain of the bacteriophage T4 short tail fibre is described in detail in **chapter 3.1**. The trimeric protein has a novel knitted trimeric metal-binding fold and contains the receptor-binding domain. This receptor-binding domain recognises and binds irreversibly to the core region of the host cell LPS. We propose where the LPS-binding region is located. In **chapter 3.2**, preliminary results are described obtained from a new crystal form of the short tail fibre. In these crystals the electron density for the short tail fibre is more ordered compared to the structure determined in chapter 3.1.

Chapter 4 describes the crystallographic studies of a dimeric double-headed potato serine protease inhibitor of the Kunitz family. These are the first crystallisation conditions

described for a dimeric double-headed inhibitor. To date, only monomeric single-headed or monomeric double-headed inhibitors have been crystallised.

The protein structure of recombinant human lactoferrin produced in the milk of transgenic cows is identical to that of natural human lactoferrin, despite a differential N-linked glycosylation. These results confirm the validity of transgenic cows to produce recombinant human proteins and are described in **chapter 5**.

This thesis is finalised with a summary and conclusions, which are described in **chapter 6**.

1.2

Human T cell receptor - CD3 complex

The T cell receptor (TCR) is present on T cells, which are a subset of lymphocytes defined by their development in the thymus and by the expression of the T cell receptor. The function of the T cell receptor is to sense the presence of intracellular pathogens which is a crucial step towards the initiation of a specific immune response aimed at the eradication of such pathogens. TCRs are heterodimeric receptors that are cell surface expressed in association with the proteins of the CD3 complex [2]. There are various subsets of T cells. The two major subsets are the CD8⁺ and CD4⁺ T cells. The cytotoxic CD8 cells can kill infected target cells thereby preventing replication of intracellular pathogens. The helper CD4 cells are crucial for the initiation and regulation of immune responses as well as providing help for B cells which results in antibody production. After the T cells have developed in the thymus they go into the bloodstream, from where they migrate through the peripheral lymphoid organs *e.g.* the lymph nodes, spleen, and mucosal-associated lymphoid tissue where immune responses are induced. Subsequently, they re-enter into the blood stream, where they circulate until they encounter antigens [3].

The immune system

The human immune system has evolved to protect us from invading pathogens. The immune system can roughly be divided in two parts: the innate or natural immune system and the adaptive or specific immune system. The innate immune system is inherited by birth and is non-specific. Important components of the innate immune system are neutrophils and the complement system. Neutrophils are short-lived leucocytes that penetrate inflamed tissues and destroy pathogens upon phagocytosis. The complement system tags invading pathogens, facilitating receptor-mediated uptake by phagocytes, and also works through soluble factors produced by the adaptive immune system, the immunoglobulins. The adaptive immune system response develops and is continuously reshaped during the lifetime of an individual. The main components of the adaptive immune system are the B and T cells, lymphocytes with clonotypic cell-surface receptors

for the recognition of the foreign antigens. While the B lymphocytes produce antibodies for antigen removal from the circulation, the T lymphocytes track down and eliminate infected cells. These lymphocytes provide the life-long immunity that develops after exposure to a disease or vaccination [2].

T and B lymphocytes development

The central or primary lymphoid organs are the bone marrow and the thymus. Both B and T cells originate in the bone marrow but only the B cells mature here. The T cells migrate to the thymus for maturation. After maturation both lymphocytes enter the blood stream. From here they migrate into the peripheral lymphoid organs. These peripheral lymphoid organs are specialised to trap antigen and allow the initiation of adaptive immune responses.

The T cell precursors entering the thymus express neither the T cell receptor nor one of the two co-receptor molecules CD4 or CD8, and are called double-negative cells. During proliferation these immature thymocytes differentiate into double-positive cells, with low T cell receptor levels and both co-receptor molecules. The T cells developing in the thymus randomly rearrange their T cell receptor encoding genes creating a vast array of TCRs. However, since it is random, it generates T cells that are potentially harmful for host tissues when their TCRs recognise combinations of MHC and self-peptides. Rearrangement can also produce cells that are useless because their TCRs interact too poorly with MHC molecules. To allow the survival of only the useful T cells, a two step process called positive and negative selection is performed. During positive selection TCRs are selected that have an affinity for MHC-peptide complexes expressed in the thymus. Subsequently, through negative selection those thymocytes expressing a TCR with a too high affinity for MHC-peptide are eliminated through the induction of apoptosis. The result is a TCR repertoire with an intermediate affinity for self MHC-self peptide which can react with self MHC-foreign peptide in the periphery. Next to rearrangement and selection, the cells differentiate in one of the two lineages, the CD4 or CD8 single-positive cells. As each T cell expresses a distinct TCR, a vast array of receptors is available for recognition of foreign peptides bound to self MHC. Because of this antigen-binding specificity, the fraction of T cells that can respond to a particular antigen is very small. After an infection many specific effector lymphocytes are required, and hence the activated lymphocytes are induced to proliferate before they differentiate

into effector cells. This major feature of the adaptive immune response is called clonal expansion [2].

The development of B cells follows a specific order of stages [2,4]. The earliest B cell precursor is the early pro-B cell, which has no B cell antigen receptor expression on the cell surface. These cells further develop into late pro-B cells, pre-B cells, immature B cells and mature B cells. In these stages the immunoglobulin gene recombination occurs. Many thousands of rearrangements are possible for both heavy- and light-chain genes. Thus, rearrangement during the B cell development is continuously providing immature B cells with a highly diversified repertoire of surface immunoglobulin molecules, all acting as specific receptors for different antigens. All the processes up to the development of the immature B cells take place in the bone marrow and are independent of antigen. Immature B cells only express surface IgM and (like T cells) they undergo a selection. The immature B cells are subjected to selection for self-tolerance. The B cells that recognise self-molecules while still immature are prevented from further development. The B cells that survive this selection bear a B cell receptor (BCR) repertoire tolerant of the self-molecules and become mature. They are called "naive" until they encounter their specific antigen. During this maturation the cells migrate into the secondary lymphoid organs and the cells are induced to express IgD on their cell surface (next to IgM) through alternative splicing of heavy-chain transcripts [2].

B cell mediated immunity or humoral immune response

B cells prevent the spread of intracellular infections by secreting antibodies. The activation of naive B cells and their differentiation into antibody-secreting cells is triggered by antigen and helper T cells. The B cells first internalise the antigen and then process and degrade it. The peptides resulting from these antigen proteins are presented to the T cells via MHC class II molecules. The helper T cells recognise the peptides derived from the antigen and act through the binding of its CD40L (CD40 ligand) to CD40 on the B cell and by the directed release of cytokines. This results in an activated B cell which produces antibodies against the specific antigen. There are several ways in which these antibodies contribute to immunity. Antibodies can bind to a bacterial antigen, which can prevent the growth, and diminish bacterial adherence. This reduces spread from cell to cell, thus neutralising the bacterial invasion. Binding of the antibody to an antigen in order

to enhance phagocytosis (the internalisation of particular matter by cells *e.g.* macrophages) is called opsonisation. Thirdly, when antibodies bind to the surface of the pathogen, this can activate the complement system, enhancing opsonisation [2,3].

T cell mediated immunity

T cells identify cells that harbour pathogens or have internalised pathogens or their products. They do so with their T cell receptor-CD3 complex. The peptide fragments of pathogen derived proteins are recognised in the form of complexes of peptide and Major Histocompatibility Complex (MHC) molecules on the cell surface of the antigen presenting cell (APC). There are two types of MHC molecule, called MHC class I and MHC class II, which are closely related in structure and function, but they have a different subunit structure. Also, they differ in the source of the peptides they bind and express on the cell surface. The folding motif of the two types of MHC-classes has been revealed by X-ray crystallography and they were found to be very similar [5-8].

MHC class I

MHC class I molecules entrap peptides derived from proteins synthesised in the cytosol. MHC class I has a transmembrane heavy chain (Hc or α -chain, 45 kDa) which is non-covalently associated with the β_2 microglobulin (β_2 M, 12kDa). The α -chain consists of three immunoglobulin-like domains, the peptide binding groove is formed by the two membrane distal domains which form two α -helical segments on top of a series of β -sheets. The resulting groove can accommodate peptides that are usually 8-10 amino acids in length. Binding of the peptides is mediated by interactions between side chains of amino acids in the peptide (anchor residues) with pockets in the MHC molecule. Moreover, binding is stabilised at the free carboxy and amino end by hydrogen bonds. MHC class I captured peptides are recognised by cytotoxic T cells (CD8), which kill the infected cell.

MHC class II

MHC class II molecules bind peptides derived from proteins in intracellular membrane-bound vesicles or internalised antigen in B cells or phagocytic cells. MHC class II molecules consist of a 32 kDa α -chain and 29 kDa β -chain which both cross the

membrane. The peptide binding groove is very similar to that of class I molecules but is open at both ends. As a result, the length of the peptides bound to the class II MHC complex is not as constrained as that of class I peptides. Peptides that can bind are at least 9 amino acid residues long, but most are longer. The peptide lies in an extended conformation along the MHC class II peptide-binding groove which allows multiple interactions that contribute to binding. Moreover, the peptide's side-chains protrude into shallow and deep pockets present in the groove of the MHC class II molecules.

MHC class II molecules captured peptides are recognised by T_H1 or T_H2 type, $CD4^+$ T helper cells. T_H1 cells activate macrophages leading to destruction of the intracellular bacteria, whereas T_H2 activates B cells which subsequently proliferate and differentiate into an antibody-producing plasma cell [2,3].

The T cells recognise the peptide-MHC complex via their T cell receptor CD3 complex, of which the chains have a Fab-like structure that is similar to the B cell receptor. Contrary to the BCR, which gets excreted after antigen stimulation, the TCR stays membrane associated for recognition of peptide-MHC complexes on antigen-presenting cells.

The $\alpha\beta$ TCR-CD3 complex

There are two types of T cell receptor, the $\alpha\beta$ TCR and the $\gamma\delta$ TCR. Here we focus on the $\alpha\beta$ TCR. The $\alpha\beta$ TCR interacts with peptide antigens bound to MHC. The TCR (Figure 1.1) consists of a disulphide-linked hetero-dimer ($\alpha\beta$) that is expressed on the cell surface in association with the CD3 complex [9].

The glycosylated TCR $\alpha\beta$, which has a Fab-like structure, is responsible for the recognition of a specific antigen bound to MHC-molecules. The associated CD3 complex consists of one CD3 γ -chain, one CD3 δ -chain, two CD3 ϵ -chains, and a ζ -dimer ($\alpha\beta\gamma\epsilon\delta\epsilon\zeta_2$). Subsequently, the CD3 components, which are in close proximity and all have immunoreceptor tyrosine-based activation motifs (ITAM) in their intracellular domain, through a still not completely understood process transduce and activate the intracellular signalling pathways [10-13]. The CD3 δ , ϵ , and γ all have a large extracellular immunoglobulin (Ig)-like domain, a membrane proximal stalk region, a transmembrane helix, and the intracellular ITAM containing domain. This in contrast to ζ , which has a small extracellular domain and a large intracellular domain with three ITAMs. The CD3

components are not only required for transduction of the signal across the cell membrane, but also for the expression of the TCR heterodimers on the T cells. If one of the CD3 chains is absent, *e.g.* due to a genetic mutation, the number of T cell receptors present on the cell surface is reduced [14-16].

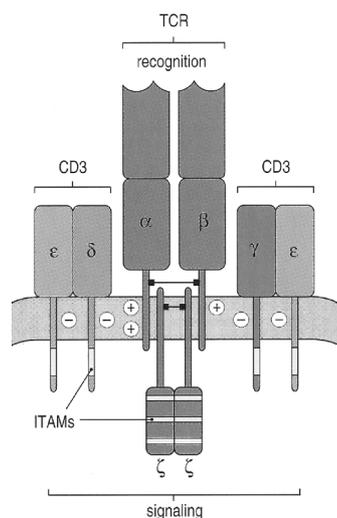
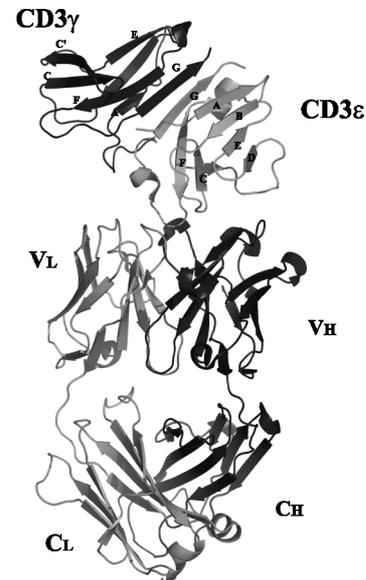


Figure 1.1. The T cell receptor is made up of antigen-recognition proteins and invariant signalling proteins. Picture adapted from [2].

Stoichiometry and assembly of TCR-CD3 complex

Several crystal structures have been elucidated for parts of the $\alpha\beta$ TCR-CD3 complex [17-20]. Garcia *et al.* [21] determined the structure of the complete extracellular fragment of a glycosylated $\alpha\beta$ TCR at 2.5 Å resolution and its orientation bound to a class I MHC-peptide complex. Garboczi *et al.* determined the structure of the complex between the human T cell receptor, viral peptide, and HLA-A2 [22]. Sun *et al.* [23] discovered the structure of an ectodomain fragment of the murine CD3 $\epsilon\gamma$ heterodimer. Recently Kjer-Nielsen *et al.* elucidated the crystal-structure of the human T cell receptor CD3 $\epsilon\gamma$ heterodimer complexed to the therapeutic mAb OKT3 [24]. They show that the binding interface between CD3 ϵ and CD3 γ is mainly formed by hydrogen bonds, salt-bridges and hydrophobic interactions. A side-to-side hydrophobic interface between the two Ig-like domains and parallel pairing of their respective C-terminal β -strands were revealed (Figure 1.2).

Figure 1.2. Structure of the OKT3 Fab/CD3 complex. Ribbon representation showing the heavy and light chain of OKT3 Fab fragment complexed to CD3. The CD3 ϵ monomer contains eight β -strands, whereas the CD3 γ subunit contains seven β -strands. Picture adapted from [24].



Due to better techniques in crystallography and nuclear magnetic resonance (NMR), more and more structures of the TCR-CD3 complex are being determined [25-31], but the precise stoichiometry of the TCR-CD3 complex still is the subject of major discussions in the field. In particular, there are still discussions going on whether α -chain and β -chain are present as one or multiple heterodimeric pairs. Punt *et al.* [32] and de la Hera [33] found that each TCR-CD3 complex contains one α TCR, one β TCR and two CD3 ϵ chains. This in contrast to San José [34], Fernández-Miquel [35], and Exley [36] who all found evidence for a double TCR heterodimer model *e.g.* $\alpha\beta\gamma\epsilon\epsilon\delta\zeta\zeta\alpha\beta$. There is even evidence for yet another stoichiometry given by Thibault and Bardos [37] who suggest the association of two TCR heterodimers with three CD3 ϵ chains in the TCR-CD3 complex. In one of the latest developments in determining the stoichiometry Call *et al.* [38] suggest that the $\alpha\beta$ TCR-CD3 complex assembled in the endoplasmic reticulum (ER) is monovalent and composed of one copy of the $\alpha\beta$ TCR, CD3 $\delta\epsilon$, CD3 $\gamma\epsilon$ and $\zeta\zeta$ -dimers.

The assembly of the TCR-CD3 complex follows discrete steps, in which the transmembrane (TM)-region of CD3 $\delta\epsilon$, CD3 $\gamma\epsilon$ and $\zeta\zeta$ -dimers play an important role. The TM domains of the receptor components have a total of nine basic/acidic residues [39]

(Figure 1.3). Three basic residues are located in the TCR TM-region, whereas each of the three signalling dimers has a pair of acidic residues in their transmembrane domain.

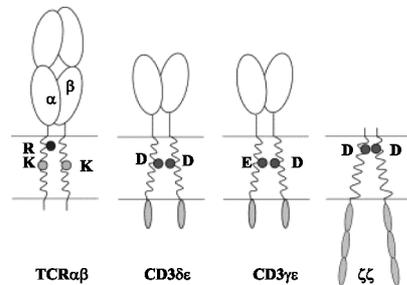


Figure 1.3. Components of the TCR-CD3. Picture from [39]. Basic amino acids are arginine (R) and lysine (K) and acidic amino acids are glutamic acid (E) and aspartic acid (D).

If TCR complexes are multivalent the apparent charge imbalance problem in the TCR-CD3 complex could be overcome. If two TCR heterodimers would be present in the complex, the number of basic residues is exactly the same as the number of acidic residues in the TM region of the CD3 chains. Several different groups examined the assembly of the TCR-CD3 complex. Brenner and colleagues [40] demonstrated the interaction between β TCR and CD3 γ by using crosslinking techniques and Geisler *et al.* showed that the $\alpha\beta$ TCR-CD3 $\delta\epsilon$ intermediate was formed in jurkat cell line deficient in expression of the CD3 γ chain [41]. Call and Wucherpfennig [39,42] proposed a molecular mechanism for the assembly of the TCR-CD3 complex. A schematic overview of the proposed assembly mechanism is shown in Figure 1.4. They showed that each basic TCR residue in the helical TM spanning domain is required for assembly with a particular signalling dimer, thereby forming a three-helix. The two lysine residues (K) are located in the middle of the TM region of $\alpha\beta$ TCR and serve as critical contact points for assembly with CD3 $\delta\epsilon$ and CD3 $\epsilon\gamma$ dimers, respectively. The $\zeta\zeta$ -dimer associates with the arginine (R) residue present in the α TCR.

There is proof that the association of CD3 $\epsilon\gamma$ with the TCR is more efficient in the presence of CD3 $\delta\epsilon$, which indicates that the kinetically preferred order in assembly of $\alpha\beta$ TCR is with CD3 $\delta\epsilon$ (1) followed by CD3 $\epsilon\gamma$ (2) and the final association is the $\zeta\zeta$ -dimer (3). However, in this proposed assembly mechanism there is a charge imbalance in the transmembrane region as it would lead to six acidic residues and only three basic ones, as written above. Call *et al.* [39] give a possible explanation why this may not be a problem.

They propose that each assembly event leads to the creation of a three-helix interface at which the basic and two acidic residues are shielded from the surrounding lipid. Partial or complete protonation of the acidic residues reduce the average net charge for the aspartic acid pair, and hence this rearrangement does not necessarily lead to a charge imbalance.

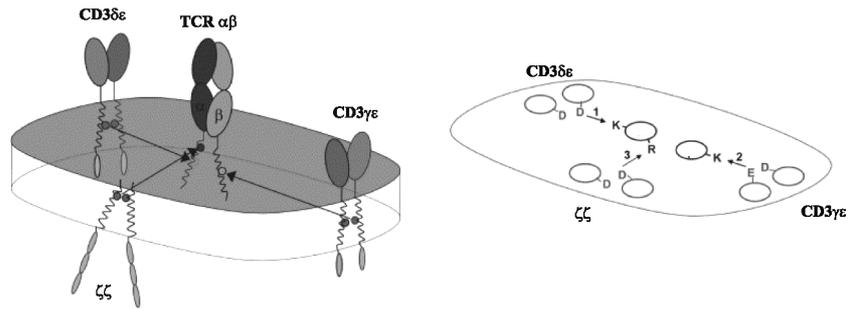


Figure 1.4. Organisation of the TCR-CD3 assembly. The important amino acid residues in each chain are given in one-letter code. D = aspartic acid, E = glutamic acid, K = lysine, and R = arginine. The assembly steps are numbered according to the order in which they occur. Figure adapted from [39].

T cell receptor signalling

As the $\alpha\beta$ TCR chains have very short intracellular domains, they lack domains which could be responsible for signalling. Instead the transmembrane CD3 domains, which are in close proximity to the TCR chains, transduce the signal from the extracellular environment into the cell [13]. The CD3 δ , ϵ , γ , and ζ -chain all have the intracellular sequence YXXI/L(4-6X)YXXI/L, termed the immuno tyrosine-based activation motif (ITAM). These ITAMs play a vital role in signalling from the CD3 complex further into the cell [43,44]. Several different models have been proposed during the last years, all describing different models of intracellular signalling events in T cells. Boniface *et al.* [45] describe that the initiation of signal transduction through the T cell receptor requires the multivalent engagement of peptide/MHC ligands for effective activation. In 2003, Alarcón *et al.* [46] proposed that changes in the interaction between CD3 subunits within the CD3 dimers and the interaction of these dimers with the TCR heterodimer could be the triggering mechanism that initiates the first activation events.

Upon TCR triggering activation of various receptor-associated protein tyrosine kinases (PTKs) of the Src family takes place, such as Fyn or Lck, upon recruitment of the

CD4 or CD8 co-receptor to the TCR-CD3 complex. By recruitment of the co-receptor to the TCR MHC-ligand complex, Lck or Fyn is brought into close proximity of the TCR-CD3 complex. Lck and Fyn have kinase activity and are associated with the cytoplasmic domain of the co-receptor CD4 and CD8. This recruitment results in phosphorylation of the CD3 ITAMs [13]. After phosphorylation of the CD3-ITAMs, ZAP-70 (an ζ -associated protein having an SH2 domain) binds to the phosphorylated ITAM of CD3 ζ . Then Lck or Fyn activates ZAP70 by phosphorylation. After activation, ZAP70 phosphorylates components of several downstream signalling pathways [2,47].

A recent review by Abraham and Weiss [48] depicts our current understanding of T cell receptor signalling that has evolved over the last years. Although thorough research has led to a better understanding of T cell signalling, structural information on the components of the signalling cascade is still largely missing. With protein production, crystallographic and NMR techniques on their current levels of sophistication, this problem may be overcome in the near future.

1.3

Bacteriophage T4

The word "phage" comes from the Greek *phagein*, meaning "to eat" and the word "bakterion" meaning "*small staff*" in Greek. A bacteriophage is a virus that infects bacteria. Bacteriophage T4 infects *Escherichia coli* (*E. coli*), an organism well known to most molecular biologists. It is one of the most complex viruses with a genome that contains 274 open reading frames out of which more than 40 encode structural proteins [49]. The phages multiply inside the bacteria by using the host's biosynthetic machinery; phages always need a host cell, as they are not capable of living without one. T4 bacteriophage is a very efficient DNA injection machine; generally a single T4 phage particle is enough to infect a host cell [50]. The virus consists of a double stranded DNA-containing head, a double-tubed tail of which the outer tail-sheath is contractile, and a baseplate to which six long tail fibres and six short tail fibres are attached (Figure 1.5). With these tail fibres the phage attaches to its host cell, after which it penetrates the cell membrane and subsequently releases its DNA into the host. Like several other phages, T4 DNA contains a modified base, which protects the DNA from the restriction system of the infected host cell.

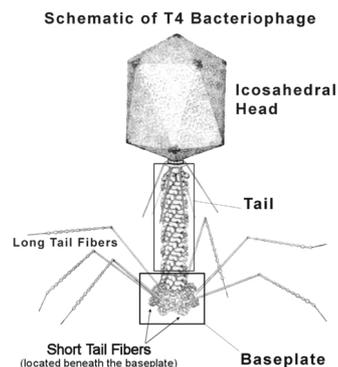


Figure 1.5. Schematic structure of bacteriophage T4. Picture from <http://www.nsf.gov/od/lpa/news/02/pr0207.htm>.

Lately, more structural details of the bacteriophage T4 became available, due to better techniques in electron microscopy and X-ray crystallography.

Bacteriophage T4 features

Bacteriophage T4 has a plating efficiency that approaches one, meaning that virtually every phage particle plated on a lawn of susceptible bacteria is capable of forming a plaque. As well as very efficient, the absorption of the phage to the host cell is very rapid. Bacteriophage T4 uses several different mechanisms to arrest the synthesis of nucleic acids and proteins of the infected host cell. Another feature is the burst size: 100-500 phage particles are produced per infected cell within 15 to 30 minutes at 37 °C [51].

The T4 life cycle

Bacteriophage T4 replicates by the so-called lytic cycle [51]. A cycle is called lytic, when new viruses are produced within the infected bacterium and the viruses lyse the infected host bacterium in order to be released. The replication cycle exists of several stages, starting with *adsorption*; the bacteriophage attaches to the receptors in the bacterial cell wall of the host cell. Next comes the *penetration*; the phages make holes in the host cell, through which they can inject their genome. Most of the phages do this by contraction of the outer tail sheath, which drives the hollow inner tail tube into the host cell. Penetration is followed by *replication*, in which T4 proteins partially shut down the macromolecular machinery of the host cell and direct the replication of the bacteriophage genome and structural components. For this replication, T4 uses the metabolic machinery of the host cell to synthesise phage enzymes and structural components. *Maturation* consists of the phage parts assembling around the genomes. Finally, during the *reinfection* step, T4 lysozyme lyses the bacterial peptidoglycan layer causing osmotic lysis, implicating the release of new T4 bacteriophages [52].

Bacteriophage T4 DNA

The bacteriophage contains 172 kb of linear double stranded DNA, made up of 5-hydroxymethyl dCMP (Hm-dCMP) building blocks instead of cytosine, as well as the normal dAMP, dGMP and dTMP nucleotides. Apart from this modification, T4 also modifies the Hm-dCMP residues by glycosylating them, after the precursor has been incorporated. Glucose is covalently bound to the T4 Hm-dCMP residues in 70 % of the cases in the α -configuration and the remaining 30 % in the β -configuration (Figure 1.6).

These modifications provide the T4 DNA with an efficient protection system against the host restriction mechanism [51].

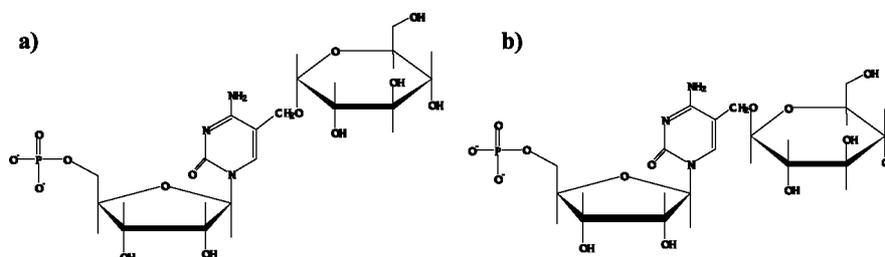


Figure 1.6. Structures of glucosylated Hm-dCMP residues in T4 DNA a) α -D-glucosyl-5-hydroxymethyl-deoxycytidine 5'-monophosphate b) β -D-glucosyl-5-hydroxymethyl-deoxycytidine 5'-monophosphate.

Molecular chaperones in T4 assembly

The assembly of the bacteriophage occurs in different, separated stages all under the control of a number of bacteriophage encoded gene products, also called chaperones [53]. The co-chaperone gp31 in a complex with the host chaperone GroEL, facilitates the folding of the T4 major capsid protein gp23, of which 960 copies are necessary during morphogenesis [54,55]. Mutations in either gp31 or GroEL genes cause the formation of amorphous aggregates in the cell, which are similar to inclusion bodies [56]. GroEL consists of monomers of approximately 60 kDa and is found in bacteria, chloroplasts, and mitochondria [57]. Native GroEL is composed of 14 identical subunits, each containing 548 amino acids, in two rings of seven monomers each. The overall shape is a "double donut" with a 125-130 Å diameter and a height of 100-155 Å, with a central 30 Å-hole [58-60]. GroEL binds to several unfolded polypeptides *in vitro*, preventing their premature aggregation and thus promoting their correct folding and oligomerisation [57].

Gp31 is a 12 kDa protein, which is similar to GroES in size and isoelectric point [61], but without significant homology between the amino acid sequences. Like GroES, it is a heptamer and forms a stable complex via its mobile loop with GroEL in the presence of Mg-ATP [55,62,63]. Keppel and co-workers [62] demonstrated that when gp31 is expressed in *E. coli*, the otherwise essential GroES can be deleted.

Gp57A is also a chaperone in the T4 assemblage; the short tail fibre (STF) protein gp12 and the long tail fibre (LTF) proteins gp34 and gp37 need this chaperone for correctly folding. The exact mechanism remains unclear. Gp57A contains 79 amino acids,

is assumed to be oligomeric and acidic (it has an excess of 9 negative charges) and its composition is somewhat strange, as it does not contain any Phe, Trp, Tyr, His, Pro or Cys [64,65].

Receptor recognition

The T4 bacteriophage recognises the host cell by its receptors: The long tail fibres recognise the outer membrane protein C (OmpC) or lipo-polysaccharide (LPS) in *E.coli* B and are responsible for the initial, reversible, attachment of the bacteriophage. After change of conformation of the baseplate, the short tail fibres extend from the baseplate and bind irreversibly to the core region of the host cell's LPS. Both LPS and OmpC are present in the outer cell wall of all Gram negative bacteria. LPSs often called endotoxins, are complex molecules with molecular weights of about 10 kDa and their compositions can vary widely between different species. The general architecture of LPS is shown in Figure 1.7 [66].

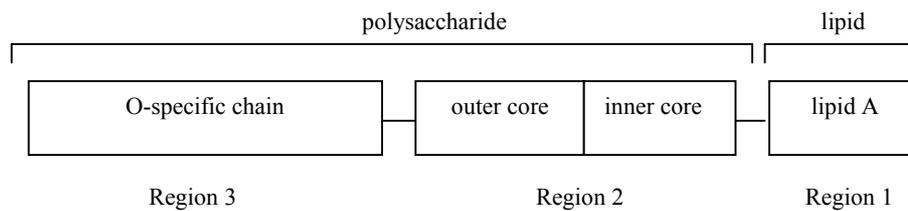


Figure 1.7. General overview of lipopolysaccharide. The polysaccharide domain is responsible for the immunogenicity of the LPS, whereas the lipid domain is responsible for the toxicity. Picture adapted from [66].

Region 1 is composed of Lipid A and is the hydrophobic, membrane-anchoring region of LPS. This domain is responsible for the toxicity of LPS. Lipid A consists of a phosphorylated N-acetylglucosamine (NAG) dimer with usually 6 saturated fatty acids attached. The structure of region 1 is highly conserved among different species. Region 2 is called the core (R) antigen or R polysaccharide, it is attached to the 6 position of one NAG and contains a short chain of sugars. Two unusual sugars are present most of the time in the core polysaccharide; heptose and 2-keto-3-deoxyoctonic acid (KDO). Region 2 is very similar among species. Region 3, called somatic (O) antigen or O polysaccharide is attached to the core polysaccharide, consisting of repeating oligosaccharide subunits.

Region 2 and 3 together are responsible for the immunogenicity of the LPS. Region 3 contains the hydrophilic domain of the LPS molecule. There are major differences in this region between different species and even between strains of Gram negative bacteria [66]. In B type *E. coli*, the distal end of the LPS has a glucose region, which is the important residue for receptor function, whereas in the K-12 strain the corresponding glucose is masked by additional glucose and galactose molecules [67,68].

OmpC, the alternative receptor for the bacteriophage T4, is a trimeric protein often called porin. Its molecular weight is 40368 Da. OmpC forms pores or channels through the outer membrane to allow passage of hydrophilic molecules. Porins allow nutrients to pass through the membrane inwards, while excluding harmful hydrophobic compounds.

Bacteriophage T4 morphology

Assembly of the T4 bacteriophage can be divided into three independent stages: head, tail and long tail fibre assembly. Some stages or assembly steps are dependent on specific chaperones (see above). An overview of the assembly of bacteriophage T4, the required stoichiometries and the necessary chaperones is given in Figure 1.8 [69].

Before the work described in this thesis, the structures of several gene products of the bacteriophage T4 or chaperones necessary in the different assembly steps were already elucidated by X-ray crystallography or NMR spectroscopy. In Table 1.1, the proteins of known three-dimensional structure are given with their corresponding location in the bacteriophage, and the resolution to which the structures were determined. As can be seen in Table 1.1, Kostyuchenko and co-workers fitted the crystallographic data of many gene products into a cryo-electron microscopy map of the base plate obtained to 12.0 Å resolution [70].

Table 1.1. Overview of the bacteriophage T4 gene products (gp) of which the structure has been determined by X-ray crystallography and EM. Denoted are their corresponding location in bacteriophage T4, PDB-code, method of structure determination, and the resolution to which the structures are determined, respectively.

Gene product	location of gp	PDB code	Method	resolution (Å)	reference
gp11	base plate - STF connector	1EL6	X-ray	2.00	[71]
		1PDF	EM	12.00	[70]
gp8	base plate	1N7Z	X-ray	2.00	[72]
		1N8O	X-ray	2.45	[72]
		1N8B	X-ray	2.90	[72]
		1PDM	EM	12.00	[70]
gp12	short tail fibre	1H6W	X-ray	1.90	[73]
		1PDI	EM	12.00	[70]
gp9	baseplate - LTF connector	1QEX	X-ray	2.30	[74]
		1S2E	X-ray	2.30	[74]
gp27	base plate - gp5 connector	1K28	X-ray	2.90	[75]
		1PDJ	EM	12.00	[70]
gp5	cell puncturing device and tail associated lysozyme	1K28	X-ray	2.90	[75]
		1PDL	EM	12.00	[70]
Wac	whisker antigen control	1RFO	NMR		[76]
		1OX3	X-ray	2.00	[77]

The different constituents of the bacteriophage T4 are described below.

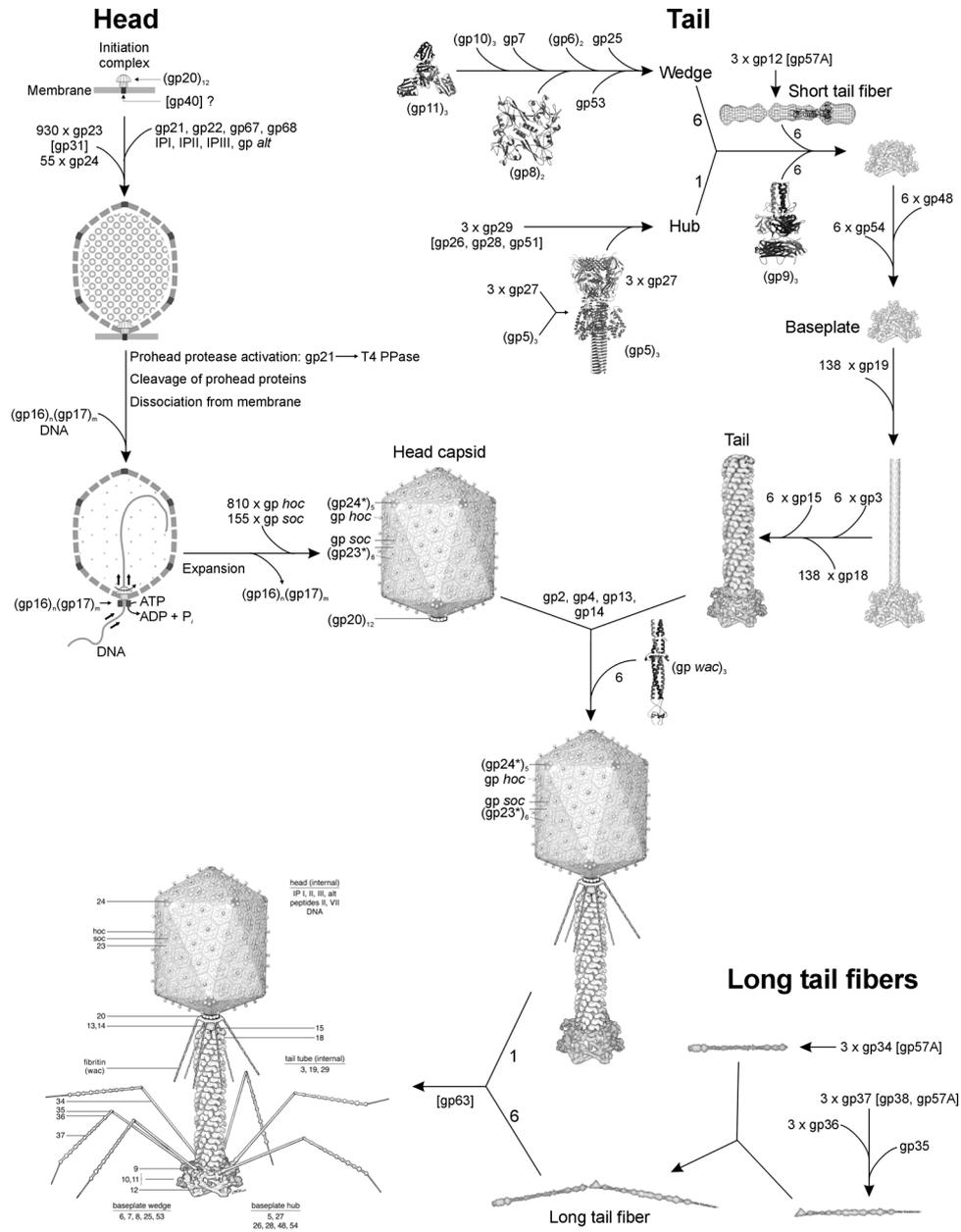


Figure 1.8. Morphogenesis of bacteriophage T4. The overall assembly can be divided into three independent stages: head, tail, and long tail fibre assembly. The chaperones and catalytic proteins are indicated in brackets near the protein, or assembly step, that requires the chaperone. Known stoichiometries are given in the subscript. Crystal structures of structural proteins are shown as ribbon diagrams. (Picture kindly supplied by P. Leiman [69]).

The double stranded-DNA containing head

The T4 head structure was determined by cryo-electron microscopy [78,79] (Figure 1.9). The head is composed of 160 hexamers of gp23* (According to phage genetics usage, gpX* signifies the product of maturation by the cleavage of gpX to gpX*) and 11 pentamers of gp24* together with *hoc* (highly antigenic outer capsid protein) and *soc* (small outer capsid protein). Together they form a shell of about 30 Å thickness encapsulating the double stranded DNA [78,80]. Whereas the *soc* protein helps to stabilise the capsid against extremes in pH, *hoc* only has a marginal effect on head stability [80,81]. Both proteins are not absolutely necessary for head morphogenesis and phage infection. The mature T4 head is elongated along the five-fold axis (Figure 1.9). The diameter was found to vary from around 973 Å along the five-fold axes to about 879 Å along the three-fold and two-fold axes. The length of the mature T4 head is around 1150 Å. The surface of the prolate icosahedron is composed of two end-caps each made of five equilateral triangular facets and connected by an elongated midsection, made of ten triangular facets. The facets of the T4 head are composed of gp23* [78,79]. The eleven vertices are occupied by pentamers of gp24*, whereas the 12th vertex is a special portal for DNA packing, tail attachment, and DNA exit. The portal protein, gp20, assembles as a dodecamer and is often called the "connector" [82,83].

Bacteriophage T4 tail

The tail of T4 consists of two concentric cylinders. The contractile outer sheath and the tube consist of gp18 and gp19, respectively. The inner cylinder, called the tail tube, is built of 144 copies of gp19 [84-86]. The outer diameter of the tail tube is 90 Å, with a 40 Å-diameter inner channel, through which the DNA passes from the T4 head to the host cell [87]. The outer cylinder is called the tail sheath, has an outer width of about 210 Å and is thought to be composed of 144 copies of gp18 [84]. However, Leiman *et al.* [88] recently determined that only 138 copies of gp18 and gp19 are present in the T4 tail. The top end of the tail tube contains gp3, which probably acts as a "paste" protein between the tail sheath and tail tube. It stabilises the tail sheath and prepares the tail for the addition of the terminal capping protein, gp15. Vianelli [50] and co-workers concluded that gp3 is an integral part of the tail, localised at the tip of the tube and capable of preventing abnormal extension of the tail tube during assembly.

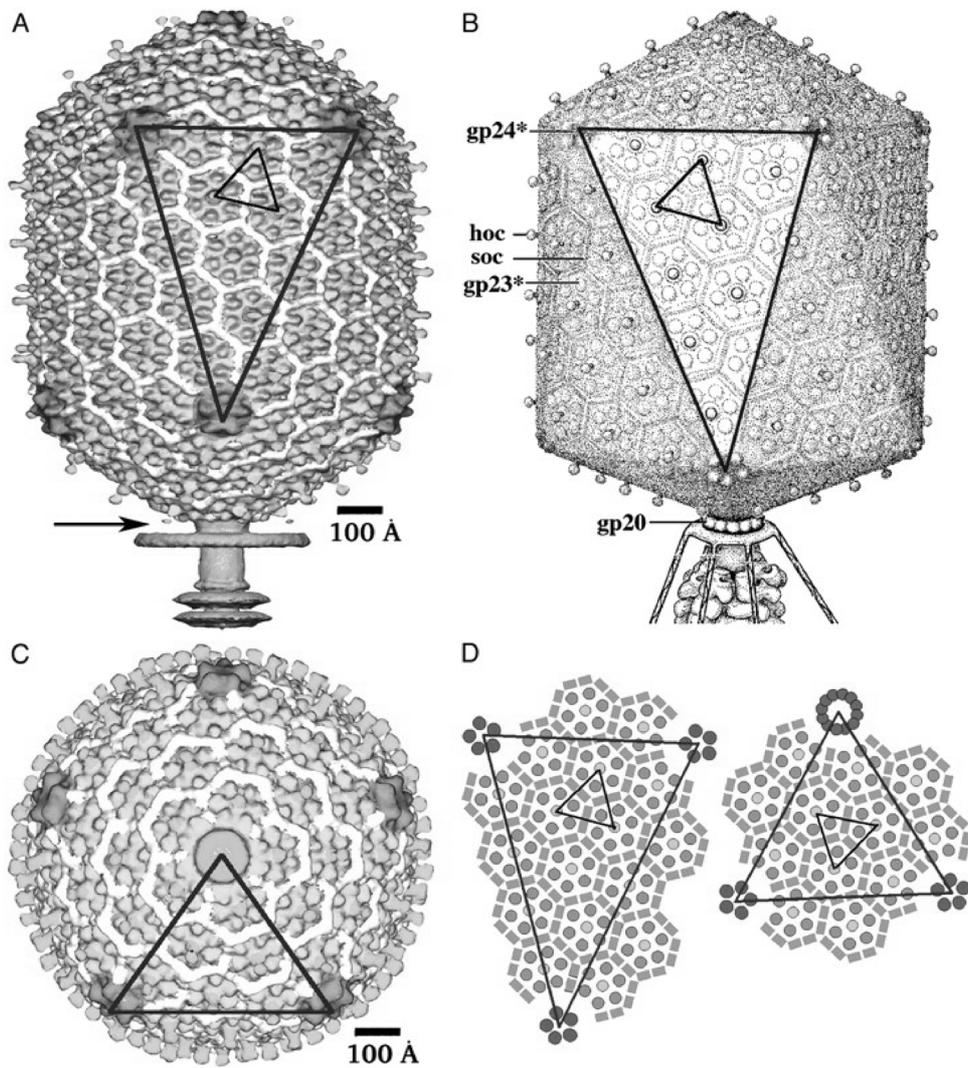


Figure 1.9. Structure of the bacteriophage T4 head. The facet triangles are shown in blue and the basic triangles are shown in black. A) Shaded surface representation of the cryo-EM reconstruction viewed perpendicular to the five-fold axis. Gp23* is shown in blue, gp24* is in magenta, *soc* is in white, and *hoc* is in yellow and the tail is in green. B) Model of the previously proposed T4 head structure. C) View of the reconstruction along the five-fold axis at the portal vertex towards the observer; the tail has been cut away at the level of the black arrow in A. Proteins are coloured as described in A. D, Left) Schematic representation of distribution of proteins in the elongated midsection facet. D, Right) Schematic representation of an end-cap facet. Proteins are coloured as in A, except for the *soc* molecules, which are shown as grey rectangles. (Picture from [79]). For colour picture see Appendix A.

Baseplate

The three-dimensional structure of the bacteriophage T4 baseplate was determined to a resolution of 12 Å by cryo-electron microscopy [70]. The baseplate contains approximately 150 sub-units of at least 16 different proteins ranging from 14 kDa to 140 kDa in size (Figure 1.10). It is a dome-like structure with down-facing pins at the vertex and has a diameter of 520 Å [70]. The baseplate is built out of six identical wedges which surround a central hub [89]. The wedges are built by the sequential assemblage of gp11, gp10, gp7, gp8, gp6, gp53 and gp25 [70].

The central hub contains gp5 and gp27 [75] (Figure 1.11). Gp27 serves as an interface (symmetry-adjuster) between the six wedges and the threefold-symmetry of the hub [90]. Two β-barrel domains of gp27 in the trimer are related by quasi-sixfold and exact three-fold symmetry [75]. Gp5 has several domains of which one is the so-called "tail lysozyme" [91]. Gp5 is the only baseplate protein that undergoes processing by proteolysis and the only one that has enzymatic activity. Its lysozyme domains digest the intermembrane peptidoglycan layer of the host's cell wall during penetration.

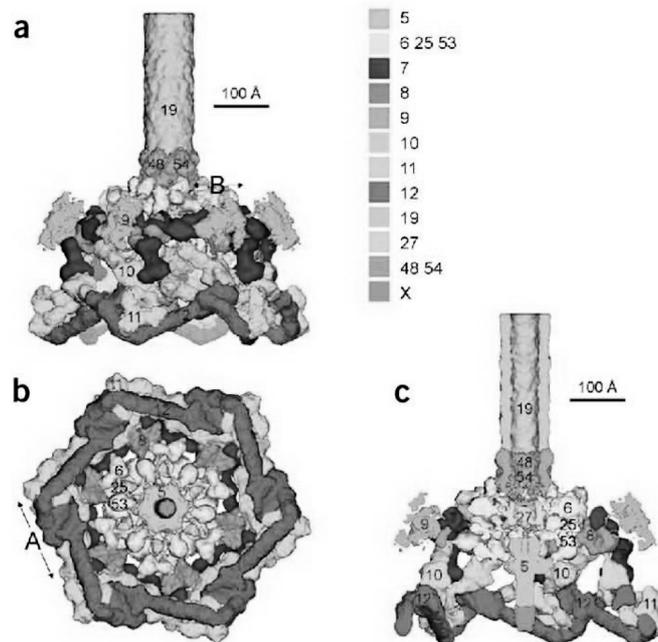


Figure 1.10. Structure of the baseplate tail tube complex. a-c) The baseplate and proximal part of the tail tube. Colours identify proteins labelled with their corresponding gene number. Unidentified protein X at the tip of gp5. a) Side view. b) End-on view. c) Cross-section. Picture adapted from [70]. For colour picture see Appendix A.

The C-terminal domain of gp5 acts as a membrane-puncturing needle and the N-terminal domain of gp5 is inserted into a cylinder formed by three gp27 monomers, which may serve as a channel for DNA ejection after cell wall penetration and dissociation of the needle part of gp5 [75] (Figure 1.11).

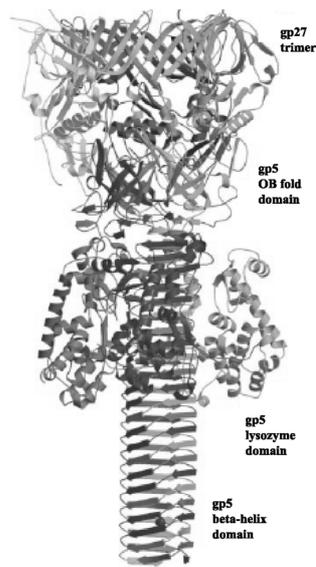


Figure 1.11. Structure of the gp5-gp27 trimeric complex, shown with its threefold axis in the plane of the paper. Picture from [75].

Long and short tail fibres

Long and short tail fibres are connected to the baseplate via gp9 (long-tail fibre connecting protein) and gp11 (short-tail fibre connecting protein), respectively. The long tail fibres are composed of gp34, gp35, gp36 and gp37. These fibres are approximately 1450 Å long and up to 40 Å in diameter. They recognise the OmpC or LPS of *E. coli* by their C-terminal domain and are responsible for the initial, reversible attachment of the bacteriophage. After at least three long tail fibres have bound, the baseplate changes conformation from the "hexagon" form to the "star" form [88,92] (Figure 1.12). In the hexagon form the short tail fibres, trimers of gp12, are incorporated into the baseplate in bent fashion, as can be seen in Figure 1.10 [75]. Upon conversion of the baseplate to the star form, STFs extend from the base plate and bind irreversible with their C-terminal domain to the core region of the host cell LPS [93]. Here they form inextensible stays,

allowing penetration of the cell envelope by the base plate hub and tail tube upon contraction of the outer tail-sheath [75].

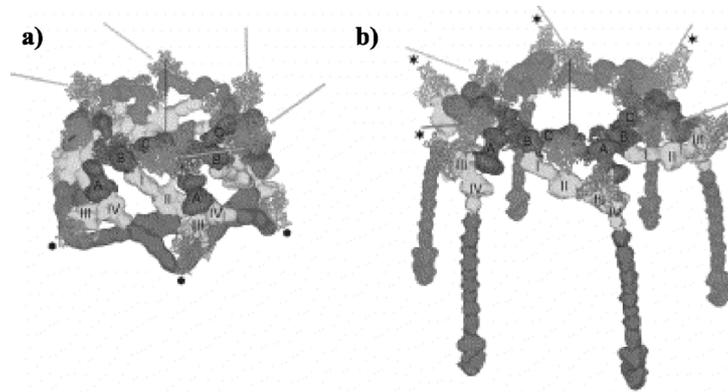


Figure 1.12. Image analysis of a) normal/hexagonal-form and b) contracted/star-form of T4 baseplate. Different colours identify different proteins as in figure 1.10; gp7 (red), gp8 (blue), gp9 (green), gp10 (yellow), gp11 (cyan), and gp12 (magenta). Directions of the long tail fibers are indicated with gray rods. Picture adapted from [88]. For colour picture see Appendix A.

The DNA injection machinery

After the C-terminal domain of the STFs have irreversibly bound to the core region of the host cell LPS [93], the tail sheath contracts, driving the rigid tail tube through the outer cell membrane, using the needle that is located at the end of the tube. The puncturing needle is formed by the gp5 C-terminal β -helix. When the β -helix comes into contact with the periplasmic peptidoglycan layer, it is thought to dissociate, activating the three lysozyme domains of gp5 (Figure 1.11). These lysozyme domains digest the peptidoglycan layer, enabling the tail tube to reach the cytoplasmic membrane of the host cell. Finally, the viral DNA is injected into the host cell cytoplasm, after which the replication, maturation and re-infection can take place.

1.4

Scope and aim of this thesis

The initial aim of my thesis research was to determine the 3D structure of the components of the human T cell receptor-CD3 complex by X-ray crystallography. I focussed on the intra- and extracellular domains of the CD3 δ and CD3 ϵ subunits. The intracellular domains of the CD3 δ and CD3 ϵ chains were synthesised to overcome problems inherent to the method of overexpression and the related problem of obtaining unfolded proteins. Subsequent NMR experiments showed these synthetic peptides to be primarily random coiled, explaining why crystallisation experiments did not (and could not) result in crystals. Next, the external parts of the CD3 δ and CD3 ϵ chain were overexpressed in different *E. coli* strains using different plasmids. Overexpression of both components succeeded, but they could not be refolded in their active conformation. Periplasmic overproduction of the CD3 ϵ extracellular domain resulted in ~ 0.5 mg of purified protein per litre of medium. At that time others published the NMR structure [23] of the ectodomains of the CD3 $\epsilon\gamma$ heterodimer. Hence I shifted the focus of my research to the interaction between CD3 γ and CD3 ϵ and the cell surface expression of the TCR-CD3 complex in human T cells, in response to mutations in CD3 γ . The reported NMR structure was used to decide which mutations in the CD3 γ extracellular domain would influence the CD3 $\epsilon\gamma$ binding interface and consequently the heterodimer formation. By inserting CD3 γ containing different mutations into a CD3 γ -deficient T cell line three phenotypes were obtained. One of these phenotypes was not able to express a CD3 $\delta\epsilon\epsilon\gamma\zeta\zeta$ complex due to a deleterious mutation in the CD3 γ -chain and therefore probably expressed a CD3 $\delta\epsilon\delta\epsilon\zeta\zeta$ complex on the cell surface. Other mutations present in the CD3 γ -chain inserted into the CD3 γ -deficient T cell line resulted in a restored cell surface expression of the CD3 $\delta\epsilon\epsilon\gamma\zeta\zeta$ complex, although some mutations with an impaired interaction between CD3 ϵ and CD3 γ .

A different but related subject of research was solving crystal structures of the T4 proteins in order to understand the mechanism of the adsorption of bacteriophage T4 prior to infection by DNA injection. Understanding the host cell recognition of T4 also has implications for development of bacteria-detection systems. This was done in

collaboration with Stefan Miller of PROFOS AG (Regensburg, Germany). A potential application of this research is the capture of bacteria or bacterial components *e.g.* LPS (endotoxin) using columns containing tails of bacteriophage T4. Endotoxin removal is important in avoiding artefacts and misinterpretation caused by endotoxin contamination in highly sensitive stimulation experiments in cell culture or animal models. An important step in this host cell recognition of the T4 bacteriophage is the attachment of the short tail fibres to the host. These short tail fibres form inextensible stays during infection and therewith they allow penetration of the host's cell envelope. Short tail fibres are formed by a single protein, gp12, which form a parallel, in-register, homotrimer of 527 residues per subunit. To investigate the host cell recognition of the bacteriophage T4, crystals were made of proteolytic fragments of the short tail fibres containing the receptor-binding domain. These structures revealed a surprising new fold, a knitted trimeric metal-binding fold. Despite crystals containing LPS remaining elusive, we mapped the LPS binding site according to the surface potential and the aromatic side-chains present on the surface of the receptor-binding domain.

In a side project with the biotech company Pharming in Leiden, I determined the structure of transgenically expressed human lactoferrin. Pharming reported the production of recombinant hLF (rhLF) in the milk of transgenic cows [94] and in comparative studies between rhLF and hLF from human milk (natural hLF) demonstrated equal biological activities. These studies revealed identical iron-binding and release properties, and despite differences in N-linked glycosylation, equal effectiveness in various infection models [94]. In spite of the presence of polymorphic sides and differences in the N-linked glycosylation, I demonstrated the three-dimensional structure of rhLF to be otherwise identical to the structure of the human milk-derived lactoferrin.

In collaboration with researchers at the University of Wageningen, I crystallised a potato serine protease inhibitor (PSPI). Protease inhibitors have regained interest because of their potent activity in preventing carcinogenesis in a wide variety of *in vivo* and *in vitro* model systems. The PSPI is a dimeric double-headed Kunitz type inhibitor of which no high-resolution structural data are available. The PSPI crystals diffracted to 1.8 Å and a native data set and several derivatives could be collected. Unfortunately, molecular replacement using different model proteins could not solve the structure. The derivatives were found to contain insufficient number of heavy atoms. The project is therefore ongoing.

References

1. www.britannica.com/nobel/micro/340_17.html.
2. Janeway, C., Travers, P., Walport, M. and Capra, J. (1996). Immunobiology, the immune system in health and disease.
3. <http://press2.nci.nih.gov/sciencebehind/immune/immune00.htm>.
4. Meffre, E., Casellas, R. and Nussenzweig, M.C. (2000). Antibody regulation of B cell development. *Nat. Immunol.* **1**, 379-385.
5. Dessen, A., Lawrence, C.M., Cupo, S., Zaller, D.M. and Wiley, D.C. (1997). X-ray crystal structure of HLA-DR4 (DRA*0101, DRB1*0401) complexed with a peptide from human collagen II. *Immunity* **7**, 473-481.
6. Fremont, D.H., Monnaie, D., Nelson, C.A., Hendrickson, W.A. and Unanue, E.R. (1998). Crystal structure of I-Ak in complex with a dominant epitope of lysozyme. *Immunity* **8**, 305-317.
7. Fremont, D.H., Hendrickson, W.A., Marrack, P. and Kappler, J. (1996). Structures of an MHC class II molecule with covalently bound single peptides. *Science* **272**, 1001-1004.
8. Smith, K.J., Reid, S.W., Harlos, K., McMichael, A.J., Stuart, D.I., Bell, J.I. *et al* (1996). Bound water structure and polymorphic amino acids act together to allow the binding of different peptides to MHC class I HLA-B53. *Immunity* **4**, 215-228.
9. Meuer, S.C., Acuto, O., Hussey, R.E., Hodgdon, J.C., Fitzgerald, K.A., Schlossman, S.F. *et al* (1983). Evidence for the T3-associated 90K heterodimer as the T-cell antigen receptor. *Nature* **303**, 808-810.
10. Wilson, I.A. and Garcia, K.C. (1997). T-cell receptor structure and TCR complexes. *Curr. Opin. Struct. Biol.* **7**, 839-848.
11. Davis, M.M., Boniface, J.J., Reich, Z., Lyons, D., Hampl, J., Arden, B. *et al* (1998). Ligand recognition by alpha beta T cell receptors. *Annu. Rev. Immunol.* **16**, 523-544.
12. Garcia, K.C., Teyton, L. and Wilson, I.A. (1999). Structural basis of T cell recognition. *Annu. Rev. Immunol.* **17**, 369-397.
13. Germain, R.N. (2001). The T cell receptor for antigen: signaling and ligand discrimination. *J. Biol. Chem.* **276**, 35223-35226.
14. Torres, P.S., Alcover, A., Zapata, D.A., Arnaud, J., Pacheco, A., Martin-Fernandez, J.M. *et al* (2003). TCR dynamics in human mature T lymphocytes lacking CD3 gamma. *J. Immunol.* **170**, 5947-5955.
15. Perez-Aciego, P., Alarcon, B., Arnaiz-Villena, A., Terhorst, C., Timon, M., Segurado, O.G. *et al* (1991). Expression and function of a variant T cell receptor complex lacking CD3-gamma. *J. Exp. Med.* **174**, 319-326.
16. Sommers, C.L., Dejarnette, J.B., Huang, K., Lee, J., El Khoury, D., Shores, E.W. *et al* (2000). Function of CD3 epsilon-mediated signals in T cell development. *J. Exp. Med.* **192**, 913-919.
17. Bentley, G.A., Boulot, G., Karjalainen, K. and Mariuzza, R.A. (1995). Crystal structure of the beta chain of a T cell antigen receptor. *Science* **267**, 1984-1987.
18. Fields, B.A., Malchiodi, E.L., Li, H., Ysern, X., Stauffacher, C.V., Schlievert, P.M. *et al* (1996). Crystal structure of a T-cell receptor beta-chain complexed with a superantigen. *Nature* **384**, 188-192.

19. Fields, B.A., Ober, B., Malchiodi, E.L., Lebedeva, M.I., Braden, B.C., Ysern, X. *et al* (1995). Crystal structure of the V alpha domain of a T cell antigen receptor. *Science* **270**, 1821-1824.
20. Reiser, J.B., Darnault, C., Guimezanes, A., Gregoire, C., Mosser, T., Schmitt-Verhulst, A.M. *et al* (2000). Crystal structure of a T cell receptor bound to an allogeneic MHC molecule. *Nat. Immunol.* **1**, 291-297.
21. Garcia, K.C., Degano, M., Stanfield, R.L., Brunmark, A., Jackson, M.R., Peterson, P.A. *et al* (1996). An alphabeta T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. *Science* **274**, 209-219.
22. Garboczi, D.N., Ghosh, P., Utz, U., Fan, Q.R., Biddison, W.E. and Wiley, D.C. (1996). Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* **384**, 134-141.
23. Sun, Z.J., Kim, K.S., Wagner, G. and Reinherz, E.L. (2001). Mechanisms contributing to T cell receptor signaling and assembly revealed by the solution structure of an ectodomain fragment of the CD3 epsilon gamma heterodimer. *Cell* **105**, 913-923.
24. Kjer-Nielsen, L., Dunstone, M.A., Kostenko, L., Ely, L.K., Beddoe, T., Mifsud, N.A. *et al* (2004). Crystal structure of the human T cell receptor CD3{epsilon}{gamma} heterodimer complexed to the therapeutic mAb OKT3. *Proc. Natl. Acad. Sci. USA* **101**, 7675-7680.
25. Koning, F., Maloy, W.L. and Coligan, J.E. (1990). The implications of subunit interactions for the structure of the T cell receptor-CD3 complex. *Eur. J. Immunol.* **20**, 299-305.
26. Borst, J., Prendiville, M.A. and Terhorst, C. (1983). The T3 complex on human thymus-derived lymphocytes contains two different subunits of 20 kDa. *Eur. J. Immunol.* **13**, 576-580.
27. Koning, F., Lew, A.M., Maloy, W.L., Valas, R. and Coligan, J.E. (1988). The biosynthesis and assembly of T cell receptor alpha- and beta-chains with the CD3 complex. *J. Immunol.* **140**, 3126-3134.
28. Koning, F., Maloy, W.L., Cohen, D. and Coligan, J.E. (1987). Independent association of T cell receptor beta and gamma chains with CD3 in the same cell. *J. Exp. Med.* **166**, 595-600.
29. Borst, J., Alexander, S., Elder, J. and Terhorst, C. (1983). The T3 complex on human T lymphocytes involves four structurally distinct glycoproteins. *J. Biol. Chem.* **258**, 5135-5141.
30. Samelson, L.E., Harford, J.B. and Klausner, R.D. (1985). Identification of the components of the murine T cell antigen receptor complex. *Cell* **43**, 223-231.
31. Huppa, J.B. and Ploegh, H.L. (1997). In vitro translation and assembly of a complete T cell receptor-CD3 complex. *J. Exp. Med.* **186**, 393-403.
32. Punt, J.A., Roberts, J.L., Kearse, K.P. and Singer, A. (1994). Stoichiometry of the T cell antigen receptor (TCR) complex: each TCR/CD3 complex contains one TCR alpha, one TCR beta, and two CD3 epsilon chains. *J. Exp. Med.* **180**, 587-593.
33. de la Hera A., Muller, U., Olsson, C., Isaaz, S. and Tunnacliffe, A. (1991). Structure of the T cell antigen receptor (TCR): two CD3 epsilon subunits in a functional TCR/CD3 complex. *J. Exp. Med.* **173**, 7-17.
34. San Jose, E., Sahuquillo, A.G., Bragado, R. and Alarcon, B. (1998). Assembly of the TCR/CD3 complex: CD3 epsilon/delta and CD3 epsilon/gamma dimers associate indistinctly with both TCR alpha and TCR beta chains. Evidence for a double TCR heterodimer model. *Eur. J. Immunol.* **28**, 12-21.

35. Fernandez-Miguel, G., Alarcon, B., Iglesias, A., Bluethmann, H., Alvarez-Mon, M., Sanz, E. *et al* (1999). Multivalent structure of an alphabeta T cell receptor. *Proc. Natl. Acad. Sci. USA* **96**, 1547-1552.
36. Exley, M., Wileman, T., Mueller, B. and Terhorst, C. (1995). Evidence for multivalent structure of T-cell antigen receptor complex. *Mol. Immunol.* **32**, 829-839.
37. Thibault, G. and Bardos, P. (1995). Compared TCR and CD3 epsilon expression on alpha beta and gamma delta T cells. Evidence for the association of two TCR heterodimers with three CD3 epsilon chains in the TCR/CD3 complex. *J. Immunol.* **154**, 3814-3820.
38. Call, M.E., Pyrdol, J. and Wucherpfennig, K.W. (2004). Stoichiometry of the T-cell receptor-CD3 complex and key intermediates assembled in the endoplasmic reticulum. *EMBO J.* **23**, 2348-2357.
39. Call, M.E. and Wucherpfennig, K.W. (2004). Molecular mechanisms for the assembly of the T cell receptor-CD3 complex. *Mol. Immunol.* **40**, 1295-1305.
40. Brenner, M.B., Trowbridge, I.S. and Strominger, J.L. (1985). Cross-linking of human T cell receptor proteins: association between the T cell idiotype beta subunit and the T3 glycoprotein heavy subunit. *Cell* **40**, 183-190.
41. Geisler, C. (1992). Failure to synthesize the CD3-gamma chain. Consequences for T cell antigen receptor assembly, processing, and expression. *J. Immunol.* **148**, 2437-2445.
42. Call, M.E., Pyrdol, J., Wiedmann, M. and Wucherpfennig, K.W. (2002). The organizing principle in the formation of the T cell receptor-CD3 complex. *Cell* **111**, 967-979.
43. Huang, Y. and Wange, R.L. (2004). T cell receptor signaling: beyond complex complexes. *J. Biol. Chem.* **279**, 28827-28830.
44. Weiss, A. (1993). T cell antigen receptor signal transduction: a tale of tails and cytoplasmic protein-tyrosine kinases. *Cell* **73**, 209-212.
45. Boniface, J.J., Rabinowitz, J.D., Wulfig, C., Hampl, J., Reich, Z., Altman, J.D. *et al* (1998). Initiation of signal transduction through the T cell receptor requires the multivalent engagement of peptide/MHC ligands [corrected]. *Immunity* **9**, 459-466.
46. Alarcon, B., Gil, D., Delgado, P. and Schamel, W.W. (2003). Initiation of TCR signaling: regulation within CD3 dimers. *Immunol. Rev.* **191**, 38-46.
47. Werlen, G. and Palmer, E. (2002). The T-cell receptor signalosome: a dynamic structure with expanding complexity. *Curr. Opin. Immunol.* **14**, 299-305.
48. Abraham, R.T. and Weiss, A. (2004). Jurkat T cells and development of the T-cell receptor signalling paradigm. *Nat. Rev. Immunol.* **4**, 301-308.
49. Kutter, E.M., Guttman, B., Mosig, G. and Rypniewski, W. (1990). T4 genomic map. *In: Genomic Maps, O'Brien S. J. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY* 1-27.
50. Vianelli, A., Wang, G.R., Gingery, M., Duda, R.L., Eiserling, F.A. and Goldberg, E.B. (2000). Bacteriophage T4 self-assembly: localization of gp3 and its role in determining tail length. *J. Bacteriol.* **182**, 680-688.
51. Mathews, C.K. (1994). An Overview of the T4 Developmental Program. *In: Molecular biology of bacteriophage T4 (ed. Karam, J.D.) American Society for Microbiology, Washington DC*, 1-8.

52. Kaiser, G. (2002). Viruses, Bacteriophage life cycles.
<http://www.cat.cc.md.us/courses/bio141/lecguide/unit2/viruses/lytic.html>.
53. Georgopoulos, C.P. and Linder, C.H. (1994). Molecular chaperones in T4 assembly. *In: Molecular biology of bacteriophage T4 (ed. Karam, J.D.) American Society for Microbiology, Washington DC*, 213-217.
54. Kurochkina, L.P. and Mesyanzhinov, V.V. (1999). Co-expression of gene 31 and 23 products of bacteriophage T4. *Biochemistry (Mosc.)* **64**, 379-383.
55. van der Vies, S.M., Gatenby, A.A. and Georgopoulos, C. (1994). Bacteriophage T4 encodes a co-chaperonin that can substitute for *Escherichia coli* GroES in protein folding. *Nature* **368**, 654-656.
56. Laemmli, U.K., Beguin, F. and Gujer-Kellenberger, G. (1970). A factor preventing the major head protein of bacteriophage T4 from random aggregation. *J. Mol. Biol.* **47**, 69-85.
57. Ellis, R.J. and van der Vies, S.M. (1991). Molecular chaperones. *Annu. Rev. Biochem.* **60**, 321-347.
58. Ludtke, S.J., Chen, D.H., Song, J.L., Chuang, D.T. and Chiu, W. (2004). Seeing GroEL at 6 Å Resolution by Single Particle Electron Cryomicroscopy. *Structure (Camb.)* **12**, 1129-1136.
59. Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D.C., Joachimiak, A., Horwich, A.L. *et al* (1994). The crystal structure of the bacterial chaperonin GroEL at 2.8 Å. *Nature* **371**, 578-586.
60. Zeilstra-Ryalls, J., Fayet, O. and Georgopoulos, C. (1991). The universally conserved GroE (Hsp60) chaperonins. *Annu. Rev. Microbiol.* **45**, 301-325.
61. Keppel, F., Lipinska, B., Ang, D. and Georgopoulos, C. (1990). Mutational analysis of the phage T4 morphogenetic 31 gene, whose product interacts with the *Escherichia coli* GroEL protein. *Gene* **86**, 19-25.
62. Keppel, F., Rychner, M. and Georgopoulos, C. (2002). Bacteriophage-encoded cochaperonins can substitute for *Escherichia coli*'s essential GroES protein. *EMBO Rep.* **3**, 893-898.
63. Landry, S.J., Taher, A., Georgopoulos, C. and van der Vies, S.M. (1996). Interplay of structure and disorder in cochaperonin mobile loops. *Proc. Natl. Acad. Sci. USA* **93**, 11622-11627.
64. Hashemolhosseini, S., Stierhof, Y.D., Hindennach, I. and Henning, U. (1996). Characterization of the helper proteins for the assembly of tail fibers of coliphages T4 and lambda. *J. Bacteriol.* **178**, 6258-6265.
65. Matsui, T., Griniuviene, B., Goldberg, E., Tsugita, A., Tanaka, N. and Arisaka, F. (1997). Isolation and characterization of a molecular chaperone, gp57A, of bacteriophage T4. *J. Bacteriol.* **179**, 1846-1851.
66. Todar, K. (2004). <http://textbookofbacteriology.net/>.
67. Henning, U. and Jann, K. (1979). Two-component nature of bacteriophage T4 receptor activity in *Escherichia coli* K-12. *J. Bacteriol.* **137**, 664-666.
68. Yu, F. and Mizushima, S. (1982). Roles of lipopolysaccharide and outer membrane protein OmpC of *Escherichia coli* K-12 in the receptor function for bacteriophage T4. *J. Bacteriol.* **151**, 718-722.
69. Leiman, P.G., Kanamaru, S., Mesyanzhinov, V.V., Arisaka, F. and Rossmann, M.G. (2003). Structure and morphogenesis of bacteriophage T4. *Cell Mol. Life Sci.* **60**, 2356-2370.

70. Kostyuchenko, V.A., Leiman, P.G., Chipman, P.R., Kanamaru, S., van Raaij, M.J., Arisaka, F. *et al* (2003). Three-dimensional structure of bacteriophage T4 baseplate. *Nat. Struct. Biol.* **10**, 688-693.
71. Leiman, P.G., Kostyuchenko, V.A., Shneider, M.M., Kurochkina, L.P., Mesyanzhinov, V.V. and Rossmann, M.G. (2000). Structure of bacteriophage T4 gene product 11, the interface between the baseplate and short tail fibers. *J. Mol. Biol.* **301**, 975-985.
72. Leiman, P.G., Shneider, M.M., Kostyuchenko, V.A., Chipman, P.R., Mesyanzhinov, V.V. and Rossmann, M.G. (2003). Structure and location of gene product 8 in the bacteriophage T4 baseplate. *J. Mol. Biol.* **328**, 821-833.
73. van Raaij, M.J., Schoehn, G., Burda, M.R. and Miller, S. (2001). Crystal structure of a heat and protease-stable part of the bacteriophage T4 short tail fibre. *J. Mol. Biol.* **314**, 1137-1146.
74. Kostyuchenko, V.A., Navruzbekov, G.A., Kurochkina, L.P., Strelkov, S.V., Mesyanzhinov, V.V. and Rossmann, M.G. (1999). The structure of bacteriophage T4 gene product 9: the trigger for tail contraction. *Structure Fold. Des.* **7**, 1213-1222.
75. Kanamaru, S., Leiman, P.G., Kostyuchenko, V.A., Chipman, P.R., Mesyanzhinov, V.V., Arisaka, F. *et al* (2002). Structure of the cell-puncturing device of bacteriophage T4. *Nature* **415**, 553-557.
76. Guthe, S., Kapinos, L., Moglich, A., Meier, S., Grzesiek, S. and Kiefhaber, T. (2004). Very fast folding and association of a trimerization domain from bacteriophage T4 fibritin. *J. Mol. Biol.* **337**, 905-915.
77. Boudko, S.P., Strelkov, S.V., Engel, J. and Stetefeld, J. (2004). Design and crystal structure of bacteriophage T4 mini-fibritin NCCF. *J. Mol. Biol.* **339**, 927-935.
78. Olson, N.H., Gingery, M., Eiserling, F.A. and Baker, T.S. (2001). The structure of isometric capsids of bacteriophage T4. *Virology* **279**, 385-391.
79. Fokine, A., Chipman, P.R., Leiman, P.G., Mesyanzhinov, V.V., Rao, V.B. and Rossmann, M.G. (2004). Molecular architecture of the prolate head of bacteriophage T4. *Proc. Natl. Acad. Sci. USA* **101**, 6003-6008.
80. Iwasaki, K., Trus, B.L., Wingfield, P.T., Cheng, N., Campusano, G., Rao, V.B. *et al* (2000). Molecular architecture of bacteriophage T4 capsid: vertex structure and bimodal binding of the stabilizing accessory protein, Soc. *Virology* **271**, 321-333.
81. Ishii, T., Yamaguchi, Y. and Yanagida, M. (1978). Binding of the structural protein soc to the head shell of bacteriophage T4. *J. Mol. Biol.* **120**, 533-544.
82. Driedonks, R.A. and Caldentey, J. (1983). Gene 20 product of bacteriophage T4. II. Its structural organization in prehead and bacteriophage. *J. Mol. Biol.* **166**, 341-360.
83. Driedonks, R.A., Engel, A., tenHeggeler, B. and van Driel. (1981). Gene 20 product of bacteriophage T4 its purification and structure. *J. Mol. Biol.* **152**, 641-662.
84. King, J. and Mykolajewycz, N. (1973). Bacteriophage T4 tail assembly: proteins of the sheath, core and baseplate. *J. Mol. Biol.* **75**, 339-358.
85. Berget, P.B. and King, J. (1983). T4 tail morphogenesis. In: *Bacteriophage T4* (C.K. Mathews, E. M. Kutter, G. Mosig and P. B. Berget, eds.) American Society for Microbiology, Washington, DC, 246-258.
86. King, J. (1968). Assembly of the tail of bacteriophage T4. *J. Mol. Biol.* **32**, 231-262.

87. DeRosier, E.M. and Klug, A. (1968). Reconstruction of three-dimensional structures from electron micrographs. *Nature* **217**, 130-134.
88. Leiman, P.G., Chipman, P.R., Kostyuchenko, V.A., Mesyanzhinov, V.V. and Rossmann, M.G. (2004). Three-dimensional rearrangement of proteins in the tail of bacteriophage T4 on infection of its host. *Cell* **118**, 419-429.
89. Kikuchi, Y. and King, J. (1975). Genetic control of bacteriophage T4 baseplate morphogenesis. III. Formation of the central plug and overall assembly pathway. *J. Mol. Biol.* **99**, 695-716.
90. Rossmann, M.G., Mesyanzhinov, V.V., Arisaka, F. and Leiman, P.G. (2004). The bacteriophage T4 DNA injection machine. *Curr. Opin. Struct. Biol.* **14**, 171-180.
91. Nakagawa, H., Arisaka, F. and Ishii, S. (1985). Isolation and characterization of the bacteriophage T4 tail-associated lysozyme. *J. Virol.* **54**, 460-466.
92. Crowther, R.A., Lenk, E.V., Kikuchi, Y. and King, J. (1977). Molecular reorganization in the hexagon to star transition of the baseplate of bacteriophage T4. *J. Mol. Biol.* **116**, 489-523.
93. Riede, I. (1987). Receptor specificity of the short tail fibres (gp12) of T-even type Escherichia coli phages. *Mol. Gen. Genet.* **206**, 110-115.
94. van Berkel, P.H., Welling, M.M., Geerts, M., van Veen, H.A., Ravensbergen, B., Salaheddine, M. *et al* (2002). Large scale production of recombinant human lactoferrin in the milk of transgenic cows. *Nat. Biotechnol.* **20**, 484-487.
95. Pouvreau, L., Gruppen, H., Van Koningsveld, G.A., Van Den Broek, L.A. and Voragen, A.G. (2003). The most abundant protease inhibitor in potato tuber (cv. Elkana) is a serine protease inhibitor from the Kunitz family. *J. Agric. Food Chem.* **51**, 5001-5005.
96. Pouvreau, L., Gruppen, H., Piersma, S.R., Van Den Broek, L.A., Van Koningsveld, G.A. and Voragen, A.G. (2001). Relative abundance and inhibitory distribution of protease inhibitors in potato juice from cv. Elkana. *J. Agric. Food Chem.* **49**, 2864-2874.

Human T Cell Receptor - CD3 complex



2.1

The impact of single amino acid substitutions in CD3 γ on the CD3 ϵ interaction and T cell receptor-CD3 complex formation

E.A.J. Thomassen, E.H.A. Spaenij-Dekking, A. Thompson, K.L. Franken, Ö. Sanal, J.P. Abrahams, M. J. D. van Tol, and F. Koning

submitted

Summary

The human T cell receptor-CD3 complex consists of at least eight polypeptide chains: CD3 $\gamma\epsilon$ - and $\delta\epsilon$ -dimers associate with the disulphide linked $\alpha\beta$ - and $\zeta\zeta$ -dimers to form a functional receptor complex. The exact structure of this complex is still unknown. We have now examined the interaction between CD3 γ and CD3 ϵ in human T cells. For this purpose we have generated site directed mutants of CD3 γ that were introduced in human T cells defective in CD3 γ -expression. Intracellular as well as cell surface expression of the introduced CD3 γ chains was determined as well as the association with CD3 δ , CD3 ϵ and the T cell receptor. Three phenotypes were observed: i) the introduction of wild type CD3 γ and CD3 γ (78Y-F) fully restored the T cell receptor assembly and expression; ii) the introduction of CD3 γ (82C-S), CD3 γ (85C-S), and CD3 γ (76Q-E) all resulted in an impaired association between CD3 γ and CD3 ϵ and a lack of cell surface expressed CD3 γ ; iii) the introduction of CD3 γ (76Q-L) and CD3 γ (78Y-A) restored the expression of TCR-CD3 $\delta\epsilon\gamma\epsilon\zeta_2$ complexes although the association between CD3 γ and CD3 ϵ was impaired. These results indicate that several amino acids in CD3 γ are essential for an optimal association between CD3 γ and CD3 ϵ and for the assembly of a cell surface expressed TCR-CD3 $\delta\epsilon\gamma\epsilon\zeta_2$ complex.

Introduction

The majority of human T cells expresses a clonotypic $\alpha\beta$ TCR heterodimer. For the full function of the receptor, the disulphide-linked $\alpha\beta$ -chains associate with CD3 δ , CD3 ϵ , CD3 γ and ζ . These latter form non-covalently linked $\delta\epsilon$ and $\epsilon\gamma$ heterodimers and disulphide-linked ζ - ζ homodimers. While each TCR-CD3 complex thus consists of a minimum of eight polypeptides, the exact stoichiometry of the complex is still under discussion [1-4]. The TCR $\alpha\beta$, which has a Fab-like structure (Fragment antigen binding), is responsible for the recognition of a specific antigen bound to MHC-molecules. Subsequently, the CD3 and ζ -components mediate signal transduction and intracellular activation [5-8]. The CD3 δ , ϵ , and γ all have a large extracellular immunoglobulin (Ig)-like domain, a membrane proximal stalk region, a transmembrane helix, and an intracellular immunoreceptor tyrosine-based activation motif (ITAM) containing domain. This is in contrast to ζ , which has a small extracellular domain and a large intracellular domain with three ITAMs. The CD3 components are not only required for transduction of the signal across the cell membrane, but also for the expression of the TCR heterodimers on the surface of T cells. In the absence of one of the CD3 chains, *e.g.* due to a deleterious mutation in one of them, a reduced number of T cell receptors is present on the cell surface [9-11].

CD3 deficiencies in man are very rare autosomal disorders. In 1986, Regueiro *et al* [12] reported a human CD3 γ deficiency [13]. This was the first primary T cell receptor immunodeficiency in human for which the genetic basis could be elucidated, MIM (Mendelian Inheritance in Man) number 186740. In 1990 another deficiency (MIM 186830) followed, reported by Thoenes *et al.*, which was later described as a CD3 ϵ deficiency [14-16]. Recently Dadi *et al.* [17] studied three cases of CD3 δ deficiency. In total three cases of human CD3 γ deficiency have been published in the mutation database, of which two are Spanish siblings, while the third patient is a Turkish male with an A to T mutation at position 242 in his CD3 γ DNA, which changes a lysine codon (AAA) to an early stop codon (TAA) (not shown & [18]).

In 2001, Sun *et al.* described mechanisms contributing to T cell receptor signalling and assembly, as revealed by the structure of the ectodomain of the murine CD3 $\epsilon\gamma$ heterodimer [19]. Mutational analysis of CD3 ϵ , focusing on the binding interface between

CD3 ϵ and γ , implicated several amino acids in CD3 ϵ and γ as being important for the domain-domain interaction. In this analysis combinations of mutations in CD3 ϵ were found to be required for strong effects on the association between CD3 ϵ and CD3 γ . More recently, the structure of the human CD3 $\epsilon\gamma$ heterodimer complexed to the OKT3 mAb was elucidated [20]. Small differences were observed between the human and the murine structure, but whether these are caused by the different origins of the heterodimers (human CD3 ϵ and CD3 γ share 41% and 43% sequence homology to their murine counterparts, respectively) or by the different methods employed (crystallography versus NMR (nuclear magnetic resonance)) is unknown.

In the present study we further analysed the binding interface of the CD3 $\epsilon\gamma$ dimer in human T cells, by introducing mutations in CD3 γ and determining the ability of such mutated CD3 γ chains to form CD3 $\gamma\epsilon$ dimers and participate in the formation of cell surface expressed T cell receptor-CD3 complexes. The results demonstrate that single amino acid alterations in CD3 γ can have a significant effect on the cell surface expression of the TCR-CD3 complex and provide further evidence that $\alpha\beta$ TCR $\delta\epsilon\delta\epsilon\zeta\zeta$ complexes can be formed even when the association of CD3 $\epsilon\gamma$ -dimers is impaired.

Results

Characterisation of TCR-CD3 expression on T cells of a patient with a CD3 γ deficiency

A male patient with a CD3 γ deficiency is a child of parents that are first cousins. At the age of 4 he was diagnosed with CD3 γ deficiency [21]. An AAA to TAA mutation at position 242 of CD3 γ leads to an early stop codon (not shown & [18]). Because of this mutation, the remaining transcript would encode a 46 amino acid long protein lacking both the transmembrane region and cytoplasmic domain. PBMCs of this patient and his healthy brother were subjected to FACS analysis (Figure 2.1A). A CD3-PE/CD4-FITC and CD3-PE/CD8-FITC analysis indicated that the patient had diminished levels of CD3 expression on the cell surface of both T cell populations. Next, T cell lines were established from the PBMCs of the patient, his brother and parents. At day 14 of the T cell establishment a CD3-PE/ $\alpha\beta$ TCR-FITC analysis demonstrated that more than 96 % of the

T cells of the healthy brother expressed an $\alpha\beta$ TCR on the cell surface, while only about 2 % of the patient's cells were $\alpha\beta$ TCR⁺ (Figure 2.1B).

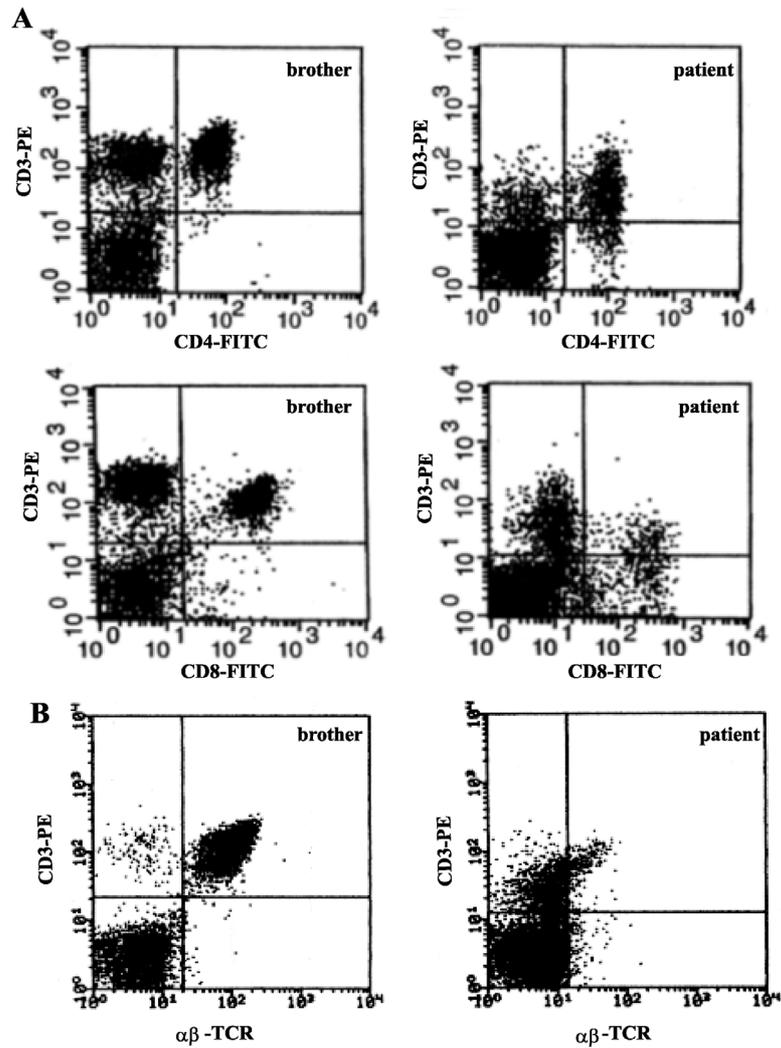


Figure 2.1. A) The FACS analysis of double stained PBMCs of the patient (right panel) and healthy brother (left panel). The upper panel shows the CD3/CD4 staining for patient and healthy brother. The lower panel shows the CD3/CD8 staining. It is clearly visible that the patient had diminished levels of CD3 expression on the cell surface of both T cell populations. B) The FACS analysis at day 14 of the T cell establishment from PBMCs. The left panel shows the CD3/ $\alpha\beta$ -TCR staining of the brother and the right panel shows the CD3/ $\alpha\beta$ -TCR analysis for the CD3 γ deficient patient. Only 2% of the patient's cells were $\alpha\beta$ TCR⁺, in contrast to 96% for the healthy brother.

To investigate the cell surface expression of the individual CD3 γ -, δ -, and ϵ -chains, the cell lines were labelled with ^{125}I and lysed in NP40 lysis buffer. Subsequently immunoprecipitations were carried out with CD3 δ , ϵ and γ -specific antibodies followed by 1-dimensional SDS-PAGE analysis (Figure 2.2). In NP40 lysisbuffer the TCR-CD3 complex dissociates into TCR $\alpha\beta$ -, $\zeta\zeta$ -, CD3 $\gamma\epsilon$ - and CD3 $\delta\epsilon$ -dimers [22]. Consequently, CD3 γ and ϵ are present in CD3 γ -immunoprecipitates, CD3 δ and ϵ in CD3 δ immunoprecipitates and all three CD3 chains in CD3 ϵ immunoprecipitates. While in the lysates of the cell lines from the healthy brother and his parents CD3 δ , ϵ and γ were present and associated as expected, the lysate of the cell line of the CD3 γ deficient patient contained very little CD3 δ and ϵ while CD3 γ was undetectable (Figure 2.2).

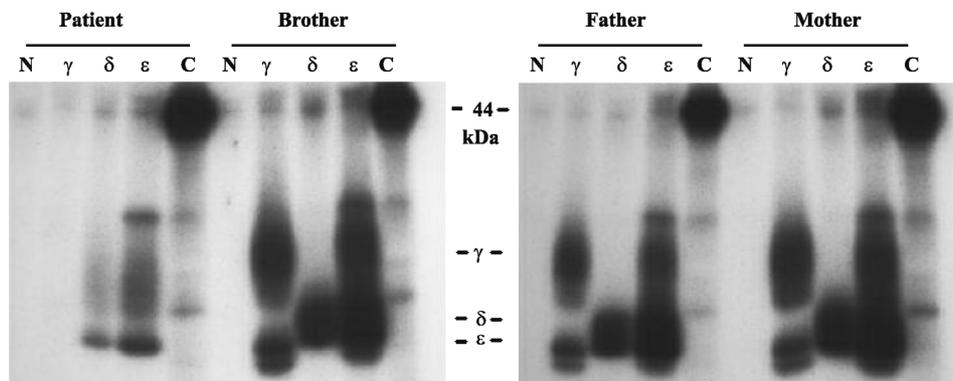


Figure 2.2. SDS-PAGE analyses of immunoprecipitates obtained from NP40 lysates after cell surface iodination of T cell lines of patient, brother, father, and mother. Antisera used were normal rabbit serum as negative control (N), anti-CD3 δ (δ), anti-CD3 ϵ (ϵ), and anti-CD3 γ (γ), and as reference (C) anti-HLA class I (only heavy chain is shown here, 44 kDa). The positions of the different chains are indicated in the middle. It can be seen that the patient's cells contained very little CD3 δ and CD3 ϵ , while CD3 γ is undetectable.

To verify the presence of TCR-CD3 components intracellularly, the T cell lines were labelled with ^{35}S -methionine/cysteine and after lysis in NP40 lysis buffer specific immunoprecipitations were carried out with CD3 γ , δ , ϵ and ζ -specific antibodies followed by SDS-PAGE analysis. This demonstrated (Figure 2.3) that the CD3 γ , δ , ϵ and ζ chains were present intracellularly in the T-cell lines obtained from the parents and the healthy brother, while the T cells of the patient contained CD3 δ , ϵ and ζ chains at levels comparable to the healthy controls but no CD3 γ . A two-dimensional non-reducing/reducing SDS-PAGE analysis of the ζ immunoprecipitate confirmed the normal

expression of a ζ -dimer (Figure 2.4). To determine the expression of the covalently linked TCR $\alpha\beta$ chains, immunoprecipitations were obtained with the anti-CD3 ϵ antibody from digitonin lysates of the metabolically labelled cells from the patient and his healthy brother (digitonin is known to preserve the subunit interactions between the TCR and CD3 complexes [23]). These immunoprecipitates were subjected to two-dimensional non-reducing/reducing SDS-PAGE analysis which revealed normal intracellular expression of the disulfide linked TCR $\alpha\beta$ dimers in the T cells of the healthy brother and the patient (Figure 2.4). Thus, the patient synthesises all TCR-CD3 chains except CD3 γ .

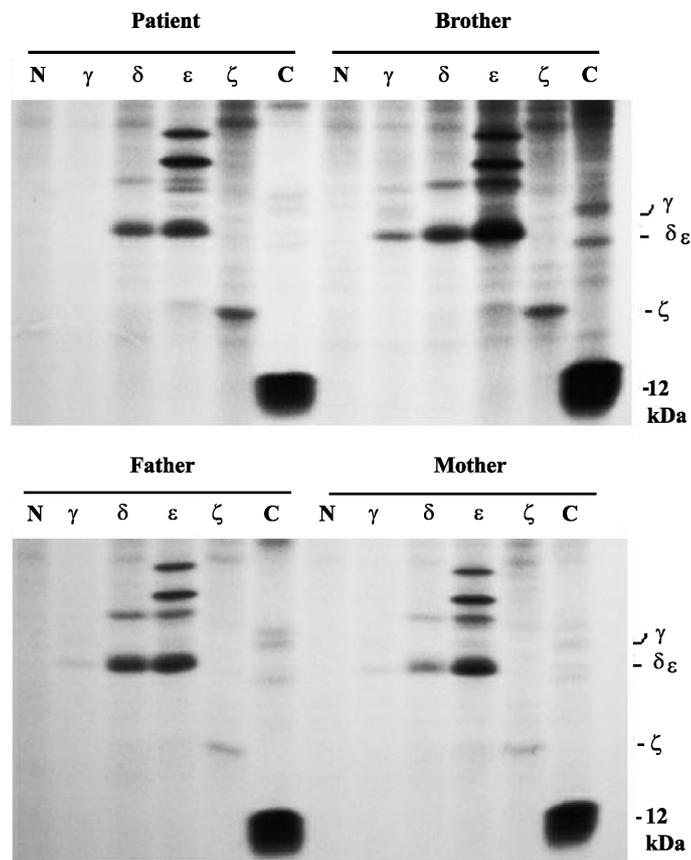


Figure 2.3. SDS-PAGE analysis of immunoprecipitates obtained from NP40 lysates after metabolic labelling of T cell lines of patient, brother, father and mother. Antisera used were normal rabbit serum as negative control (N), anti-CD3 δ (δ), anti-CD3 ϵ (ϵ), anti-CD3 ζ (ζ), and anti-CD3 γ (γ), and as reference (C) anti-HLA class I (only β_2 M is shown here, 12 kDa). The positions of the different chains are indicated on the right. This shows that the parents and healthy brother contain all the CD3 chains, in contrast to the patient which lacks the CD3 γ chain.

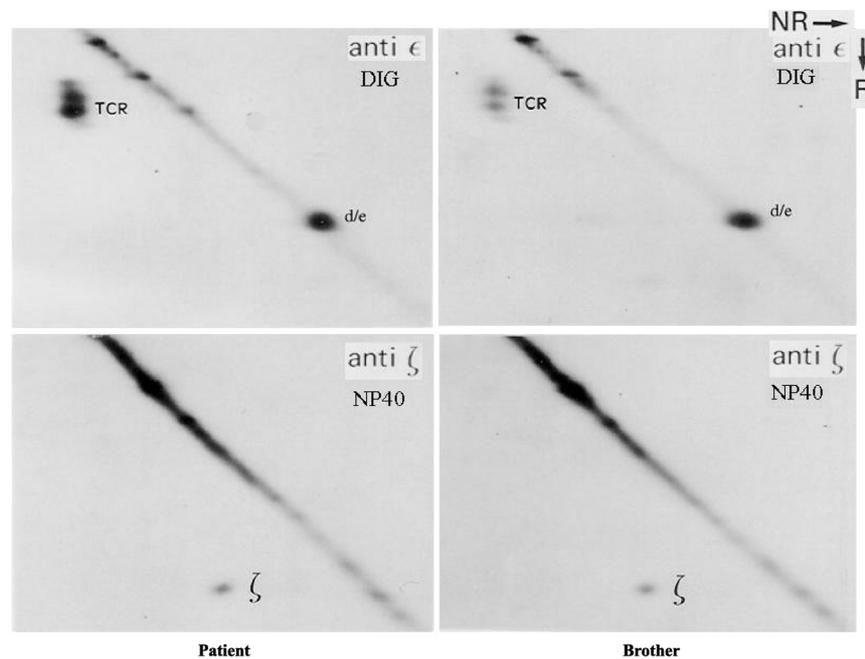


Figure 2.4. 2D (first dimension non-reducing, second reducing) SDS-PAGE analysis of anti-CD3 ϵ and anti- ζ immunoprecipitates. The T cells of patient (left panel) and the healthy brother (right panel) were metabolically ^{35}S labelled and lysates were prepared with digitonin (upper panel) or NP40 (lower panel) lysis buffer. Antisera used were anti-CD3 ϵ (upper panel), anti-CD3 ζ (lower panel). This demonstrates the normal expression of the ζ -dimer and $\alpha\beta$ TCR for the patient compared to the healthy brother.

Influence of amino acid substitutions in CD3 γ on assembly and cell surface expression of the TCR-CD3 complex

Several amino acids in the binding interface between CD3 γ and CD3 ϵ have been implicated to be important for the specific interaction between these two chains (Figure 2.5, adapted from [19]). In particular, the amino acids at position 76, 78, 82 and 85 in CD3 γ are thought to interact with amino acids in CD3 ϵ . Mutational analysis of CD3 ϵ has demonstrated that multiple replacements in the binding interface are required for abolishment of the CD3 $\gamma\epsilon$ interactions. Such an analysis has not been performed for CD3 γ . We have now taken advantage of the availability of the T cell line of the CD3 γ deficient patient to investigate this in detail. For this purpose site directed mutants of CD3 γ cDNA were generated encoding CD3 γ -chains in which the amino acids thought to

be important for the interaction with CD3 ϵ are substituted by either homologous or non-homologous amino acids (see Table 2.1). These CD3 constructs, as well as wild type CD3 γ , were stably introduced into the patient's T cell line together with the reporter gene GFP as described [24,25] and GFP⁺ cells were selected by FACS-sorting. These GFP⁺ T cell lines and the control patient T cells were analysed for expression of the TCR-CD3 complex by FACS analysis using a CD3 ϵ specific antibody. Moreover, the intracellular- and cell surface expression of the CD3 γ , δ , ϵ -chain and the interaction between these chains were determined by SDS-PAGE analysis of CD3 δ , CD3 ϵ and CD3 γ immunoprecipitates carried out with ³⁵S-labeled and ¹²⁵I-labeled cell lysates of the cell lines, respectively.

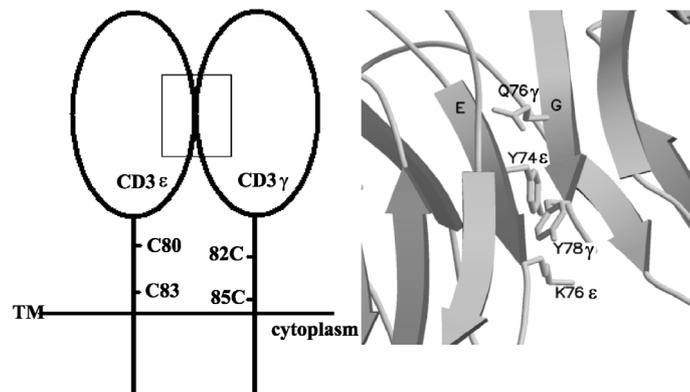


Figure 2.5. Schematic overview of the CD3 $\epsilon\gamma$ binding interface. The important residues are represented by their one-letter amino acid code and their corresponding residue number, followed by the specific monomer. The right figure, which is a zoom-in of the left picture, is adapted from [12]. The figures were prepared with MOLSCRIPT [26] and Raster3D [27].

Three phenotypes were observed. First, the introduction of wild type CD3 γ fully restored the CD3 $\gamma\epsilon$ interaction and TCR-CD3 expression (Figure 2.6, and Table 2.1 for summary). Similarly, the introduction of CD3 γ in which the tyrosine at position 78 is replaced for a phenylalanine (78 Y-F) resulted in restoration of CD3 γ protein expression, CD3 $\gamma\epsilon$ interaction and TCR-CD3 expression. We therefore conclude that the 78 Y-F mutation allows the expression of a TCR-CD3 $\delta\epsilon\gamma\epsilon\zeta_2$ complex on the cell surface. In contrast, the 82 C-S (Figure 2.6, see Table 2.1 for summary), 85 C-S and 76 Q-E mutations (not shown, see Table 2.1 for summary) gave rise to a different phenotype. In

all these cases apparent normal cell surface expression of the TCR-CD3 complex was observed but no CD3 γ was observed on the cell surface and, consequently, no interaction between CD3 γ and CD3 ϵ was detectable (Figure 2.6). To determine if the mutated CD3 γ -chain is synthesised, immunoprecipitation with the CD3 γ specific antibody were carried out using ³⁵S-labeled lysates of the transfectants. The results indicate that CD3 γ is synthesised (Figure 2.7) but that it can not associate with CD3 ϵ since the $\alpha\beta$ TCR and CD3 ϵ could not be detected in the CD3 γ specific immunoprecipitates carried out with DIG lysates. These results show that CD3 γ is expressed intracellularly, but is not capable to associate with the CD3 ϵ chain in order to form a cell-surface $\alpha\beta$ TCR-CD3 $\delta\epsilon\gamma\epsilon\zeta_2$ complex. Finally, the 76 Q-L and 78 Y-A mutations gave rise to a third phenotype: while CD3 γ protein was present in the cell surface expressed complex (Figure 2.6, see Table 2.1 for summary), the association with CD3 ϵ was impaired since no CD3 γ could be detected in the CD3 ϵ immunoprecipitate. Longer exposures also failed to show the presence of CD3 γ (not shown). The presence of CD3 γ on the cell surface, however, indicates that these cells do express TCR-CD3 $\delta\epsilon\gamma\epsilon\zeta_2$ complexes.

Table 2.1. Analysis of CD3 γ expression, heterodimer formation and subsequent TCR-CD3 complex cell surface expression of the different mutations, wild type and the patient.

mutation	76Q-E	76Q-L	78 Y-F	78 Y-A	82 C-S	85 C-S	wt	patient
phenotype	2	3	1	3	2	2	1	
γ intra	n/a	n/a	n/a	+	n/a	+	+(b)	-
γ cell surface	-*	+	++	+	-	-	++	-
$\epsilon\gamma$ cell surface	-	+/-	++	+/-	-	-	++	-
$\epsilon\delta$ cell surface	+	+	++	+	+	+	++	+
complex cell surface	$\epsilon\delta\epsilon\delta$	$\epsilon\gamma\epsilon\delta$	$\epsilon\gamma\epsilon\delta$	$\epsilon\gamma\epsilon\delta$	$\epsilon\delta\epsilon\delta$	$\epsilon\delta\epsilon\delta$	$\epsilon\gamma\epsilon\delta$	$\epsilon\delta\epsilon\delta$

* If the level of CD3 γ or CD3 $\epsilon\gamma$ is denoted as “-“, it is possible that undetectable amounts are present at the cell surface, n/a = not analysed, wt = wild type, full CD3 γ inserted, b = brother was used as reference for a healthy person. Phenotype 1 = fully restored TCR assembly and expression, 2 = impaired association between CD3 γ and CD3 ϵ and lack of cell surface expressed TCR-CD3 $\delta\epsilon\gamma\epsilon\zeta_2$ complex, and 3 = restored CD3 $\delta\epsilon\gamma\epsilon\zeta_2$ complexes although impaired association between CD3 γ and CD3 ϵ .

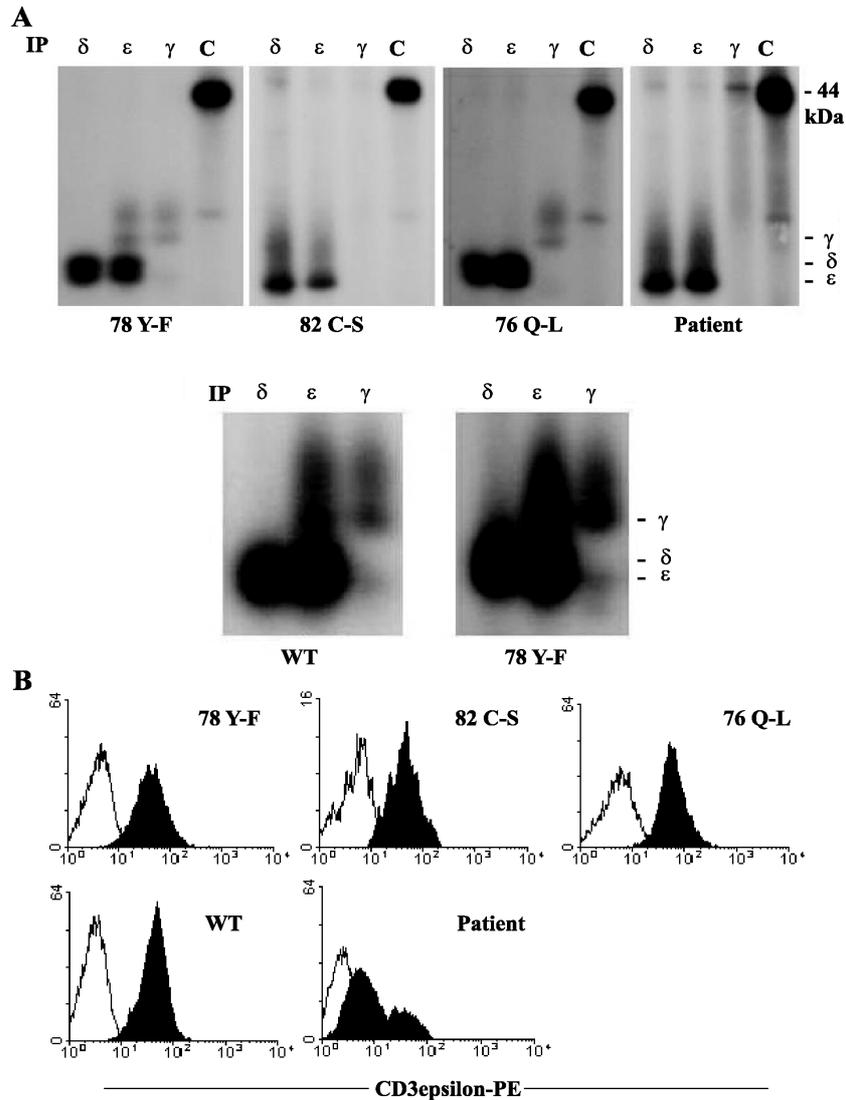


Figure 2.6. A) SDS-PAGE analysis of immunoprecipitates obtained from NP40 lysates after cell surface iodination of patient and the cells contain the CD3 γ with different mutations. Antisera used were, anti-CD3 δ (δ), anti-CD3 ϵ (ϵ), and anti-CD3 γ (γ) and anti HLA class I (c, only heavy chain is shown). Analysis of the WT was done in a different experiment causing differences in autoradiography time and radioactivity of ^{125}I . To show that the 78Y-F mutation has the same expression levels compared to the WT, the 78Y-F is shown twice, once with the WT (lower panel) and once with all phenotypes (upper panel). B) Histograms of patient and mutant GFP $^+$ T cells. Unstained cells are represented by a white histogram, and cells after staining with anti-CD3-PE are represented by a black histogram. The abscissa gives the fluorescence intensity in a logarithmic scale, the ordinate gives the relative cell number.

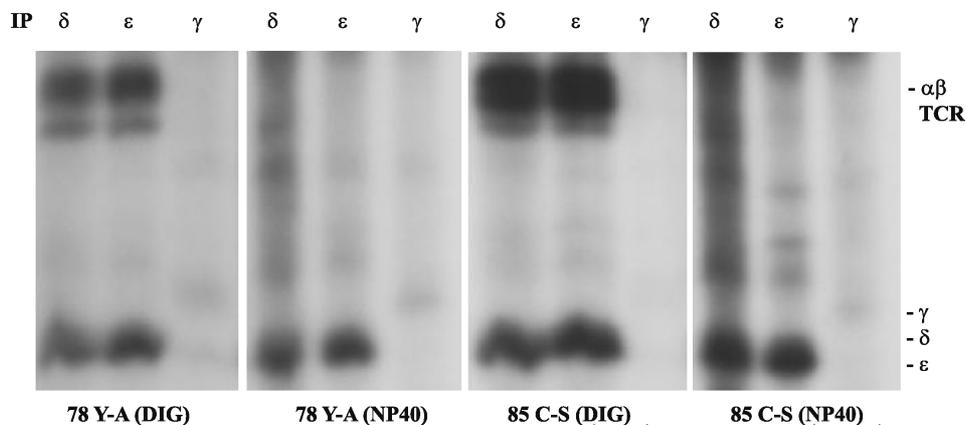


Figure 2.7. SDS-PAGE analysis of immunoprecipitates obtained from NP40 and DIG lysates after ^{35}S metabolic labelling of CD3 γ mutants. Antisera used were, anti-CD3 δ (δ), anti-CD3 ϵ (ϵ), and anti-CD3 γ (γ). The positions of the different chains are indicated on the right. This shows that CD3 γ is expressed intracellularly even in mutants that lack cell surface CD3 γ expression (85 C-S).

Discussion

The interaction of the T cell receptor with its MHC-peptide ligand is a crucial step towards the initiation of adaptive immune responses. Due to the short cytoplasmic tails of the T cell receptor α - and β -chains, however, these can not transduce the signal over the cell membrane. This is accomplished by the T cell receptor associated CD3 γ , δ , ϵ and ζ -chains. The mechanism by which this is performed is still an ill understood process. Although several studies have indicated that the minimal T cell receptor-CD3 complex contains 8 chains, $\alpha\beta\text{TCR-CD3}\delta\epsilon\gamma\zeta\zeta$, the exact stoichiometry of the complex is still not clear, nor is the way in which all the individual components are arranged in the complex. A NMR study has for the first time provided information on the structure of part of the extracellular domains of a murine CD3 $\gamma\epsilon$ -complex [19]. In this study several amino acids located in the interface between CD3 ϵ and CD3 γ were implicated as being important for the association. Mutational analysis of residues in the stalk region of CD3 ϵ indicated that combinations of mutations were required for disturbing the interaction of CD3 ϵ with CD3 γ . We have now generated mutants of CD3 γ and stably introduced these in T cells from a patient with a deleterious mutation in the CD3 γ gene, resulting in aberrant TCR-

CD3 expression. TCR-CD3 complexes in such a patient are most likely composed of TCR-CD3 $\delta\epsilon\zeta\zeta$ complexes instead of TCR-CD3 $\delta\epsilon\gamma\zeta\zeta$ as has been previously suggested by Geisler [28] and Pérez-Aciego *et al.* [10]. In agreement with previous studies we find that the reconstitution with wt CD3 γ completely restores the cell surface expression levels of the TCR, CD3 δ , CD3 ϵ , and CD3 γ [9,29]. In contrast, replacement of either one of the two cysteine residues in the proximal stalk region of CD3 γ to serine abrogates the interaction between CD3 ϵ and CD3 γ . This is in disagreement with Sun *et al.* [19] who suggested that these conserved cysteines in the stalk region facilitate the CD3 $\epsilon\gamma$ pairing but are not required for the association of these CD3 chains.

The CD3 γ (78 Y-F) mutation did not effect the CD3 $\gamma\epsilon$ association and thus led to the cell-surface expression of the TCR-CD3 $\delta\epsilon\gamma\zeta\zeta$ complex comparable to that observed with wt CD3 γ . A striking result is that 3 mutations, 82 C-S, 76 Q-E, and 85 C-S all resulted in cell surface expression of the TCR-CD3 complex equivalent to the cells expressing wt CD3 γ but in the apparent absence of detectable amounts of CD3 γ itself as evidenced by a lack of (co)immunoprecipitation with the CD3 ϵ and the CD3 γ specific antibody. Intracellularly, however, the expression of these mutated CD3 γ chains could be detected as shown for mutant 85 C-S. It is known that the large majority of intracellularly synthesised TCR and CD3 chains are degraded before they can be assembled in a functional TCR-CD3 complex [30]. Possibly a weak interaction of mutant CD3 γ with CD3 ϵ limits the rate of degradation of CD3 ϵ and leads to a higher cell surface expression of a TCR-CD3 $\delta\epsilon\zeta\zeta$ complex compared to the CD3 γ deficient patient, as has been seen by FACS analysis. Alternatively, the culture conditions may have selected transfectants expressing a higher number of receptors on their cell surface. Irrespective of this, our results demonstrate the importance of the two cysteine residues in the proximal stalk region for the association between CD3 γ and CD3 ϵ .

Finally, the CD3 γ (76Q-L) and CD3 γ (78Y-A) mutations led to intracellular expression and transport of CD3 γ to the cell surface, presumably in a TCR-CD3 $\delta\epsilon\gamma\zeta\zeta$ complex, in spite of an undetectable association with CD3 ϵ . This result indicates that a tight association between CD3 γ and CD3 ϵ is not required for normal expression of an intact TCR-CD3 complex.

From the crystal structure of the CD3 $\epsilon\gamma$ heterodimer complexed with the mAb OKT3 [20] we can make two assumptions about the impact of the CD3 γ mutations. First, the 76Q to E mutation excludes the possibility of hydrogen-bond formation at position Q 76^{Oe1}. However, since all the other interactions, including main-chain hydrogen bonding (see Table 3 of [20]) at the interface will be maintained, it can be concluded that the hydrogen bond formation at Q 76^{Oe1} is very important for the heterodimer formation. Second, the 78Y to F mutation has little impact on the hydrogen bond pattern and the Van der Waals interactions, resulting in a normal association of the heterodimer. This in contrast to the replacement of 78Y to A mutation, which does change the Van der Waals interactions with Leu90^{C δ 1}, Leu92^{C δ 1} and Val16^{C β , C γ 1, C γ 2} of CD3 ϵ (see Table 3 of [20]). This loss of Van der Waals interactions explains the observed loose interaction of CD3 $\epsilon\gamma$.

In conclusion, our results have identified amino acids in the extracellular domain of CD3 γ to play a key role in the association between CD3 γ and CD3 ϵ which is a prerequisite for the cell surface expression of TCR-CD3 $\delta\epsilon\gamma\epsilon\zeta\zeta$ complexes.

Materials and Methods

PBMC isolation and establishment of T cell lines

PBMC were isolated from heparinised blood obtained from patient, his parents and a healthy older brother, by Ficoll density gradient centrifugation. For the generation of T cell lines, PBMC (5×10^6) were cultured in the presence of allogeneic, irradiated (3000 rad) PBMC (10×10^6), 50 U/ml rIL2 (Chiron, Amsterdam, The Netherlands) and 5 μ g/ml PHA (Murex Diagnostics, Dartford, UK) in RPMI 1640 (Gibco) supplemented with 2 mM glutamine, 100 U/ml penicillin, 100 μ g/ml streptomycin and 10 % FCS (Gibco). Twice weekly half of the supernatant was replaced by fresh medium containing 50 U/ml rIL2 and once in a period of three weeks allogeneic, irradiated PMBC were added. After a culture period of 2 months the doses of rIL2 was enhanced to 100 U/ml and cells were harvested a month later.

Derivatisation of mutants and construction of retroviral vectors

A bicistronic vector, LZRSpBMN-linker-IRES-GFP, was used as described in Heemskerk *et al.* [24] with the gene of interest linked to a downstream internal ribosomal

entry site (IRES) and a marker gene (Green fluorescent protein, GFP) that allows independent translation of the products of both genes in the transduced cells. The vector was adapted for the gateway cloning system (Invitrogen, Breda, The Netherlands). Site directed mutagenesis was performed using the polymerase chain reaction according to Landt [31] using a CD3 γ cDNA clone [32] as template with custom-made primers. The mutations were made at AA-position 78, tyrosine (TAT) was changed to a phenylalanine (TTT) and an alanine (GCT), at AA-position 76 where the glutamine (CAA) was mutated to a glutamate (GAA) or a leucine (CTA), and at AA-position 82 and 85 where the cysteines (TGT) were mutated to serines (TCT). The resulting PCR products were cloned into the vector and clones were confirmed by automated fluorescent sequencing.

Generation of retroviral supernatant and retroviral transduction

Phoenix cells, a 293T-based amphotropic retroviral packaging cell line [25] were transfected via the calcium phosphate precipitation method (Life Technologies, Gaithersburg, MD, USA), after 2 days 2 μ g/ml puromycin (Clontech, Palo Alto, CA, USA) was added. At 10-14 days after transfection, 6 x 10⁶ cells per 10 cm² petridish were plated (Beckton Dickinson, Meylan, France) in 10 ml IMDM supplemented with 10 % FBS without puromycin. After 24 hours the medium was refreshed and at the following day retroviral supernatant was harvested, which was frozen at -70 °C. Non-tissue culture treated plates (Beckton Dickinson, Meylan, France) were coated with retronectin (Takara, Shiga, Japan) for two hours at room temperature. Then, the coated wells were blocked with 2 % human serum albumin (CLB, Amsterdam, the Netherlands). T cells were added, and after 30 minutes culture medium was removed, and thawed retroviral supernatant was supplemented. Viral supernatant together with cells were incubated overnight, followed by washing, and finally transferred to cell culture bottles in normal tissue culture medium. After 3 days, transduced cells were sorted by fluorescence-activated cell sorting on a FACSVantage (Becton Dickinson, Mountain View, CA, USA) on basis of high level GFP expression. After several weeks of culture, the GFP⁺ populations of the several mutant T cells were used for the experiments.

Cell culturing

Patient cells and the mutant cells were cultured in Iscove's modified Dulbecco's medium (IMDM, Gibco BRL) supplemented with 3 mM glutamine, 10 % normal human serum (NHS) and 100 U rIL2/ml. Medium was refreshed every three days and the cells were restimulated every 10-14 days with 100 U rIL2, 1 µg/ml PHA and 10×10^6 /ml irradiated (2500 rad) PMBC, as feeder cells.

Antibodies

The antibodies against CD3 δ , CD3 ϵ , and CD3 γ were all rabbit antibodies and directed against peptides corresponding to the Carboxy termini of the human CD3 chains, The anti-CD3 ζ serum was raised against the predicted N-terminus of the human ζ chain as has been described before [33]. The anti-HLA class I antibody (W6/32) [34], which was used as a control, is a mouse monoclonal antibody.

Cell surface iodination and metabolic labelling

For cell surface iodination 10×10^6 cells were washed three times in PBS and resuspended in 30 µl lactoperoxidase (250 U/ml, Sigma-Aldrich) solution. 1 mCi Na¹²⁵I (NEN) was added to the cells followed by the addition of 10 µl 0.05 % H₂O₂/PBS (Sigma-Aldrich) with 5 minutes intervals. After the labelling was stopped and the free iodine was removed by washing the cells three times with 500 µl 2 mM KI/PBS, the cells were solubilised in 750 µl lysisbuffer containing 0.5 % NP40 (Sigma-Aldrich) and protease inhibitors (Complete, EDTA-free tablets, Roche Diagnostics). After 30 minutes incubation on ice, the lysates were centrifuged 15 minutes (13000 x g) at 4 °C. For ³⁵S metabolic labelling, 7×10^6 vital cells were washed with 10 ml RPMI without methionine and cysteine, centrifuged and resuspended in 10 ml of the same medium. Cells were incubated for 30 minutes at 37 °C / 5% CO₂, centrifuged and the resulting pellets were resuspended in 10 ml RPMI without Met/Cys, 100 U/ml IL2 and 1 mCi ³⁵S Met/Cys. After 4 hours incubation at 37 °C / 5% CO₂ cells were centrifuged and washed twice with PBS. The cells were solubilised in 750 µl lysisbuffer containing 0.5 % NP40 or 1 % digitonine, and protease inhibitors. After 30 minutes incubation on ice, the lysates were centrifuged 15 minutes (13000 x g) at 4 °C.

Immunoprecipitation and SDS-PAGE analysis

The lysates, after labelling with ^{125}I or ^{35}S , were precleared by adding 100 μl protein A sepharose CL4B (PAS beads, Amersham Pharmacia Biotech) gently shaking at room temperature followed by the addition of 75 μl normal rabbit serum, and shaken at room temperature for one hour. After the removal of the beads, specific immunoprecipitations with CD3 δ , CD3 ϵ , CD3 ζ , CD3 γ antibodies, normal rabbit serum (NRS) or anti-HLA class I were performed by adding 5 μl antiserum (2 μl for anti-HLA class I) to 100 μl precleared lysate (1D SDS-PAGE) or 200 μl precleared lysate (2D SDS-PAGE) for 60 minutes at room temperature. The PAS beads were washed four times with 500 μl lysis buffer and resuspended in 100 μl 1x sample loading buffer (20 mM Tris-HCl, pH 6.8, 0.01% bromophenol blue, 1% SDS, 10% glycerol, 1% β -mercaptoethanol) and analysed on a reducing 12 % SDS-PAGE. For 2 dimensional SDS-PAGE, the PAS beads were resuspended in loading buffer without 2-mercaptoethanol, and loaded onto the first non-reducing 12% cylindrical-gel. After the first dimension, the gel was incubated for 10 minutes in Laemmli buffer with β -mercaptoethanol and placed on top of the reducing second dimension 12 % SDS-PAGE gel. After drying of the gels autoradiography was performed at $-80\text{ }^{\circ}\text{C}$ using Kodax or Fuji scientific imaging films.

Fluorescence activated cell sorter (FACS) analysis

For single stained cells, the cells were washed with PBS/0.5 % BSA and labelled with mouse anti-human anti-CD3-PE (Becton/Dickinson) at $4\text{ }^{\circ}\text{C}$ for one hour. Followed by washing with PBS/0.5 % BSA and fixating with 0.5 % paraformaldehyde. In the case of CD3/CD4 and CD3/CD8 double staining, the cells were stained with anti-human anti-CD3-PE and washed with PBS/0.5 % BSA and labelled with mouse anti-human anti-CD4-FITC or anti-CD8-FITC antibody. For CD3/ $\alpha\beta$ -TCR double staining cells were stained with anti-human anti-CD3-PE and washed with PBS/0.5 % BSA, followed by labelling with non-directly conjugated anti- $\alpha\beta$ -TCR antibody for one hour at $4\text{ }^{\circ}\text{C}$. After washing the cells were stained with GAM-FITC antibody, washed and fixated.

Acknowledgement

The authors are indebted to Y.M. Kooy-Winkelaar for help with cell culturing and FACS analysis, and P.J. van den Elsen for providing us with the CD3 γ cDNA, Dr. G. Noles for the retroviral vectors. We thank the Drs B.O. Roep and P.J. van den Elsen for critically reading the manuscript.

References

1. Manolios, N., Letourneur, F., Bonifacio, J.S. and Klausner, R.D. (1991). Pairwise, cooperative and inhibitory interactions describe the assembly and probable structure of the T-cell antigen receptor. *EMBO J.* **10**, 1643-1651.
2. San Jose, E., Sahuquillo, A.G., Bragado, R. and Alarcon, B. (1998). Assembly of the TCR/CD3 complex: CD3 epsilon/delta and CD3 epsilon/gamma dimers associate indistinctly with both TCR alpha and TCR beta chains. Evidence for a double TCR heterodimer model. *Eur. J. Immunol.* **28**, 12-21.
3. Exley, M., Wileman, T., Mueller, B. and Terhorst, C. (1995). Evidence for multivalent structure of T-cell antigen receptor complex. *Mol. Immunol.* **32**, 829-839.
4. Fernandez-Miguel, G., Alarcon, B., Iglesias, A., Bluethmann, H., Alvarez-Mon, M., Sanz, E. *et al* (1999). Multivalent structure of an alphabetaT cell receptor. *Proc. Natl. Acad. Sci. USA* **96**, 1547-1552.
5. Wilson, I.A. and Garcia, K.C. (1997). T-cell receptor structure and TCR complexes. *Curr. Opin. Struct. Biol.* **7**, 839-848.
6. Davis, M.M., Boniface, J.J., Reich, Z., Lyons, D., Hampl, J., Arden, B. *et al* (1998). Ligand recognition by alpha beta T cell receptors. *Annu. Rev. Immunol.* **16**, 523-544.
7. Garcia, K.C., Teyton, L. and Wilson, I.A. (1999). Structural basis of T cell recognition. *Annu. Rev. Immunol.* **17**, 369-397.
8. Germain, R.N. (2001). The T cell receptor for antigen: signaling and ligand discrimination. *J. Biol. Chem.* **276**, 35223-35226.
9. Torres, P.S., Alcover, A., Zapata, D.A., Arnaud, J., Pacheco, A., Martin-Fernandez, J.M. *et al* (2003). TCR dynamics in human mature T lymphocytes lacking CD3 gamma. *J. Immunol.* **170**, 5947-5955.
10. Perez-Aciego, P., Alarcon, B., Arnaiz-Villena, A., Terhorst, C., Timon, M., Segurado, O.G. *et al* (1991). Expression and function of a variant T cell receptor complex lacking CD3-gamma. *J. Exp. Med.* **174**, 319-326.
11. Sommers, C.L., Dejarnette, J.B., Huang, K., Lee, J., El Khoury, D., Shores, E.W. *et al* (2000). Function of CD3 epsilon-mediated signals in T cell development. *J. Exp. Med.* **192**, 913-919.
12. Regueiro, J.R., Arnaiz-Villena, A., Ortiz, d.L., Martin Villa, J.M., Vicario, J.L., Pascual-Ruiz, V. *et al* (1986). Familial defect of CD3 (T3) expression by T cells associated with rare gut epithelial cell autoantibodies. *Lancet* **1**, 1274-1275.

13. Arnaiz-Villena, A., Timon, M., Corell, A., Perez-Aciego, P., Martin-Villa, J.M. and Regueiro, J.R. (1992). Brief report: primary immunodeficiency caused by mutations in the gene encoding the CD3-gamma subunit of the T-lymphocyte receptor. *N. Engl. J. Med.* **327**, 529-533.
14. Soudais, C., de Villartay, J.P., Le Deist, F., Fischer, A. and Lisowska-Grospierre, B. (1993). Independent mutations of the human CD3-epsilon gene resulting in a T cell receptor/CD3 complex immunodeficiency. *Nat. Genet.* **3**, 77-81.
15. Soudais, C., de Villartay, J.P., Le Deist, F., Fischer, A. and Lisowska-Grospierre, B. (1993). Genetic analysis of the human CD3-epsilon gene in a T cell receptor/CD3 immunodeficiency. *Immunodeficiency* **4**, 117-119.
16. Thoenes, G., Le Deist, F., Fischer, A., Griscelli, C. and Lisowska-Grospierre, B. (1990). Immunodeficiency associated with defective expression of the T-cell receptor-CD3 complex. *N. Engl. J. Med.* **322**, 1399
17. Dadi, H.K., Simon, A.J. and Roifman, C.M. (2003). Effect of CD3delta deficiency on maturation of alpha/beta and gamma/delta T-cell lineages in severe combined immunodeficiency. *N. Engl. J. Med.* **349**, 1821-1828.
18. Zapata, D.A., Pacheco-Castro, A., Torres, P.S., Millan, R. and Regueiro, J.R. (2000). CD3 immunodeficiencies. *Roifman, C.M. (ed.), Immunology and Allergy Clinics of North America: Inherited T cell Immunodeficiencies. B W Saunders Co.* **20**, 1-17.
19. Sun, Z.J., Kim, K.S., Wagner, G. and Reinherz, E.L. (2001). Mechanisms contributing to T cell receptor signaling and assembly revealed by the solution structure of an ectodomain fragment of the CD3 epsilon gamma heterodimer. *Cell* **105**, 913-923.
20. Kjer-Nielsen, L., Dunstone, M.A., Kostenko, L., Ely, L.K., Beddoe, T., Mifsud, N.A. *et al* (2004). Crystal structure of the human T cell receptor CD3{epsilon}{gamma} heterodimer complexed to the therapeutic mAb OKT3. *Proc. Natl. Acad. Sci. USA* **101**, 7675-7680.
21. Sanal, O., Yel, L., Ersoy, F., Tezcan, I. and Berkel, A.I. (1996). Low expression of T-cell receptor-CD3 complex: a case with a clinical presentation resembling humoral immunodeficiency. *Turk. J. Pediatr.* **38**, 81-84.
22. Pacheco-Castro, A., Alvarez-Zapata, D., Serrano-Torres, P. and Regueiro, J.R. (1998). Signaling through a CD3 gamma-deficient TCR/CD3 complex in immortalized mature CD4+ and CD8+ T lymphocytes. *J. Immunol.* **161**, 3152-3160.
23. Van Neerven, J., Coligan, J.E. and Koning, F. (1990). Structural comparison of alpha/beta and gamma/delta T cell receptor-CD3 complexes reveals identical subunit interactions but distinct cross-linking patterns of T cell receptor chains. *Eur. J. Immunol.* **20**, 2105-2111.
24. Heemskerk, M.H., Blom, B., Nolan, G., Stegmann, A.P., Bakker, A.Q., Weijer, K. *et al* (1997). Inhibition of T cell and promotion of natural killer cell development by the dominant negative helix loop helix factor Id3. *J. Exp. Med.* **186**, 1597-1602.
25. Kinsella, T.M. and Nolan, G.P. (1996). Episomal vectors rapidly and stably produce high-titer recombinant retrovirus. *Hum. Gene Ther.* **7**, 1405-1413.
26. Esnouf, R.M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta.* **D55**, 938-940.
27. Merritt, E.A. and Bacon, D.J. (1997). Raster3D; Photorealistic Molecular Graphics. *Methods In Enzymology* **277**, 505-524.

28. Geisler, C. (1992). Failure to synthesize the CD3-gamma chain. Consequences for T cell antigen receptor assembly, processing, and expression. *J. Immunol.* **148**, 2437-2445.
29. Sun, J.Y., Pacheco-Castro, A., Borroto, A., Alarcon, B., Alvarez-Zapata, D. and Regueiro, J.R. (1997). Construction of retroviral vectors carrying human CD3 gamma cDNA and reconstitution of CD3 gamma expression and T cell receptor surface expression and function in a CD3 gamma-deficient mutant T cell line. *Hum. Gene Ther.* **8**, 1041-1048.
30. Koning, F., Lew, A.M., Maloy, W.L., Valas, R. and Coligan, J.E. (1988). The biosynthesis and assembly of T cell receptor alpha- and beta-chains with the CD3 complex. *J. Immunol.* **140**, 3126-3134.
31. Landt, O., Grunert, H.P. and Hahn, U. (1990). A general method for rapid site-directed mutagenesis using the polymerase chain reaction. *Gene* **96**, 125-128.
32. Krissansen, G.W., Owen, M.J., Verbi, W. and Crumpton, M.J. (1986). Primary structure of the T3 gamma subunit of the T3/T cell antigen receptor complex deduced from cDNA sequences: evolution of the T3 gamma and delta subunits. *EMBO J.* **5**, 1799-1808.
33. Koning, F., Maloy, W.L. and Coligan, J.E. (1990). The implications of subunit interactions for the structure of the T cell receptor-CD3 complex. *Eur. J. Immunol.* **20**, 299-305.
34. Barnstable, C.J., Bodmer, W.F., Brown, G., Galfre, G., Milstein, C., Williams, A.F. *et al* (1978). Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens-new tools for genetic analysis. *Cell* **14**, 9-20.

2.2

Analysis of intracellular CD3 δ and CD3 ϵ synthetic peptides

Summary

The T cell receptor-CD3 complex contains at least eight chains, the disulphide-linked $\alpha\beta$ TCR, and the associating CD3 $\gamma\epsilon$, CD3 $\delta\epsilon$ and ζ_2 dimers. The structural information on these components is limited. This is not due to problems with crystallography or NMR, these techniques got better over the years, but to problems of expressing and isolating the different chains, heterodimers or the whole TCR-CD3 as such. To overcome these problems, synthetic peptides corresponding to the intracellular domains of CD3 δ and CD3 ϵ were synthesised. Crystallisation and NMR studies were performed in order to determine the three-dimensional structures. Unfortunately, crystallisation experiments for CD3 δ only produced "sea urchin" type of crystals of poor quality and micro- or macro seeding did not give diffracting crystals. Folding experiments using NMR show that the synthetic peptides are primarily random coiled.

Introduction

Synthetic peptides were synthesised corresponding to the cytoplasmic domains of the CD3 δ and CD3 ϵ . The CD3 δ is 44-mer, with a molecular weight and a calculated pI of 4995.38 Da and 6.93, respectively. The CD3 ϵ is a 55-mer, with a molecular weight and a pI of 6179.03 Da and 11.52, respectively. These intracellular domains are quite distinctive in their charge. It is thought that the intracellular domains bind to form a heterodimer in the TCR-CD3 complex.

Results

Crystallisation

In order to determine the three-dimensional structure of the intracellular domains of CD3 δ and CD3 ϵ crystallisation experiments were performed. The peptides were dissolved in water at various concentrations (2.5-25 mg ml⁻¹). Different Hampton screens were set up by mixing equal volumes of 1 μ l reservoir solution and protein solution. Unfortunately, none of these conditions resulted in the formation of suitable crystals. After several rounds of macro- and micro seeding for the CD3 δ peptide, the best that was obtained were "sea urchins" (Figure 2.8).

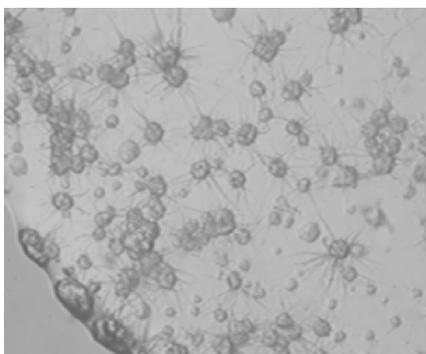


Figure 2.8. Sea urchins of CD3 δ intracellular peptide (25mg/ml) in a hanging drop-vapour experiment. The reservoir solution contains 1.2M ammonium sulphate at pH 5.5.

No set up for CD3 ϵ peptide resulted in any crystal or sea urchin formation. The CD3 ϵ peptide has a high pI and Hampton screens are not available in the higher pH range (around 11) so a screen was made in this pH range. Unfortunately this did also not initiate any crystal formation. A possible explanation for the lack of crystal formation is that both the peptides are (partially) unfolded.

Folding and binding studies on CD3 δ and CD3 ϵ peptides using NMR

In order to determine whether the synthetic peptides used in the crystallisation experiments are folded, NMR experiments were performed. The synthetic peptides were dissolved in water, at a concentration of approximately 1 mM. The pH was adjusted to 5.6 and 6.2 for CD3 δ and CD3 ϵ , respectively. Of both intracellular peptides a 1D proton NMR spectrum (Figure 2.9) was recorded at 298 K on a Bruker DMX-600. The spectra of the single peptides show poor dispersion and narrow line width. Therefore it was concluded that the peptides are primarily random coil.

After these experiments, the peptides were mixed and left for 30 minutes at room temperature. A 1D proton NMR was recorded (Figure 2.9), the spectrum of the combined peptides was a superimposition of the two spectra of the single peptides, leading to the conclusion that no interaction or folding could be measured in this NMR experiment. Therefore further experiments on these domains were abandoned.

Discussion and conclusion

Both the CD3 δ and CD3 ϵ synthetic peptides have no internal structure shown by 1D proton NMR and are therefore unable to form crystals. Mixing the two synthetic peptides did not change the delta-positions of the peaks in the 1D proton NMR spectrum. The combined spectrum is a superimposition of the two spectra of the single peptides, therefore it was concluded that the peptides did not interact at the given experimental conditions. Mixing the peptides does not result in folding of the individual peptides, prohibiting significant interaction between the intracellular domains. Whether the lack of folding and binding of the two peptides after mixing is caused by the absence of the transmembrane and/or extracellular domains or by the fact that the single peptides were primarily unfolded is unknown.

Acknowledgements

We thank W.E. Benckhuijsen and J.W. Drijfhout for synthesis of the peptides and S. van Wetswinkel and G. Siegal for assisting with the NMR experiments.

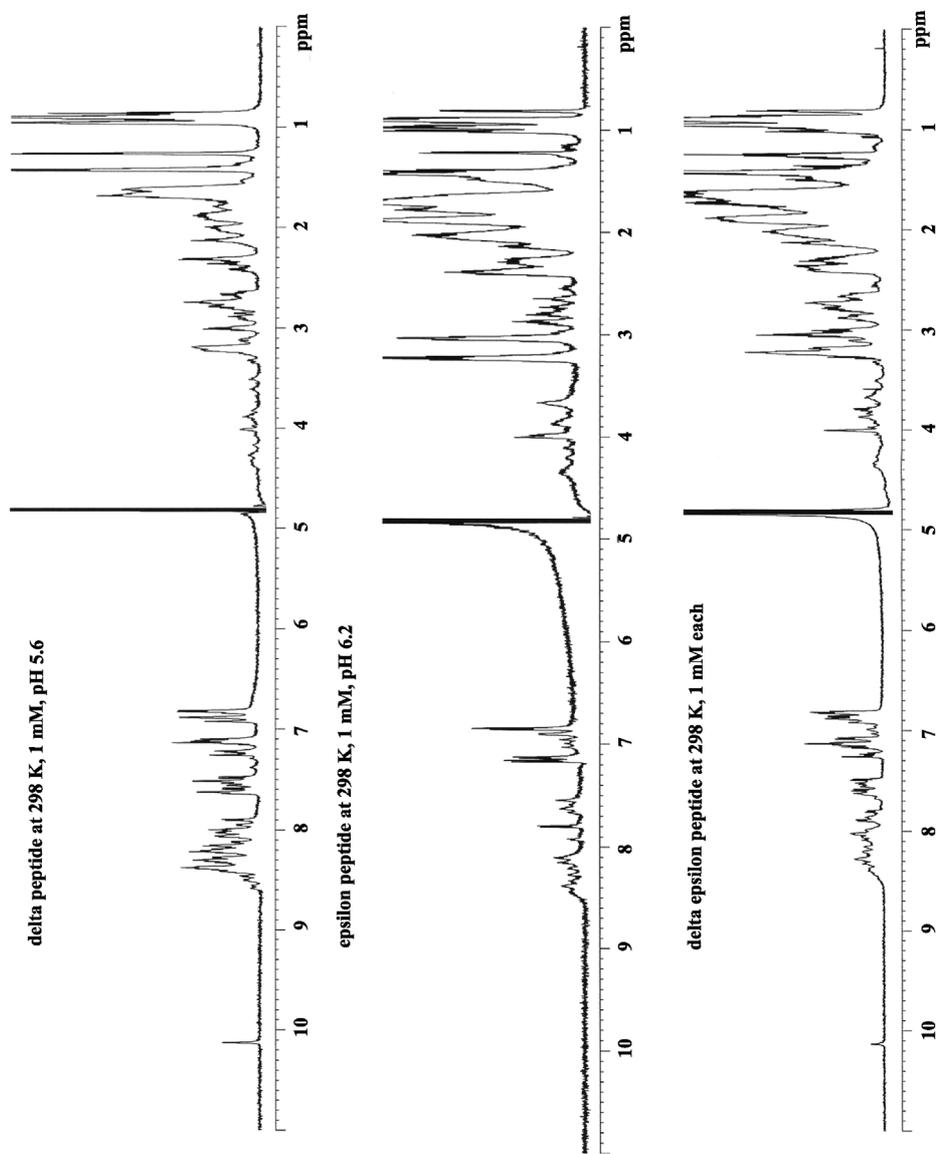
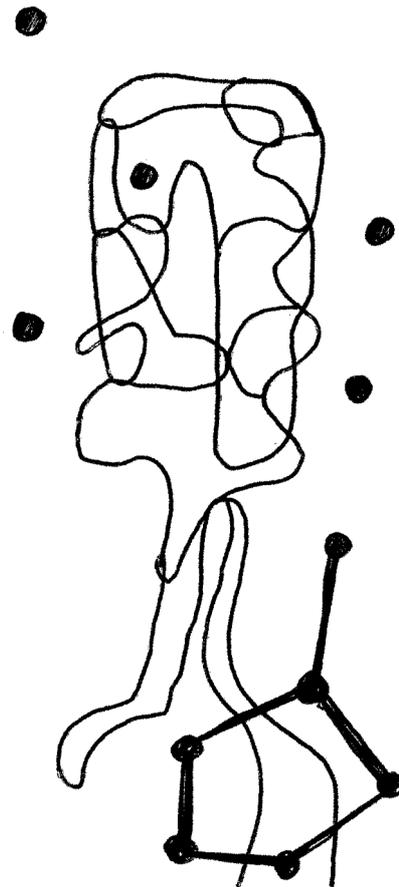


Figure 2.9. 1D proton NMR of CD3 δ and CD3 ϵ intracellular synthetic peptides and the CD3 $\delta\epsilon$ mix at 298K. Poor dispersion and narrow line width leading us to conclude that no interaction or folding could be measured in this NMR experiment.

Bacteriophage T4



3.1

The structure of the receptor-binding domain of the bacteriophage T4 short tail fibre reveals a knitted trimeric metal-binding fold

Ellen Thomassen, Gerrit Gielen, Michael Schütz, Guy Schoehn, Jan Pieter Abrahams, Stefan Miller and Mark J. van Raaij

Published in: *J. Mol. Biol.* (2003) **331**, 361-373

Summary

Adsorption of T4 bacteriophage to the *Escherichia coli* host cell is mediated by six long and six short tail fibres. After at least three long tail fibres have bound, short tail fibres extend and bind irreversibly to the core region of the host cell lipo-polysaccharide (LPS), serving as inextensible stays during penetration of the cell envelope by the tail tube. The short tail fibres consist of a parallel, in-register, trimer of gene product 12 (gp12). X-ray crystallography at 1.5 Å resolution of a protease-stable fragment of gp12 generated in the presence of zinc chloride reveals the structure of the C-terminal receptor-binding domain. It has a novel "knitted" fold, consisting of three extensively intertwined monomers. It reveals a metal-binding site, containing a zinc ion coordinated by six histidine residues in an octahedral conformation. We also suggest an LPS-binding region.

Introduction

Bacteriophage T4 is a very efficient DNA injection machine [1], consisting of a DNA-containing head, a double-tubed tail of which the outer tail-sheath is contractile, and a base-plate to which six long tail fibres and six short tail fibres are attached. The long tail fibres recognise the outer membrane protein C (OmpC) protein or LPS of *E. coli* B and are responsible for the initial, reversible, attachment of the virion. After at least three long tail fibres have bound, the baseplate changes conformation from the "hexagon" form to the "star" form [2]. In the hexagon form, the short tail fibres are incorporated in the baseplate in bent fashion, as shown by cryo-electron microscopy [3]. Upon conversion of the baseplate to the star form, the short tail fibres extend and bind irreversibly to the core region of the host cell LPS [4]. Here they form inextensible stays, allowing penetration of the cell envelope by the baseplate "hub" and tail tube upon contraction of the outer tail-sheath [3]. When gene product (gp12) is absent, baseplates spontaneously convert to the star form [5] and *T2* phages have the unusual property of contracting without releasing their DNA upon reversibly adsorbing to cells [6].

Short tail fibres are composed of a single protein, gp12 [7] forming a parallel, in-register, homo-trimer of 527 residues per subunit. They are attached to the baseplate *via* their N-termini, while the C-terminal globular domain has been proposed to bind to the bacterial host cell [8], although to our knowledge no mutation or deletion experiments have been reported to prove this. Correct folding into trimers of gp12 requires the chaperone protein gp57 [9]; gp12 trimers are incorporated into the baseplate as one of the last steps, together with gp9 [10]. Gp12 resists dissociation by sodium dodecylsulphate (SDS) at room temperature and monomerisation requires heating of the polypeptide chains [11].

Previous structural studies by electron microscopy [8] show gp12 to form an elongated trimeric complex with a slender shaft (24 nm x 4 nm) capped by an arrow-shaped, more globular domain (11 nm x 6 nm). Estimates of its mass by scanning transmission electron microscopy confirmed the trimeric nature of the protein. These studies also show that the gp12 trimer is frequently bent roughly halfway its length, a feature consistent with its folded-away state in the baseplate.

Previously a 33 kDa proteolytic fragment generated in the presence of EDTA [12] was crystallised. This fragment consisted of residues 85-396 and 518-527. However, in the crystal structure [13] residues 85-245 were invisible due to static disorder. The ordered

residues, 246-396 and 518-527, revealed three new folding motifs. Residues 246-289 form an N-terminal region, held together by intertwined strands and a central, mainly hydrophobic core. Residues 290-327 form a central right-handed triple-stranded β -helix, which is connected to the "collar" domain by three short α -helices (residues 333-341). The collar is formed by three small globular domains, each containing amino acid residues 343-396 and 518-527 and consisting of six β -strands and an α -helix in a hitherto unseen topology.

By omitting EDTA and adding divalent zinc ions during proteolysis we could ensure that amino acids 397-517 remained part of the complex upon limited proteolysis, resulting in a 45 kDa fragment. Here, we report the structure of this C-terminal domain at 1.5 Å resolution and show that it is the receptor-binding domain.

Results and Discussion

Preparation of a proteolytic 45 kDa fragment of the bacteriophage T4 short tail fibre

The gp12 full-length protein was co-expressed with its chaperone gp57 [9] and purified. The purified protein was incubated at 56 °C and subsequently treated with proteases in the presence of divalent zinc ions. SDS-PAGE (not shown) revealed that a fragment of around 45 kDa was obtained, instead of a 33 kDa fragment that is generated in the presence of EDTA [12]. N-terminal sequence analysis (results not shown) showed that both the 45 kDa and 33 kDa fragment start at residue Leu85. The 33 kDa fragment, however, lacks amino acid residues 397-517. Inclusion of these residues leads to a predicted mass of 47072 Da; therefore, we assumed that the 45 kDa fragment contains amino acid residues 85-527 of native gp12, without the internal deletion of residues 397-517 of the 33 kDa fragment.

Crystallisation and structure determination

Crystals of the 45 kDa fragment of gp12 with unit cell dimensions of 50.78 x 50.78 x 435.6 Å, grew under the same conditions as those of the 33 kDa fragment [12]. Both proteins crystallised in space group P321 with similar short *a* and *b* cell axes. However, the long *c* cell axis of the 45 kDa fragment is substantially larger, the crystal contacts are

different and the crystals are not isomorphous. The crystals contain one molecule in the asymmetric unit (i.e. one-third of a trimer), have a solvent content of 64 % (v/v) and a Wilson temperature factor of 14 Å².

Molecular replacement trials using the known structure of residues 246-396 plus 518-527 of the 33 kDa fragment (PDB-code 1H6W [13]) were not successful, so two datasets of a single-site mercury derivative were collected and the structure was determined by single isomorphous replacement using the anomalous signal. Both derivative crystals had virtually identical mercury content, and although the structure could most likely have been elucidated with just one derivative dataset, we decided to include both as they were of similar quality (see Table 3.1). The resulting map showed good electron density for residues 330-527 (see Figure 3.1), but not for residues 85-329. Nevertheless, we confirmed residues 85-329 to be present in the crystal by SDS-PAGE of crystals washed in mother liquor (results not shown). As we knew the structure of residues 246-334 from the crystal structure of the 33 kDa fragment of gp12, we included these in the model in a refinement trial. However, even adding just five extra residues to the N-terminus of the model led to lower agreement with the measured data. As none of the residues 85-329 could be refined satisfactorily, they were probably not well ordered in the crystal.

The Hg-derivatised amino acid was found to be Cys372. In the 33 kDa fragment both Cys362 and Cys372 were derivatised, but in the 45 kDa fragment solvent access to Cys362 appeared to be prevented by the extra domain consisting of residues 397-517. Having solved the structure, we repeated molecular replacement trials using just residues 330-397 and 518-527 from the 33 kDa fragment structure [13]. This led to the correct solution, so with hindsight the structure could also have been determined by molecular replacement, had we guessed that in crystals of the 45 kDa fragment residues up to 329 were disordered. The final model of the monomer contains 198 residues of 443 present in the asymmetric unit. The final model had good stereochemistry and low R-factors (Table 3.1) and contains 1560 protein atoms, a zinc ion on the threefold axis, two sulphate ions, one citric acid molecule and 512 water molecules in the asymmetric unit (Table 3.1).

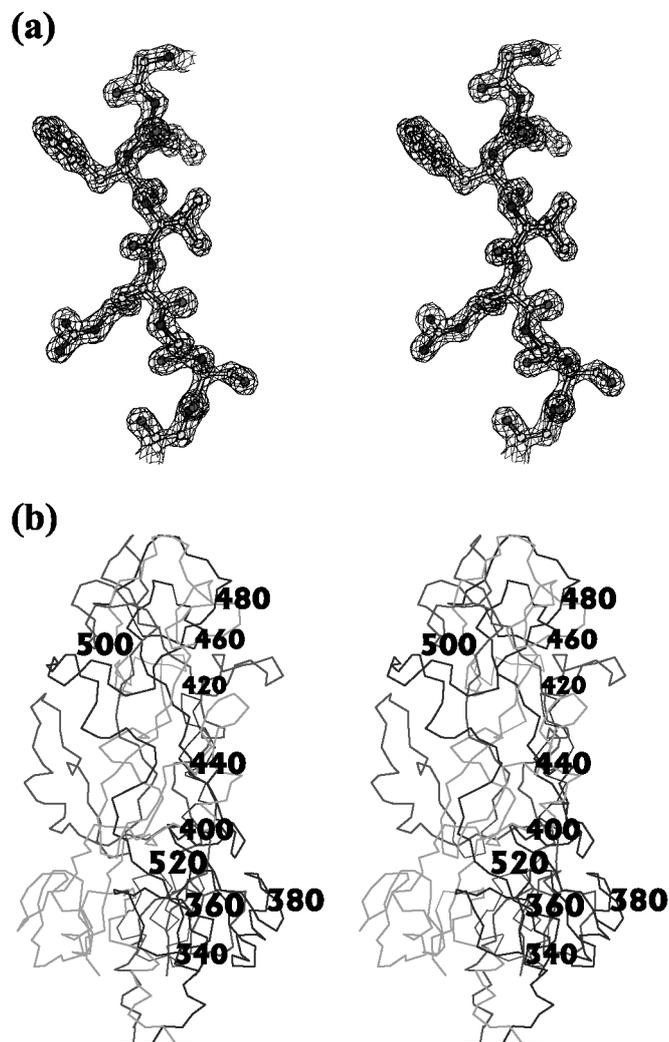


Figure 3.1. Structure of the ordered region of the proteolytic fragment of gp12 generated in the presence of zinc ions. (a) Stereo-view of part of the crystallographic electron density. Amino acids Gly397 (top), Leu398, Phe399, Val400, Arg401, Gly402, Ser403 and Gly404 (bottom) are shown in ball-and-stick representation surrounded by the final $2F_o - F_c$ electron density map contoured at 1.5σ . (b) Stereo-view of the structure. In the blue monomer, every 20th residue is labelled. The other monomers are coloured red en green. The N-terminal "neck" is at the bottom, the "collar" domain in the middle and the C-terminal receptor-binding domain at the top. Figures 1, 2, 4 and 6 were prepared with BOBSCRIPT [14], which is based on MOLSCRIPT [15]. For a colour picture see appendix A.

Table 3.1. Crystallographic data and refinement statistics

	Native	Derivative set1	Derivative set2
<i>A. Data collection</i>			
Beam-line	ESRF ID29	ESRF BM14	ESRF BM14
Wavelength (Å)	0.946	0.867	0.867
Detector	ADSC Q4R	130 mm MAR CCD	130 mm MAR CCD
Resolution range (Å)	15.0-1.5 (1.58-1.50) ^a	99-2.74 (2.83-2.74)	99-2.73 (2.83-2.73)
Multiplicity	6.6 (3.1)	5.6 (3.3)	6.1 (2.2)
Completeness (%)	89.8 (53.3) ^b	98.0 (90.4)	96.6 (78.0)
Rsym ^c (%)	10.6 (28.4)	8.6 (13.1)	9.9 (18.9)
<i>B. Phasing</i>			
Number of Hg-sites		1	1
Phasing power ^{d,f} (centric/acentric)		2.70 / 1.61	2.61 / 1.57
Rcullis ^{e,f} (centric/acentric)		0.53 / 0.57	0.54 / 0.59
<i>C. Refinement</i>			
Resolution range (Å)		15.0-1.50 (1.58-1.50)	
No. of reflections used in refinement		94341 (6339)	
No. of reflections used for R-free		1922 (116)	
R-factor ^g		0.143 (0.20)	
R-free		0.152 (0.22)	
No. of protein / water atoms		1560 / 512	
Average B-value protein / solvent (Å ²)		16.8 / 42.7	
Ramachandran statistics ^h (%)		84.9 / 15.1 / 0.0 / 0.0	
R.m.s. deviations ⁱ (bonds, Å / angles, °)		0.018 / 1.6	

^a Values in parentheses are for the highest resolution bin, where applicable.

^b Data in the higher resolution shell are less complete because of data collection on a square detector.

^c $R_{\text{sym}} = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i |I_{hi}|$, where I_{hi} is the intensity of the i^{th} measurement of the same reflection and $\langle I_h \rangle$ is the mean observed intensity for that reflection.

^d Phasing power = $(|F_H| / (|F_{PH}| - |F_P + F_H|))$, where F_H , F_{PH} and F_P are the heavy atom, derivative and native structure factors, respectively.

^e $R_{\text{cullis}} = (|F_{PH}| - |F_P + F_H|) / (|F_{PH}| + |F_P|)$.

^f Phasing statistics were produced with MLPHARE [16].

^g $R = \sum ||F_{\text{obs}}(hkl)| - |F_{\text{calc}}(hkl)|| / \sum |F_{\text{obs}}(hkl)|$.

^h According to the program PROCHECK [17]. The percentages are indicated of residues in the most favoured, additionally allowed, generously allowed and disallowed regions of the Ramachandran plot, respectively.

ⁱ Estimates provided by the program REFMAC [18].

Electron microscopy of the 45 kDa fragment of gp12

To further characterise the crystallised 45 kDa fragment, we performed electron microscopy and three-dimensional reconstruction on negatively stained protein from dissolved crystals (see Figure 3.2(a)) and compared the resulting three-dimensional maps to those of full-length protein and the 33 kDa fragment obtained before [12]. This showed that, compared with the full-length protein, the 45 kDa fragment only misses an N-terminal region, while the 33 kDa fragment misses an additional C-terminal domain (Figure 3.2). We did not attempt a very precise and high-resolution reconstruction as we already had the crystal structure. However, the reconstruction was of sufficient quality to assign the polarity of the fibre without ambiguity and to place the atomic model in the electron microscopy map, even if some pseudo-symmetrisation may have taken place along the long axis of the fibre.

Previously, the C-terminal domain was shown to have an arrow-shaped form [8], but in our electron microscopy reconstruction it has a more bi-lobed peanut shape. The bi-lobed shape is also observed in our previous reconstruction of full-length gp12 and is probably due to some internal flexibility, which smeared out the density at both ends of the particles. A space-filling representation of our current crystal structure shows a shape in between these two extremes: a bi-lobed shape in which the extreme lobe is more slender. The flexibility of the 45 kDa fragment appears to be more pronounced than that of the 33 kDa fragment and full-length protein, also leading to the disappearance of some ordered density at the N-terminal side compared to the 33 kDa fragment.

Bacterial cell binding of the 45 kDa and 33 kDa fragment

As it is generally assumed that the C-terminal domain is involved in LPS-binding, we investigated whether the 33 kDa and 45 kDa fragments behaved differently in binding to bacterial cells. To measure cell-binding activity, proteins were immobilised in microplate wells and were allowed to bind to bacteria. Unbound bacteria were washed off and bound bacteria were detected by their intrinsic beta-galactosidase activity (Figure 3.3). Full length gp12 and the 45 kDa fragment were both capable of binding to bacteria, while the 33 kDa fragment did not bind, showing that the receptor-binding domain was absent from the 33 kDa fragment but present in the 45 kDa fragment. Full-length gp12 appeared more active than the 45 kDa fragment in binding, which may be explained by the hydrophobic virus-binding N-terminus of full length gp12. The hydrophobic parts having

a higher affinity for the micro-plate wells may have led to more C-terminal domains being exposed to the solution.

In summary, the activity data indicated that the most important receptor-binding residues are contained in the C-terminal region removed upon proteolysis in the presence of EDTA but are retained upon proteolysis in the presence of zinc. This region, containing residues 397-517, are therefore referred to as the receptor-binding domain.

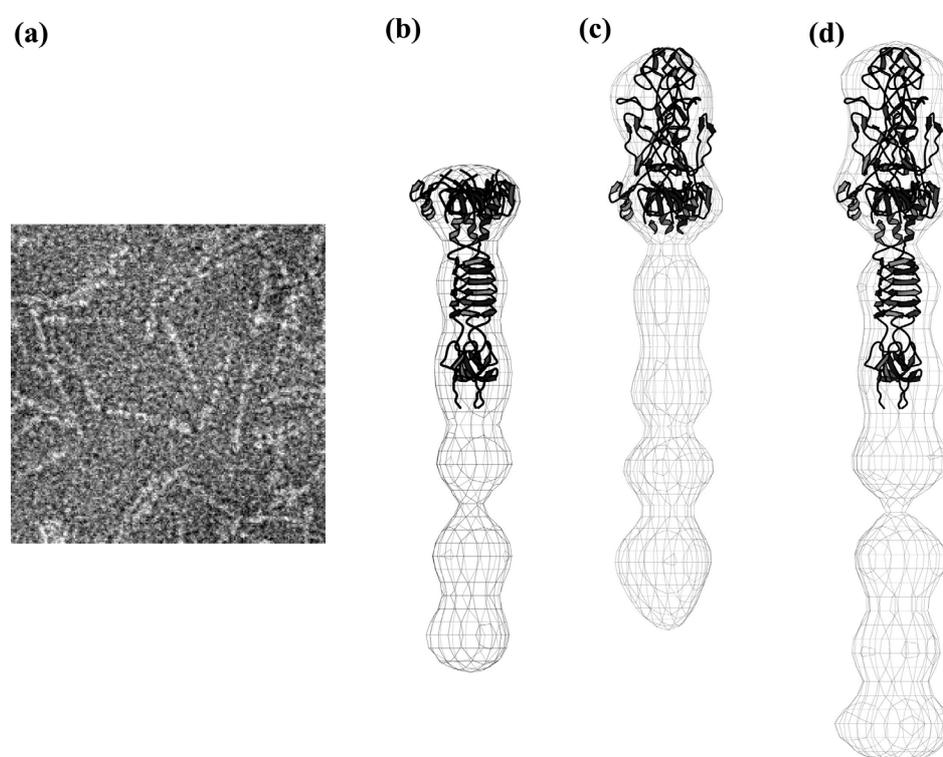


Figure 3.2. Crystal structures fitted into their corresponding electron microscopy maps. (a) Part of a field of negatively stained 45 kDa fragment fibres from a dissolved crystal. The size of the field is 75 x 75 nm. (b) Averaged three-dimensional negative stain electron microscopy map of the 33 kDa fragment with the structure of ordered region of the 33 kDa fragment fitted into it [9]. (c) Averaged three-dimensional negative stain electron microscopy map of the 45 kDa fragment with the structure of ordered region fitted into it (this work). (d) Averaged three-dimensional negative stain electron microscopy map of full-length gp12 with a composite structure of the ordered regions of the 33 and 45 kDa fragments fitted into it. The composite structure was constructed by superimposing residues 340-395 and 520-527, present in both the 33 kDa and 45 kDa fragments.

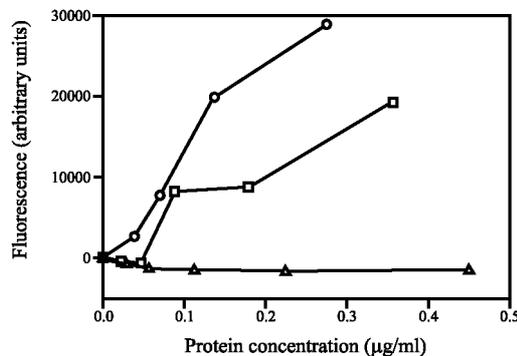


Figure 3.3. Assay of bacterial cell-binding of gp12 and its proteolytic fragments. Shown is fluorescence caused by hydrolysis of methylumbelliferyl-beta-galactoside due to *E. coli* bacteria bound to immobilised gp12 at the concentrations indicated on the x-axis. See Materials and Methods for further details. Circles, full-length gp12; triangles, 33 kDa fragment missing N-terminal and C-terminal domains and squares, 45 kDa fragments missing only N-terminal domains.

Overview of the structure

The structure can be divided into two parts, the part already present in the previously reported structure of the 33 kDa fragment [13], and the previously unresolved receptor-binding domain (Figures 3.2(c) and 3.4). The first can be further sub-divided into a "neck" and "collar", while the latter can be further sub-divided into a "head" and "bonnet" (see below). We refer to the triple alpha-helix of residues 333-341 as the neck, connecting the body of the fibre (unresolved in this structure) to its C-terminal collar and receptor-binding domains. Amino acid residues 342-396 plus 518-527 form the collar, whilst residues 397-517 form the extensively intertwined receptor-binding domain (Figure 3.4).

The surface area of the monomer in our structure is 16800 Å², of which 9900 Å² (58 %) is buried in the trimer. This large proportion of buried surface explained the extraordinary stability of the trimer, which does not dissociate into monomers in SDS-PAGE without boiling (for more information on the stability of gp12 see ref. 14).

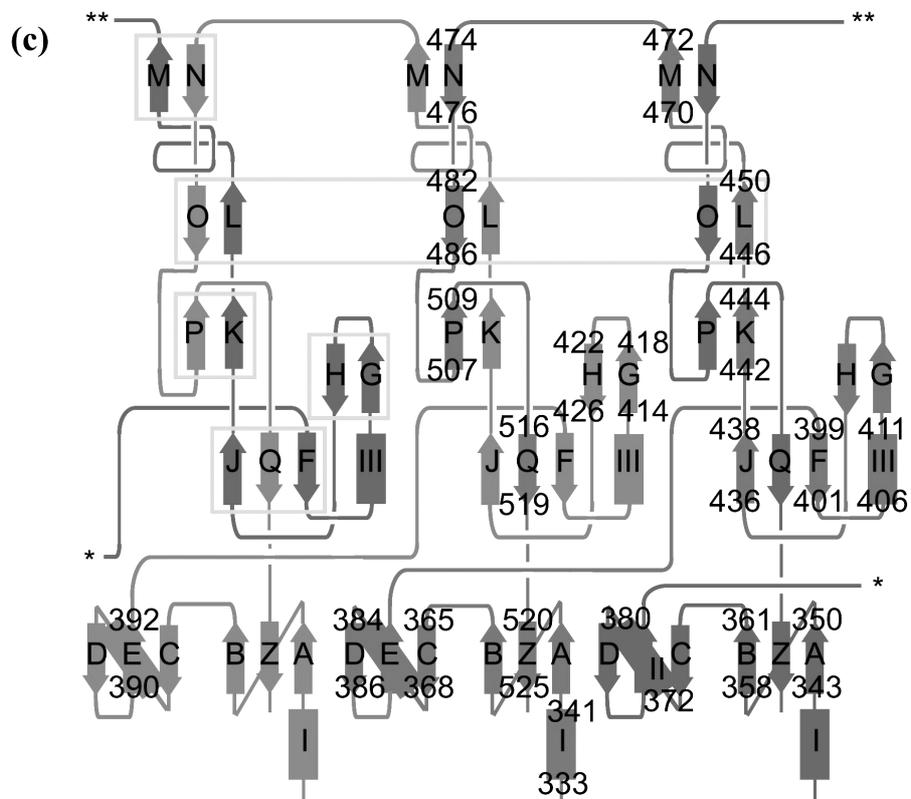
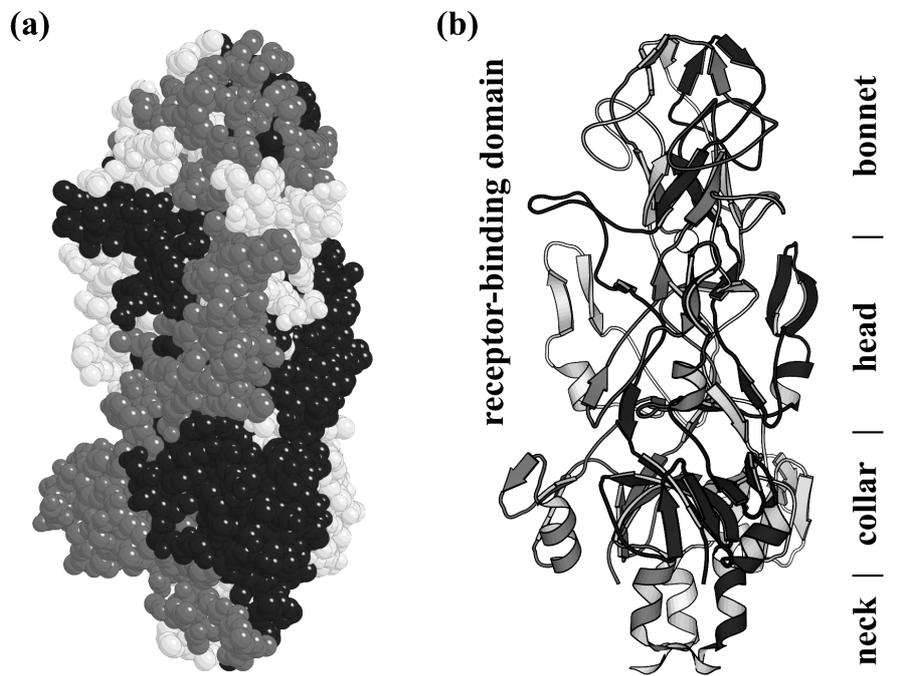
The collar domain

The collar domain is a small globular domain containing six beta-strands and an alpha-helix (bottom of Figure 3.4(b)) with a weak structural homology with bacteriophage T4 gp11 [19] as described before [13]. The six beta-strands form a sandwich of two three-

stranded anti-parallel beta-sheets (bottom of Figure 3.4(c)). When the main-chain atoms of the collar domain (residues 342-395 and 520-527) from the current structure were superimposed on those of the smaller 33 kDa fragment, they overlapped very well with a root-mean-square displacement of 0.5 Å for the monomer and 0.7 Å for the trimer. Differences of more than one Å were limited to a few residues at the top of the collar. Presumably, the absence of the receptor-binding domain in the smaller fragment was the cause of these differences. Of each monomer in the collar domain, a relatively modest 1200 Å² (27%) of its surface area is buried. The collar domain is the only domain in which the three monomers are not intertwined. This observation suggests that folding of the T4 short tail fibre may start with this domain (see below).

Legend to Figure 3.4 which is on the next page

Figure 3.4. Structure of the ordered region of the 45 kDa fragment. **(a)** Space-filling model of the structure illustrating the extensive intertwining in the receptor-binding domain. **(b)** Ribbon diagram in the same orientation as (a). At the bottom the neck region (the alpha-helical residues 336-339), in the middle the collar region (residues 340-396 plus 518-527) and at the top the metal-containing receptor-binding domain (residues 397-517), divided into head and bonnet as described in the text. **(c)** Topology diagram. Alpha-helices are marked I, II and III, beta-strands A to Q (excluding I) and Z. Begin and end residues of the secondary structure elements are numbered in one of the monomers (not always the same one to avoid overlaps). Grey boxes surround the six-stranded beta-barrel domain and the other, smaller beta-sheets. In the three-dimensional structure, the blue loops marked with a * connect, as do those marked **. With a small thought experiment the reader can convince him- or herself that the structure is knitted rather than knotted: move the green JKL-strands to the left under the red OP-strands, then lift to the right over the red OP-strands. Do this for all three monomers and the "knit" is resolved. For colour picture see appendix A.



The receptor-binding domain: head and bonnet

The receptor-binding domain as a whole can be compared to a flower bud with 12 petals in 3-fold symmetry. At the bottom, residues 406-432 form the first petal, just above that residues 489-504 form a second petal, above that residues 450-470 a third and at the top a fourth petal consisting of residues 470-480 is present. The three JQF beta-sheets (Figure 3.4(b) and (c)) form a base and the rest of the amino acids are all in more or less extended conformation at the centre of the trimer (Figure 3.4). In the receptor-binding domain, the strands are knitted together in a highly complex fashion and 7400 Å² (60%) of the surface area of the monomer is buried in the trimer. The receptor-binding domain can be sub-divided into "head" (residues 397-446 and 487-517) and "bonnet" sub-domains (residues 447-486), with a metal-binding site on the border between the two.

The receptor-binding domain has hardly any regular secondary structure elements like alpha-helices or beta-strands, despite 90% of its main chain torsion angles being in most favoured regions of the Ramachandran plot and none being in the generously allowed or disallowed regions (according to PROCHECK [17]). Many inter-domain hydrogen bonds (main-chain - main-chain, main-chain - side-chain and side-chain - side-chain) stabilise the trimeric protein structure. At the top of the head domain, in the central region of the trimer, a metal ion, presumably zinc, is octahedrally coordinated by Nε2 of six histidine residues (His445 and His447 from each monomer).

Residues 397-403 form an extended strand transverse to the fibre axis, contributing Leu398, Phe399 and Val400 to the hydrophobic core and Arg401 and Ser403 to the outside of the molecule. The strand is followed by a long "loop" (404-432), buffeting against the core of the molecule (the first petal of the flower bud). In some respects the loop resembles a beta-hairpin, but regular beta-strand contacts are rare and there are deviations and inserted structures, like a distorted helix (amino acids 406-414). N404 and O404 form hydrogen bonds with O432 and N431, respectively, while hydrogen bonds are also present between O414 and N426, O416 and N424 and N418 and O422. Residues 418 to 421 form a non-classical reverse turn. At the end of this petal is another reverse turn, with Gly435 at position $i + 3$. This turn is followed by an extended region going upwards, consisting of residues 436-450 and containing the metal-binding histidine residues 445 and 447 (see below).

Just above the metal-binding site, residues 447-450 together with 482-486 form a highly distorted anti-parallel beta-barrel-like structure (the OPOPOP-sheet) in the trimer;

and together with the connecting regions (451-481) they form the bonnet. Main-chain - main-chain hydrogen bonds are found between the main-chain oxygen of residue 447 and the main-chain nitrogen of residue 484 of a second subunit, between N449 and O482 of the same second subunit and between O448 and N485 of a third subunit and between N450 and O483 of the same third subunit. The residues above this structure are knotted together, but because they are de-knotted again in the domain underneath, overall the trimer is knitted rather than knotted (see Figure 3.4(c)). Residues 450-470 form a loop capped by a reverse turn (458-461), encircling residues 478-484 of a neighbouring monomer.

Residues 470-472 form a small two-stranded beta-sheet with residues 474-476 of a neighbouring monomer. Inter-chain hydrogen bonds are present between N470 and O476, O470 and N474 and N472 and O474. The residues in between, Arg472 and Lys473, form short loops at the top of the molecule, making the top positively charged (see below). Residues 480-488 form an extended strand downwards at an angle of around 45°, traversing near the centre of the trimer (Thr484), culminating in the second petal, a loop of residues 489-504. Residues 492-496 form a small distorted helix in this loop. The following residues (506-517) meander downwards towards the collar domain, encircling amino acids 439-443 of a neighbouring monomer.

The central trimer axis is mainly hydrophobic, in the collar domain with contributions from Val331, Val332, Ile341, Met347 and Met348 (from bottom to top). In the receptor-binding domain amino acids projecting side chains towards the central trimer axis are Phe399, Pro514, Thr512, His445, His447 (the latter two forming the metal-binding site), Thr484, Phe451, Thr471 and Lys472, again from bottom to top.

The LPS-binding site

The binding site for the LPS core region is currently unknown, although our proteolysis experiments show that it is contained in the domain consisting of residues 397-517. Based on the chemical nature of LPS core, we can make some suggestions. Figure 3.5(a) shows the primary structure of the LPS core region, consisting of sugar residues with attached phosphate groups. We can assume phosphate groups bind to positive charges on the protein surface while the sugars bind to aromatic side-chains [20-23].

Figure 3.5(b) shows the electronic surface potential of the protein, while in 5(c) the aromatic side-chains are mapped on the surface. The trimer in general, and its top in particular, is positively charged, which could contribute to its affinity for the negatively charged *E. coli* membrane. Symmetrically along the sides of the trimer there are three grooves which are flanked by, or contain positively charged residues and aromatic amino acids. Two solvent molecules, a citric acid and sulphate ion, form salt bridges with Arg465 and Arg464, respectively, perhaps mimicking LPS core phosphate groups. We suggest that each of these grooves may bind an LPS molecule, analogous to the adenovirus fibre binding to three receptor molecules at once [24]. This would lead to very tight binding and a very low off-rate, especially if we consider that six short fibres can bind simultaneously to the *E. coli* membrane, each through three receptor binding sites. This also would fit to the observation that binding of the short fibres to the bacterial cell in the second host recognition step renders the modus from a reversible to an irreversible (or at least quasi-irreversible) attachment of the phage [1]. The sequence of the homologous bacteriophage T4-like strain AR1 [25] has 64 % overall and 63 % receptor-binding domain identity and provides additional clues. Assuming the short tail fibre of this phage also binds to the same LPS core molecule, we could eliminate some residues from consideration, as their basic or aromatic character is not conserved in AR1 gp12. This would leave the basic putative LPS-binding residues Lys446, Lys422, Arg504, Arg424, Arg513 and the aromatics Tyr454, Phe451, Trp477, Phe468, Phe460, Phe420, Tyr488, Tyr444, His408 and Tyr433. The basic residues cluster in a region behind the bottom petal, suggesting it may function as a kind of LPS clamp. The putative aromatic receptor-binding residues are more scattered (Figure 3.5(b) and (c)).

Of course, we cannot rule out other amino acids being important for binding, a conformational change taking place upon LPS core binding or an alternate mode of receptor binding by the short tail fibre of strain AR1. Nevertheless, the suggested residues are a starting point for site-directed mutagenesis and / or other biochemical experiments to elucidate the location of the receptor-binding site. We are pursuing crystal soaking and co-crystallisation experiments with LPS core analogues to obtain structural information, but so far these have not yielded useful data (Chapter 3.2).

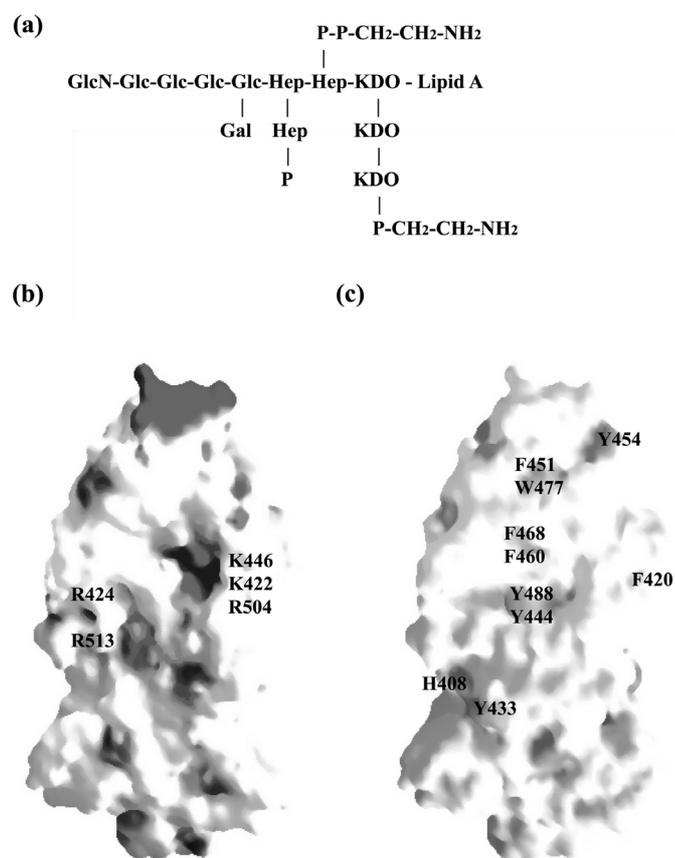


Figure 3.5. Possible LPS binding residues. (a) Primary structure of the LPS of *E. coli* B core region plus lipid A. Glc stands for glucose, GlcN for glucosamine, Hep for heptose, P for ortho-phosphate and KDO for 2-keto-3-deoxyoctonate. (b) Electronic surface potential diagram of the trimer. Positive charges are marked in blue, negative in red. Putative receptor binding basic amino acids (see text) are labelled. Please note that although the five labelled residues appear to cluster in two groups, they are in fact all five very close to each other due to the 3-fold symmetry. (c) Surface diagram showing in green aromatic side-chains that may be involved in LPS binding. Labels identify the amino acids. Figures 5 (b) and (c) were made with the program GRASP [26] with the protein in the same orientation for both. The protein is shown tilted forwards to afford a better view of the top of the trimer. Residues are labelled by their one-letter amino acid code. For colour picture see appendix A.

The metal

In the centre of the receptor-binding domain, on the 3-fold axis, we found a metal ion, octahedrally coordinated by the Nε2 of His445 and His447 from each monomer (Figure 3.6). An X-ray fluorescence scan of the crystal used for data collection at beam-line ID29 of the ESRF confirmed the metal ion to be zinc, which was anticipated, as zinc chloride was used in the preparation of the 45 kDa proteolytic fragment (see Material and Methods). The zinc ion inside the gp12 trimer confirms data of Zorzopulos & Kozloff [27], indicating gp12 to be a zinc metallo-protein. They estimated gp12 to contain one zinc atom per monomer. In our structure, we found one zinc atom per trimer (although we cannot exclude other zinc binding sites in the parts of gp12 missing from our structure). Further experiments are required to determine the gp12-zinc stoichiometry and to show whether other metals can substitute for zinc in the protein. Preliminary results suggested that divalent calcium, manganese and copper ions can substitute for zinc ions in the proteolysis experiments to obtain the 45 kDa fragment, while magnesium ions led to a 33 kDa fragment similar to the one obtained with EDTA (Thomassen *et al.*, unpublished results).

The Zn ion is well-ordered with a temperature factor of 7 \AA^2 and also the coordinating residues are present in well-defined conformations with temperature factors of between 6 and 8 \AA^2 . The refined distances of the zinc ion to the Nε2 of His445 and His447 are 2.22 and 2.25 Å, respectively. These distances are larger than found normally for zinc coordination by histidine [28]. A possible explanation lies in the octahedral coordination of the zinc in our structure. In protein structures solved so far, zinc is coordinated tetrahedrally [28]. The His445 Nε2 - zinc - His445 Nε2 angles are around 90.5° , while His447 Nε2 - zinc - His447 Nε2 angles are around 89° . The His445 Nε2 - zinc - His445 Nε2 angles are around 87, 94 and 175° , i.e. very close to perfect octahedral coordination. There is no evidence for alternate conformations of the zinc ion or its ligands in the electron density.

As the bacteriophage T4 short tail fibre does not appear to have enzymatic activity, unlike the P22 tailspike [29], the role of the zinc ion is probably purely structural. Zinc clearly increases the stability of the C-terminal part of gp12 against proteases, but may also increase the stability of the C terminus in general.

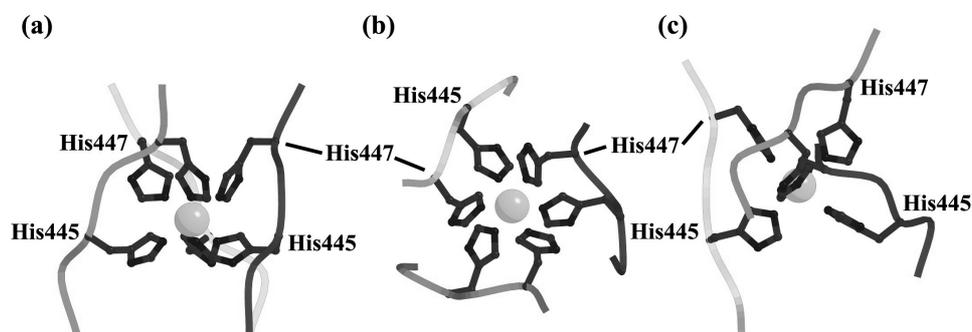


Figure 3.6. The zinc ion in the centre of the receptor-binding domain. Shown are the main chains of the three monomers (residues 443-448 in each case) in yellow, red and green and the side-chains of the ligating histidine residues. The coordination is octahedral with a Zn-His445 Ne2 distance of 2.22 Å and a Zn-His447 Ne2 distance of 2.25 Å. For angles see the text. **(a)** Side view. **(b)** View from the top down the three-fold axis. **(c)** View down one of the four-fold axes of the octahedron formed by the Ne2 atoms of the six ligating histidine residues. For colour picture see Appendix A.

Comparison with known protein structures

A search of the protein structure database using the program DALI [30] with the structure of amino acids 397-517 did not show up any related structure, while the collar domain has some limited structural homology with the central domain of T4 gp11 [13]. There are, however, some proteins of unknown structure which show limited sequence similarity. A BLAST search [31] using residues 330-527 found the collar region to be a putative conserved domain in a microcystin dependent protein (mdpB), of which the function is not known [32]. We also discovered similarities with several lambda prophage side tail fibre protein homologues from the *E. coli* genome [33] and the *Rhizobium* protein RhiB, a protein proposed to be involved in interactions of *Rhizobium* with its plant host [34]. In all cases most homology was with the collar domain, although in the case of RhiB a putative His-Xaa-His (His-Ala-His in this case) metal-binding "motif" could be identified C-terminal to the homologous region. As the functions of these proteins remain to be fully elucidated, the possibility that they are just inserted phage genes cannot be ruled out.

Homology was also detected with the bacteriophage long tail fibre protein gp37 [35], which, apart from binding to OmpC, also binds to *E. coli* B LPS [1]. Again, most homology was found within the collar domain, with several His-Xaa-His motifs C-terminal to the homologous sequence. Further homologues that turned up were fibre

proteins from other bacteriophages; some of these proteins also contained His-Xaa-His sequences C-terminal to their putative collar domains. Further biochemical and structural experiments will reveal whether these homologues also contain metal ions.

The disordered region

The inability to visualise residues 85-329 remained puzzling, but was also observed for the smaller 33 kDa fragment [13], in which residues 85-245 are invisible. Furthermore, adding the known structure of residues 246-329 to the model led to a worse model, indicating the region was really disordered, or, but this we consider unlikely, has a different structure in the larger fragment. The nature of the disorder is unknown, but the 3-fold symmetry of the protein suggests that perhaps crystal contacts with the next layer may be made in three different, symmetrically equivalent, ways. As the disordered, fibrous region of the structure is thinner than the globular C-terminal collar and receptor-binding domains (see Figure 3.2), there may be enough space within the crystal lattice for these contacts to be distributed randomly around the three possibilities, smearing out crystallographic electron density. To visualise the fibrous domains at high resolution, it may be necessary to crystallise artificial constructs in which the globular domains are replaced by a smaller trimerisation domain like the bacteriophage T4 fibritin "foldon" [36].

Folding of the C-terminal domain of gp12

From the structure, it seems probable that the collar domain may act as a folding nucleus, as it is the only domain in which the three monomers do not intertwine. Upon trimerisation of this domain, the remaining C-terminal receptor-binding domain and the N-terminal domains can then intertwine. As the zinc ion occupies a central position in the receptor-binding domain, it seems likely zinc (or other metal) ions play a role in its folding, although successful refolding experiments were also done in presence of 1 mM EDTA [9]. Perhaps gp12 competed effectively with EDTA for zinc. *In vivo*, the molecular chaperone gp57 is necessary for folding of gp12 (as well as for the long tail fibre proteins gp34 and gp37), although the mechanism for this chaperone action is not known. Data from refolding experiments suggests that addition of gp57 suppresses unproductive side-reactions of refolding gp12. Based on the structural data presented here, we would suggest

gp57 may act in keeping the unfolded gp12 monomers apart until the collar domains trimerise, preventing aggregation.

Domain swapping in the receptor-binding domain?

The intertwining or knitting of the receptor-binding domain (residues 397-517) could also be considered as a particular case of three-dimensional "domain swapping" (reviewed by Rousseau *et al.* [37] and Liu *et al.* [38]), in which residues 399-472, containing beta-strands F-M and alpha-helix III have been "swapped" to the neighbouring monomer (or, which is equivalent, where residues 474-519, containing beta-strands N-Q, have swapped to the neighbour on the other side), leading to cyclic trimerisation.

However, we do not know of a homologous protein where the receptor-binding domain is a monomer, or of a homologous bacteriophage short tail fibre where the receptor-binding domain is not intertwined. Also, the extent of inter-monomer contacts in the receptor-binding domain and in the N-terminal and triple-stranded beta-helix domains makes it unlikely that the monomer exists other than as a very unstable folding intermediate. Therefore, we do not know if the domain evolved as an intertwined domain from its origins or whether at some time during its evolution three-dimensional domain swapping took place. In this respect it is interesting to note that the collar domain shows some limited structural homology to a domain of gp11 [19] (explained by van Raaij *et al.* [13]), the protein to which it binds in the baseplate, raising the possibility that gp12 evolved from gp11, which does not show domain swapping.

Perspectives

The structure presented in this thesis is an important step towards understanding the molecular details of receptor-binding of bacteriophage T4 and suggests experiments to determine the receptor-binding site and mechanism. It also reveals a surprising new fold. The as yet unsolved structures of the bacteriophage T4 long tail fibre proteins and fibres of other viruses may turn up more surprises.

Material and Methods

Expression and purification of gp12

For expression of full length gp12, the plasmid pT4g57g12 was transformed into *E. coli* strain JM109(DE3) (Promega) and cultures (total volume 4 litre) were grown to $A_{600\text{ nm}} = 0.6-1.0$. Cultures were cooled to 22 °C or below, expression was induced by addition of 1 mM IPTG and carried out overnight at 16 °C. Cells were harvested by centrifugation, resuspended in 100 ml of cold TE-buffer (40 mM Tris-HCl pH 8.0, 10 mM EDTA) containing 0.2 M ammonium sulphate and lysed using a French press. After centrifugation, ammonium sulphate was added to the soluble fraction to 0.8 M and some impurities were pelleted by centrifugation.

The supernatant was brought to 1.4 M ammonium sulphate, precipitating gp12. After centrifugation, the gp12 precipitate was re-dissolved in 50 ml of TE-buffer containing 0.8 M ammonium sulphate and loaded onto a 75 ml high substitution phenyl Sepharose column (Pharmacia), which was equilibrated with the same buffer. The protein was eluted with a gradient of 0.8 M to 0 M ammonium sulphate in TE-buffer, gp12 elutes at around 200 mM ammonium sulphate. The gp12 containing fractions were pooled (total volume 120 ml), ammonium sulphate was added to a final concentration of 1.8 M, precipitating gp12. The precipitated gp12 was stored as a suspension at 4 °C.

Proteolysis experiments

The precipitated gp12 in 1.8 M ammonium sulphate was pelleted by centrifugation and re-dissolved in 20 ml of HN-buffer [10 mM N-(2-hydroxyethyl)piperazine-N'-(2-ethane) sulphonic acid-sodium hydroxide, 150 mM sodium chloride, pH 7.4] containing 10 mM zinc chloride. The sample was incubated for 30 minutes at 56 °C and cooled to 37 °C, 10 mg trypsin and 10 mg chymotrypsin powder (Sigma) were added and the sample was incubated for 20 minutes at 37 °C. The heat and trypsin/chymotrypsin treatment was then repeated. The sample was purified by gel-filtration on a Pharmacia Sephacryl S200 column which was equilibrated with TE-buffer containing 200 mM sodium chloride, the fractions that contained the 45 kDa fragment were pooled (total volume 37 ml), precipitated with ammonium sulphate (final concentration 1.8 M) and stored at 4 °C.

Electron microscopy

Electron microscopy of the 33 kDa fragment and full-length gp12 has been described [12]. Micrographs of 45 kDa fragment from crystals dissolved in HN-buffer were taken the same way; the protein at a concentration of about 0.1 mg/ml was negatively stained with sodium silicotungstate at pH 7.0. The bending of the 45 kDa fibre showed more variation than for the 33 kDa fragment or the full-length gp12. The three best images of the sample were digitised on Zeiss scanner (Photoscan TD) with a pixel size of 14 μm (3.5 \AA at the sample scale). Six hundred particles were manually selected from the corresponding images and boxed into 128 x 128 pixels square boxes using Ximdisp [16]. All particles were band-pass filtered between 200 and 25 \AA without contrast transfer function correction and then normalised to the same mean and standard deviation. Processing was done using the SPIDER image-processing package [39]. We generated a starting model by back-projecting one fibre and by applying the known three-fold symmetry along the axis of the fibre. This model was re-projected along the axis perpendicular to the three-fold axis every 15 degrees covering the asymmetric unit (120°). The raw images were aligned against the re-projections by cross-correlation and then averaged inside their corresponding image classes. The eight class averages were then used to calculate a new three-dimensional reconstruction of the fibre by back projection. After 20 cycles of this procedure, the reconstruction parameters were stable. The resolution of the final reconstruction was estimated by Fourier shell correlation to be around 45 \AA .

Crystal structures were fitted into the averaged electron microscopy maps using the fast Fourier transform accelerated six-dimensional exhaustive search module (COLORES) of the program SITUS [40]; default values for all parameters were used.

Bacterial cell binding assay

To generate the 33 kDa fragment for cell-binding analysis, limited proteolysis of gp12 (1 ml) was performed in the presence of 10 mM EDTA as described [12] at 37 °C after partial unfolding at 56 °C by adding trypsin (80 μl of a 10 mg/ml solution). After a second trypsin addition to achieve homogenous proteolysis, a sample of the reaction mixture was analysed using SDS-PAGE. The 45 kDa fragment was generated the same way, but in presence of 10 mM zinc chloride and absence of EDTA.

The 33 kDa and 45 kDa fragments were subsequently purified by gel filtration chromatography on a Superdex 75 preparation grade matrix (Amersham Biosciences, Uppsala, Sweden) using PBS-buffer (10 mM sodium phosphate pH 7.4; 150 mM sodium chloride). Fractions containing the 33 kDa and 45 kDa fragments were analysed using SDS-PAGE, pooled and concentrated by ammonium sulphate precipitation as described above.

To analyse the bacteria-binding capabilities of the 33 kDa and 45 kDa fragments in comparison to native gp12, a cell binding assay was performed. Various amounts of the proteins in 100 µl PBS-buffer were immobilised in wells of micro-plates (Nunc Maxisorb Fluorescence, Nunc, Roskilde, Denmark) by passive adsorption for 12 h at 37 °C. Excess material was removed by washing the wells three times with 200 µl of PBS. Bound protein was measured by anti-p12 rabbit antibody (kindly provided by York-Dieter Stierhof and Ulf Henning, Max-Planck-Institut für Biologie, Tübingen) and anti-rabbit IgG-alkaline phosphatase conjugate (Sigma, St. Louis MO, USA). For cell-binding assays, wells were incubated with 100 µl of a tenfold diluted, IPTG-induced stationary culture of *E. coli* D21. After incubation for one hour at 25 °C unbound bacteria were removed by washing three times with 200 µl PBS supplemented with 0.5 % Tween 20 (Sigma, St. Louis MO, USA). Bound bacteria were detected by measuring their intrinsic beta-galactosidase activity with the fluorescence substrate methylumbelliferyl-beta-galactoside (Sigma, St. Louis MO, USA) as described by Davies *et al.* [41].

Crystallisation

The ammonium sulphate-precipitated 45 kDa fragment was re-dissolved in HN-buffer to a final concentration of 10-20 mg/ml. Protein concentrations were determined using UV absorption measurements at 280 nm; the absorption of a 1 mg/ml solution of gp12 or its proteolytic 45 kDa fragment was assumed to be 1.0 at a path-length of 1 cm. Crystals of the fragment were obtained by vapour diffusion against reservoirs containing 15-35 % (v/v) tertiary butanol buffered with 100 mM sodium citrate at pH 5.6 and in the presence of 0-10 % (v/v) glycerol. Single crystals of up to 0.1 x 0.1 x 0.5 mm were obtained. Crystals were transferred to 10 % (v/v) glycerol in crystallisation buffer for data collection if necessary.

Crystallographic data collection and processing.

Native data were collected on beam-line ID29 at a wavelength of 0.946 Å at the European Synchrotron Radiation Facility (ESRF). The crystal was flash-frozen and kept at 100 K during data collection. The crystal had a mosaic spread of 0.255°. The total rotation angle was 180° (360 images of 0.5°). Reflections were integrated with the programme MOSFLM [42] and processed using programs of the CCP4 suite [43]. Measurements differing more than 6 standard deviations from the weighted mean were rejected (180 of 629124 measurements or 0.03 %). For data statistics see Table 3.1. For collection of derivative data, crystals were soaked overnight in methylmercury acetate. Datasets to 2.8 Å resolution of two crystals soaked in methylmercury chloride were measured at the BM14-UK MAD beam-line at the ESRF at a wavelength of 0.867 Å (Table 3.1). Data of the derivative data was processed using HKL2000 [44]; the statistics can also be found in Table 3.1.

Structure solution and refinement

One mercury site was identified by the SOLVE programme [45], using data between 20 and 2.8 Å. The position, occupancy and temperature factor of the sites were refined and phasing was done also using SOLVE. Solvent flattening and automated building of a partial model proceeded using RESOLVE [46] and data between 20 Å and 2.2 Å. When input into the ARP-WARP auto-trace mode [47] using data to 1.5 Å resolution, a model resulted containing residues 335 to 526. Rebuilding of the model and addition of extra amino acid residues (330-334 and 527) was done with O [48]. Refinement was done using the REFMAC programme [18]. Water molecules were built using ARP [49].

Coordinates. Coordinates and structure factors have been deposited in the Protein Data Bank (accession codes 1OCY and R1OCYSF, respectively).

Acknowledgements

We thank A. Coetsee for help with derivative data collection and processing. We are grateful to R. Amons for N-terminal sequence analysis, R. Ravelli, M. Walsh, E. Mitchell and other members of the EMBL-ESRF Joint Structural Biology Group for

providing crystallographic data collection facilities and help therewith, and to R.A.G. de Graaff and N.S. Pannu for discussions. M.J.v.R. was supported by an “Akademie-onderzoeker” fellowship of the Dutch Royal Academy of Science. Finally, we thank the anonymous referees for helpful comments.

References

1. Goldberg, E., Grinius, J. and Letellier, L. (1994). Recognition, attachment and injection. *In: Molecular biology of bacteriophage T4 (ed. Karam, J.D.) American Society for Microbiology, Washington DC*, 347-356.
2. Crowther, R.A., Lenk, E.V., Kikuchi, Y. and King, J. (1977). Molecular reorganization in the hexagon to star transition of the baseplate of bacteriophage T4. *J. Mol. Biol.* **116**, 489-523.
3. Kanamaru, S., Leiman, P.G., Kostyuchenko, V.A., Chipman, P.R., Mesyanzhinov, V.V., Arisaka, F. *et al* (2002). Structure of the cell-puncturing device of bacteriophage T4. *Nature* **415**, 553-557.
4. Riede, I. (1987). Receptor specificity of the short tail fibres (gp12) of T-even type *Escherichia coli* phages. *Mol. Gen. Genet.* **206**, 110-115.
5. Berget, P.B. and King, J. (1983). T4 tail morphogenesis. *In: Bacteriophage T4 (C.K. Mathews, E. M. Kutter, G. Mosig and P.B. Berget, eds.) American Society for Microbiology, Washington, DC*, 246-258.
6. King, J. (1968). Assembly of the tail of bacteriophage T4. *J. Mol. Biol.* **32**, 231-262.
7. Mason, W.S. and Haselkorn, R. (1972). Product of T4 gene 12. *J. Mol. Biol.* **66**, 445-469.
8. Makhov, A.M., Trus, B.L., Conway, J.F., Simon, M.N., Zurabishvili, T.G., Mesyanzhinov, V.V. *et al* (1993). The short tail-fiber of bacteriophage T4: molecular structure and a mechanism for its conformational transition. *Virology* **194**, 117-127.
9. Burda, M.R. and Miller, S. (1999). Folding of coliphage T4 short tail fiber in vitro. Analysing the role of a bacteriophage-encoded chaperone. *Eur. J. Biochem.* **265**, 771-778.
10. Eiserling, F.A. and Black, L.W. (1994). Pathways in T4 morphogenesis. *In: Molecular biology of bacteriophage T4 (ed. Karam, J.D.) American Society for Microbiology, Washington DC*, 209-212.
11. King, J. and Laemmli, U.K. (1971). Polypeptides of the tail fibres of bacteriophage T4. *J. Mol. Biol.* **62**, 465-477.
12. van Raaij, M.J., Schoehn, G., Jaquinod, M., Ashman, K., Burda, M.R. and Miller, S. (2001). Identification and crystallisation of a heat- and protease-stable fragment of the bacteriophage T4 short tail fibre. *Biol. Chem.* **382**, 1049-1055.
13. van Raaij, M.J., Schoehn, G., Burda, M.R. and Miller, S. (2001). Crystal structure of a heat and protease-stable part of the bacteriophage T4 short tail fibre. *J. Mol. Biol.* **314**, 1137-1146.
14. Esnouf, R.M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Cryst. D* **55**, 938-940.
15. Kraulis, P.J. (1991). MOLSCRIPT; a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

16. Crowther, R.A., Henderson, R. and Smith, J.M. (1996). MRC image processing programs. *J. Struct. Biol.* **116**, 9-16.
17. Laskowski, R.A. (1993). PROCHECK; a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.
18. Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S. and Dodson, E.J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Cryst.* **D55**, 247-255.
19. Leiman, P.G., Kostyuchenko, V.A., Shneider, M.M., Kurochkina, L.P., Mesyanzhinov, V.V. and Rossmann, M.G. (2000). Structure of bacteriophage T4 gene product 11, the interface between the baseplate and short tail fibers. *J. Mol. Biol.* **301**, 975-985.
20. Vyas, N.K. (1991). Atomic features of protein-carbohydrate interactions. *Curr. Opin. Struct. Biol.* **1**, 732-740.
21. Baxa, U., Steinbacher, S., Miller, S., Weintraub, A., Huber, R. and Seckler, R. (1996). Interactions of phage P22 tails with their cellular receptor, Salmonella O-antigen polysaccharide. *Biophys. J.* **71**, 2040-2048.
22. Lee, Y.C. (1997). Fluorescence spectrometry in studies of carbohydrate-protein interactions. *J. Biochem. (Tokyo)* **121**, 818-825.
23. Hilge, M., Gloor, S.M., Rypniewski, W., Sauer, O., Heightman, T.D., Zimmermann, W. *et al* (1998). High-resolution native and complex structures of thermostable beta-mannanase from *Thermomonospora fusca* - substrate specificity in glycosyl hydrolase family 5. *Structure* **6**, 1433-1444.
24. Lortat-Jacob, H., Chouin, E., Cusack, S. and van Raaij, M.J. (2001). Kinetic analysis of adenovirus fiber binding to its receptor reveals an avidity mechanism for trimeric receptor-ligand interactions. *J. Biol. Chem.* **276**, 9009-9015.
25. Liao, C.P. and Syu, W., Jr. (2002). Analysis of the baseplate region of phage AR1 that specifically infects *Escherichia coli* O157:H7. *J. Microbiol. Immunol. Infect.* **35**, 269-271.
26. Nicholls, A., Sharp, K.A. and Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281-296.
27. Zorzopulos, J. and Kozloff, L.M. (1978). Identification of T4D bacteriophage gene product 12 as the baseplate zinc metalloprotein. *J. Biol. Chem.* **253**, 5543-5547.
28. Auld, D.S. (2001). Zinc coordination sphere in biochemical zinc sites. *Biometals*. **14**, 271-313.
29. Steinbacher, S., Seckler, R., Miller, S., Steipe, B., Huber, R. and Reinemer, P. (1994). Crystal structure of P22 tailspike protein: interdigitated subunits in a thermostable trimer. *Science* **265**, 383-386.
30. Holm, L. and Sander, C. (1996). Mapping the protein universe. *Science* **273**, 595-603.
31. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. *et al* (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389-3402.
32. Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S. *et al* (2000). Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res.* **7**, 331-338.
33. Blattner, F.R., Plunkett, G., III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M. *et al* (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453-1474.

34. Cubo, M.T., Economou, A., Murphy, G., Johnston, A.W. and Downie, J.A. (1992). Molecular characterization and regulation of the rhizosphere-expressed genes rhiABCR that can influence nodulation by *Rhizobium leguminosarum* biovar *viciae*. *J. Bacteriol.* **174**, 4026-4035.
35. Oliver, D.B. and Crowther, R.A. (1981). DNA sequence of the tail fibre genes 36 and 37 of bacteriophage T4. *J. Mol. Biol.* **153**, 545-568.
36. Miroshnikov, K.A., Marusich, E.I., Cerritelli, M.E., Cheng, N., Hyde, C.C., Steven, A.C. *et al* (1998). Engineering trimeric fibrous proteins based on bacteriophage T4 adhesins. *Protein Eng.* **11**, 329-332.
37. Rousseau, F., Schymkowitz, J.W. and Itzhaki, L.S. (2003). The unfolding story of three-dimensional domain swapping. *Structure (Camb.)* **11**, 243-251.
38. Liu, Y. and Eisenberg, D. (2002). 3D domain swapping: as domains continue to swap. *Protein Sci.* **11**, 1285-1299.
39. Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M. *et al* (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* **116**, 190-199.
40. Chacon, P. and Wriggers, W. (2002). Multi-resolution contour-based fitting of macromolecular structures. *J. Mol. Biol.* **317**, 375-384.
41. Davies, C.M., Apte, S.C. and Peterson, S.M. (1995). Beta-D-galactosidase activity of viable, non-culturable coliform bacteria in marine waters. *Lett. Appl. Microbiol.* **21**, 99-102.
42. Leslie, A.G. (1999). Integration of macromolecular diffraction data. *Acta Cryst.* **D55**, 1696-1702.
43. Collaborative Computational Project Number 4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760-763.
44. Otwinowski, Z. and Minor, W. (1997). Processing of x-ray diffraction data collected in oscillation mode. *Methods In Enzymology* **276**, 307-326.
45. Terwilliger, T.C. and Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Cryst.* **D55**, 849-861.
46. Terwilliger, T.C. (2000). Maximum-likelihood density modification. *Acta Cryst.* **D56**, 965-972.
47. Perrakis, A., Morris, R. and Lamzin, V.S. (1999). Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458-463.
48. Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* **A47**, 110-119.
49. Lamzin, V.S. and Wilson, K.S. (1997). Automated refinement for protein crystallography. *Methods In Enzymology.* **277**, 269-305.

3.2

Crystallisation and preliminary crystallographic studies of a new crystal form of the bacteriophage T4 short tail fibre

Summary

Six long and six short tail fibres mediate adsorption of T4 bacteriophage to the *Escherichia coli* host cell. After at least three long tail fibres have bound, short tail fibres extend and bind irreversibly to the core region of the host cell lipo-polysaccharide (LPS), serving as inextensible stays during penetration of the cell envelope by the tail tube. The short tail fibre consists of a parallel, in-register, trimer of gene product 12 (gp12). Parts of the structure of the short tail fibre were determined before, the so-called 33 kDa and 45 kDa fragments, of which the latter contains the receptor-binding domain (see chapter 3.1 of this thesis). None of these structures had ordered electron density for about half of the amino acid residues, although we did confirm their presence in both crystals by N-terminal sequencing. Here we present a partial structure of a new crystal form of the 45 kDa fragment, which contains all the residues present in the 33 kDa and 45 kDa fragment (amino acids 246 to 527), and also shows low resolution electron density for the residues invisible in the structure described in chapter 3.1 (some or all of the amino acids 85-245). The structure presented here confirms the hypothesis of the composite structure of the ordered regions of the 33kDa and 45kDa given in chapter 3.1 Figure 3.2d.

Introduction

Previously, a 33 kDa proteolytic fragment was generated in the presence of EDTA (ethylenediaminetetra-acetic acid) and crystallised [1]. This fragment consisted of residues 85-396 and 518-527, however in the crystal structure [2] residues 85-245 were invisible due to static disorder. By omitting EDTA and adding divalent zinc ions during proteolysis we could ensure that amino acids 397-517 are included in the so-called 45 kDa fragment [3]. The 45 kDa fragment contains the C-terminal receptor-binding domain however, residues 85-329 are absent. The amino acids 330-396 and 518-527 were the only residues visible in both the 33 kDa and 45 kDa fragments. See Figure 3.7 for an overview of both structures.

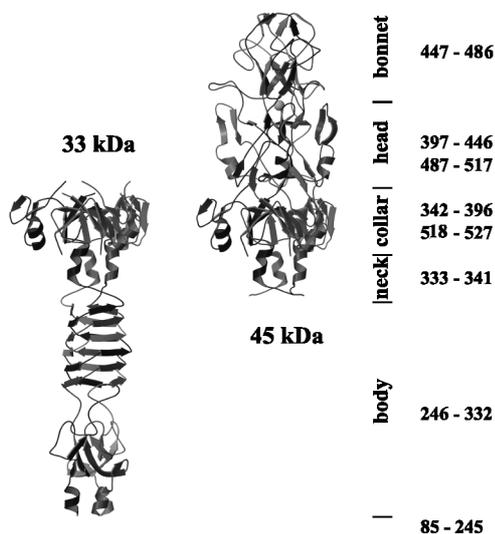


Figure 3.7. The crystal structures of the 33 kDa (PDB 1HW6) and 45 kDa (PDB 1OCY) fragments. The corresponding amino acids and domains are depicted on the right.

Here we report a new crystal form of a proteolytic fragment of gp12. This crystal form was obtained in experiments performed to determine the binding location of the LPS to the short tail fibre receptor-binding domain (for further details on the receptor-binding domain see Chapter 3.1). Unfortunately no LPS did bind to the receptor-binding domain, however, a different crystal form was obtained. This crystal form has a larger part of the electron density ordered, amino acid residues 246-527 have clear electron density. For amino acid residues 85-245 that were not visible at all in the earlier determined structures, low resolution electron density is now present. The N-terminal domain of full length gp12 contains six times the same 17-residues sequence repeat [2]. One of these repeats (amino

acids 255-271) is present in the well ordered part of the 33 kDa structure so we searched for the other remaining repeats in the less ordered electron density via molecular replacement. Three possible repeats were found but did not really improve the density or Rfactors. Unfortunately the low resolution electron density is not ordered enough to enable building of any of the amino acids 85 to 245.

Results

Protein purification

The gp12 full-length protein was co-expressed with its chaperone gp57 and purified as has been described before, see chapter 3.1 of this thesis and ref. 3. The purified protein was incubated at 56 °C and treated with proteases in the presence of divalent zinc ions. SDS-PAGE revealed that a fragment of around 45 kDa was obtained. The protein (further referred to as the 45kDa fragment) was precipitated with ammonium sulphate and stored at 4 °C until further use.

Crystallisation and structure determination

The ammonium sulphate precipitated 45 kDa fragment was re-dissolved in HN-buffer (10 mM N-(2-hydroxyethyl)piperazine-N''-(2-ethane) sulphonic acid-sodium hydroxide, 150 mM sodium chloride, pH 7.4) to a concentration of 10-20 mg/ml. Protein concentrations were determined using UV absorption measurements at 280 nm; the absorption of a 1 mg/ml solution of the 45kDa fragment was assumed to be 1.0 at a path-length of 1 cm. In order to determine where the lipo-polysaccharide binds on the short tail fibre receptor-binding domain, crystallisation trials were set up in the presence of LPS. The LPS from *E.coli* mutant D21f1 was isolated and purified according to Boman & Monner and Prehm *et al.* [4,5]. Crystals of the 45 kDa fragment were obtained by vapour diffusion (at 295 K) against a reservoir solution containing 35 % (v/v) tertiary butanol buffered with 100 mM sodium citrate pH 5.6, 10 % (v/v) glycerol, and LPS (see Figure 3.8). The exact LPS concentration was difficult to determine due to the presence of impurities. The crystallisation conditions are the same as described in chapter 3.1 [3]. However, the protein crystallised in a different space group, R32, with single crystals growing as hexagon-shaped plates (Figure 3.8).

Data collection and processing

Native data were collected on beam-line ID29 at a wavelength of 0.9756 Å at the European Synchrotron Radiation Facility (ESRF). The crystal was flash-frozen and kept at 100 K during data collection. The total rotation angle was 269° (538 images of 0.5°). Reflections were integrated with the program MOSFLM [6] and processed using programs of the CCP4 suite [7]. The 45 kDa fragment crystallised in the rhombohedral space group R32 with unit cell parameters $a = b = 45.49$, $c = 1376.97$ Å (hexagonal setting). The overall completeness of this data set was 79.0 %, with a completeness of 54.1 % in the highest resolution bin (2.42 - 2.30 Å). Data in the higher resolution shell are less complete because of the data collection on a square detector. The overall completeness is rather low, because the long axis was parallel to the beam, causing many overlaps. The overall redundancy and R_{sym} were 10.4 and 0.080, respectively. The crystals contain 1 molecule in the asymmetric unit (i.e. one-third of a trimer), have a solvent content of 57.5 % (v/v) and a Wilson temperature factor of 36 Å². Data collection and processing details are summarised in Table 3.2.



Figure 3.8. Crystals of the short tail fibre protein in the presence of LPS. The protein crystallised in space-group R32.

Structure solution and refinement

Molecular replacement trials using the known structure of residues 246-396 plus 518-527 of the 33 kDa fragment (PDB-code 1H6W [2]) and the residues 330-527 of the 45 kDa fragment (PDB-code 1OCY [3]) gave a solution with an Rfactor of 0.58 and a correlation coefficient of 0.237. Rigid body refinement by REFMAC [8] using 19833 reflections resulted in an Rfactor of 0.443 and an Rfree of 0.447, for the resolution range of 40 to 2.3 Å. Amino acids 246-284 were removed from the refinement because these were not well ordered. Restrained refinement was carried out using REFMAC [8]. The top of the molecule appears to be somewhat rearranged and the amino acid residues 469-476

were rebuilt manually using O [9]. Several restrained refinement runs including amino acid residues 280-284, and manually rebuilding amino acid residues 246-279, resulted in an R-factor and Rfree of 0.31 and 0.36, respectively. An overview of the partial structure is shown in Figure 3.9. A mask was built using SOLOMON [10] and density modification was performed using SOLOMON and DM [11]. Despite the fact that a well-defined mask was obtained, the density modification was ineffective. At 4 Å resolution there is density visible for the remaining N-terminal residues 85-245, but it is not straightforward to build them.

Table 3.2. Data collection and processing statistics. Data statistics of the outer resolution shell (2.42 - 2.30 Å) are given in parentheses, where applicable

Data collection	45kDa fragment in space group R32
Crystal dimensions (mm)	0.2 x 0.2 x 0.05
Wavelength (Å)	0.9756
Resolution range (Å)	76.70 - 2.30
Crystal system	Rhombohedral
Space group	R32
Unit cell parameters (Å)	$a = b = 45.49, c = 1376.97$
Total No. reflections	215985
No. unique reflections	21737
Multiplicity	10.4 (5.1)
Rsym ^a	0.080 (0.26)
Completeness (%)	79.0 (54.1)
Average I/σ(I)	5.9 (2.7)
Solvent content (%)	57.5
V _M (Å ³ / Da)	2.9

^a $R_{sym} = \frac{\sum_h \sum_i |I_{hi} - \langle I_h \rangle|}{\sum_h \sum_i I_{hi}}$, where I_{hi} is the intensity of the i th measurement of the same reflection and $\langle I_h \rangle$ is the mean observed intensity for that reflection

In an attempt to build extra residues, a two-dimensional search was done with the program AMORE maintaining residues 279-527 (the most well-ordered amino acids) fixed, and translating and rotating residues 255-271 (one sequence repeat) in small intervals (0.001 x cell-axis c and 3°, respectively) along z, which is the three-fold axis of

the crystal and of the trimeric structure. This way, apart from the original 255-271 repeat, three more possible repeats were found which looked to fit reasonably into the density. However, incorporating the respective amino acid changes followed by refinement did not really improve the density or the Rfactors (the minimum Rfactor obtained was 0.31 with an Rfree of 0.35).

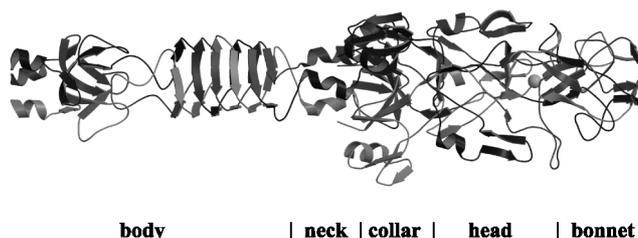


Figure 3.9. Crystal structure of short tail fibre (gp12) of bacteriophage T4, amino acid residues 246-527. The head and bonnet form the receptor-binding domain. A grey sphere represents the zinc ion present on the 3-fold axis. The picture was prepared with MOLSCRIPT [12].

Discussion and future prospects

The short tail fibre (gp12) of bacteriophage T4 crystallised into an other space group (R32) as before (P321), having more of the density ordered in the crystals compared to the two partial structures determined earlier [2,3]. A view of the partial crystal structure is shown in Figure 3.9. This structure contains amino acid residues 246-527, which can be divided in the "body" (246-332), "neck" (333-341), "collar" (342-396 and 518-527), "head" (397-446 and 487-517), and "bonnet" (447-486). The metal binding site, which has a zinc-ion on the 3-fold axis, is between the "head" and the "bonnet" as before [3]. We could not identify extra electron density located near to the receptor-binding domain; therefore it appears no LPS has bound.

Amino acid residues 330-527 of the 1OCY structure were superimposed with the residues 330-527 of the structure described here, using LSQKAB from the CCP4 program suite [7]. This showed that the top region of the receptor-binding domain, the "bonnet", has minor differences in the main-chain positions for the residues 469-476 (Figure 3.10). Whether the differences in main-chain position of residues 469-476 are caused by the difference in crystal contacts or by the LPS present in the crystallisation conditions, is not known.

The Rfactors suggest that there is still improvement possible in the missing unsolved domain of residues 85-245. Because we were not able to build these residues into the electron density map, and solvent flattening did not improve the electron density map, a possibility would be to improve the phases using data collection on heavy atom-derivatised crystals. From the structure determined before [2,3], we know that the 45 kDa fragment of gp12 has at least one residue (Cys372), which can react with methylmercury acetate, this residue is solvent accessible in the crystals as well. Hg-derivatisation as done for the 45 kDa fragment may therefore be a good strategy to try to improve the phases. Collecting a native data set with the long axis of the crystal almost parallel to the rotation axis would improve the completeness of the native data set.

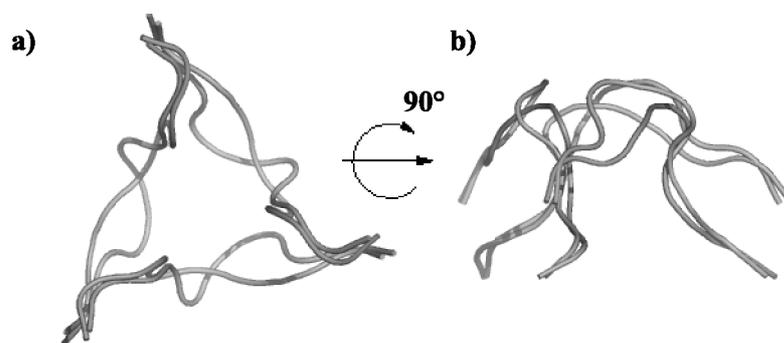


Figure 3.10. Superposition of the $C\alpha$ atoms (amino acid 468-477) of a part of the "bonnet" for the two gp12 trimers in different space groups. a) Top view on the bonnet along the three-fold axis. Both gp12 trimers in different colours, R32 (blue) and P321 (green). b) Rotated 90° , same colours as in a. It shows clearly that the top region of the receptor-binding domain, the "bonnet", has differences in the main-chain positions for the residues 469-476. Picture was made with PyMOL [13]. For colour picture see Appendix A.

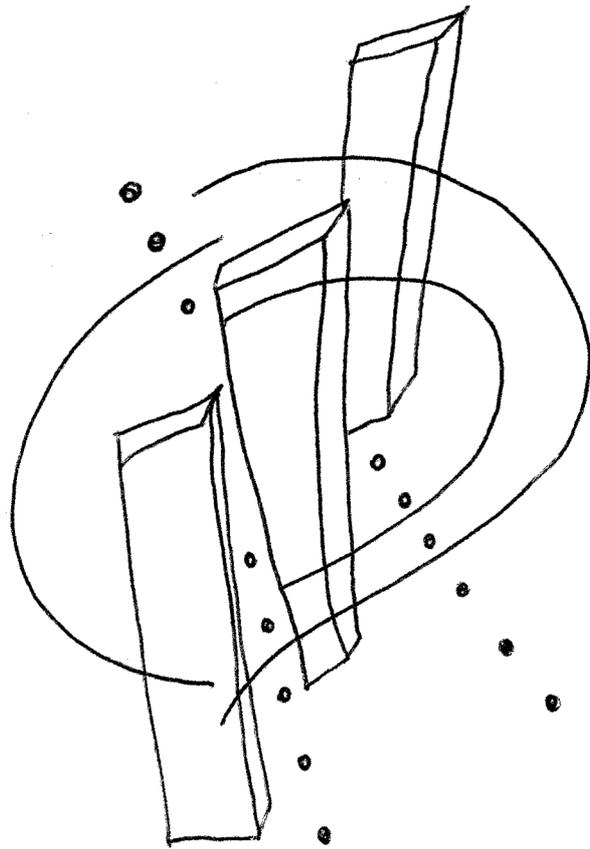
Acknowledgements

We are grateful to B. Shepard and other members of the EMBL-ESRF Joint Structural Biology Group for providing crystallographic data collection facilities and help therewith. We would like to thank Stefan Miller for providing us with the LPS and R.A.G. de Graaff for helpful discussions.

References

1. van Raaij, M.J., Schoehn, G., Jaquinod, M., Ashman, K., Burda, M.R. and Miller, S. (2001). Identification and crystallisation of a heat- and protease-stable fragment of the bacteriophage T4 short tail fibre. *Biol. Chem.* **382**, 1049-1055.
2. van Raaij, M.J., Schoehn, G., Burda, M.R. and Miller, S. (2001). Crystal structure of a heat and protease-stable part of the bacteriophage T4 short tail fibre. *J. Mol. Biol.* **314**, 1137-1146.
3. Thomassen, E., Gielen, G., Schutz, M., Schoehn, G., Abrahams, J.P., Miller, S. *et al* (2003). The structure of the receptor-binding domain of the bacteriophage T4 short tail fibre reveals a knitted trimeric metal-binding fold. *J. Mol. Biol.* **331**, 361-373.
4. Boman, H.G. and Monner, D.A. (1975). Characterization of lipopolysaccharides from *Escherichia coli* K-12 mutants. *J. Bacteriol.* **121**, 455-464.
5. Prehm, P., Jann, B., Jann, K., Schmidt, G. and Stirm, S. (1976). On a bacteriophage T3 and T4 receptor region within the cell wall lipopolysaccharide of *Escherichia coli* B. *J. Mol. Biol.* **101**, 277-281.
6. Leslie, A.G. (1999). Integration of macromolecular diffraction data. *Acta Cryst.* **D55**, 1696-1702.
7. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760-763.
8. Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S. and Dodson, E.J. (1999). Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Cryst.* **D55 (Pt 1)**, 247-255.
9. Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst.* **A47**, 110-119.
10. Abrahams, J.P. and Leslie, A.G.W. (1996). Methods used in the structure determination of bovine mitochondrial F₁ ATPase. *Acta Cryst.* **D52**, 30-42.
11. Cowtan, K. (1994). *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography.* **31**, 34-38.
12. Esnouf, R.M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Cryst.* **D55**, 938-940.
13. DeLano, W.L. (2002). The PyMOL Molecular Graphics System on World Wide Web <http://www.pymol.org>.

Potato Serine Protease Inhibitor



4.1

Crystallisation and preliminary X-ray crystallographic studies on a serine protease inhibitor from the Kunitz family

Ellen A.J. Thomassen, Laurice Pouvreau, Harry Gruppen and Jan Pieter Abrahams

Published in: *Acta Cryst* (2004) D60, 1464-1466

Summary

Interest in protease inhibitors has been renewed because of their potent activity in preventing carcinogenesis in a wide variety of *in vivo* and *in vitro* model systems. Potato tubers contain a wide range of such protease inhibitors. In cv. *Elkana* potato tubers, protease inhibitors represent about 50 % of the total amount of soluble protein. Potato Serine Protease Inhibitor (PSPI), one of the isoforms of the most abundant group of protease inhibitors, is a dimeric double-headed Kunitz-type inhibitor. No high-resolution structural information on this type of inhibitor has so far been obtained, as all structures currently known are of the monomeric single-headed or monomeric double-headed types. Crystals were grown in 0.1M HEPES pH 7.5, 10 % PEG8000 and 8 % ethyleneglycol complemented with 9 mM 1-s-octyl- β -D-thioglucoiside or 0.1 M glycine. Data were collected from a single crystal under cryo-conditions to 1.8 Å resolution. The protein crystallises in space-group P2₁, unit cell parameters $a = 54.82$, $b = 93.92$, $c = 55.44$ Å and $\beta = 100.7^\circ$; the scaling R_{sym} is 0.044 for 45456 unique reflections.

Introduction

The majority of known and characterised protease inhibitors are serine protease inhibitors, they have been classified in at least 13 different families according to their sequence homology [1]. Protease inhibitors are denoted Kunitz-type inhibitor when particular conserved amino acids are present in their primary structures, such as four cysteinyl residues forming two disulphide bridges [1]. The Kunitz-type inhibitors are relatively small (approximately 20 kDa) and are mostly monomeric and single-headed, containing one reactive site that inhibits the protease. Kunitz-type inhibitors are pseudo-substrates that bind directly to the binding site of the protease. To date, only monomeric single-headed (*e.g.* winged bean chymotrypsin inhibitor) or monomeric double-headed inhibitors (*e.g.* soybean trypsin inhibitor) have been crystallised. PSPI is a dimeric double-headed Kunitz-type inhibitor with two subunits that differ in size (15 kDa for the larger subunit and 6 kDa for the smaller one [2]). The subunits share about 22 % sequence homology with the soybean trypsin inhibitor, the reference inhibitor for the Kunitz-type [3]. Here, we describe the crystallisation conditions of a dimeric Kunitz-type inhibitor, which inhibits trypsin, chymotrypsin and human leukocyte elastase [4].

Material and Methods

Purification

Potatoes of *cv. Elkana* (Avebe B.A., Veendam, The Netherlands) were stored at 277 K in the dark at a relative humidity of 95-100 % for a maximum period of six months. The potatoes were chopped into large pieces and mixed in the presence of sodium bisulphite (0.5 g/kg potatoes) to prevent oxidation of phenolic compounds. Potato juice was prepared as described previously [4]. PSPI 6.1 (the number representing the isoelectric point), one of the isoforms of PSPI group, was purified as described previously [4]. PSPI 6.1 is a dimeric protein of 20 273 Da [2], in which the 2 sub-units are disulphide linked; it shows approximately 22 % sequence homology with the Kunitz soybean trypsin inhibitor. An additional chromatofocusing step was included, using a Poly-buffer Exchanger 74 column (60 x 1.6 cm). The fraction containing PSPI 6.1 was loaded onto the column, which had been pre-equilibrated with 0.025 M imidazole-HCl buffer (pH 7.4) and the protein was eluted using a Polybuffer 74-HCl (pH 5.0) (dilution factor 1:8)

(Amersham Biosciences, Uppsala, Sweden). The polybuffer was removed using a HP phenyl sepharose column (10 x 2.6 cm). The chromatofocusing step resulted in the removal of minor contaminants. After purification, PSPI 6.1 was dialysed at 277 K against 10 mM Tris-HCl buffer (pH 8.0) and subsequently stored (253 K) at a concentration of 1 mg ml⁻¹.

Crystallisation

The purified PSPI 6.1 was concentrated to 10 mg ml⁻¹ in 10 mM Tris-HCl buffer (pH 8.0), using a centriprep 10 kDa ultracentrifugation device (Millipore, Bedford, MA, USA).

Crystallisation experiments were performed by the sitting-drop vapour-diffusion method at 295 K. Each drop was formed by mixing equal volumes (1 µl) of 10 mg ml⁻¹ PSPI 6.1 solution and reservoir solution. Crystals were grown (Figure 4.1) in 0.1 M HEPES pH 7.5, 10 % PEG8000 and 8 % ethyleneglycol complemented with 9 mM 1-s-octyl-β-D-thioglucoiside (Hampton Detergent screen 1) or 0.1 M glycine (Hampton Additive screen 2).

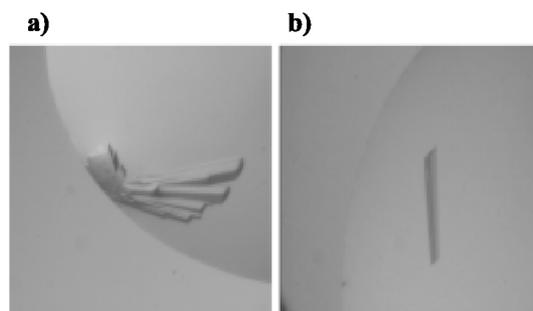


Figure 4.1. Crystals of Potato Serine Protease Inhibitor (PSPI 6.1) in 0.1 M HEPES pH 7.5, 10 % PEG8000 and 8 % ethyleneglycol with detergent or additive. a) 9 mM 1-s-octyl-β-D-thioglucoiside b) 0.1 M glycine

X-ray data collection and processing

A single crystal was soaked in reservoir solution containing 25 % glycerol for 10 s and mounted on a nylon-fibre loop (Hampton), after which it was flash frozen in a liquid nitrogen stream at 100 K. The data were collected at beamline ID29 of the European Synchrotron Radiation Facility (ESRF) on an ADSC Q4R detector. The crystal-to-

detector distance was 200 mm and 100° of 1° oscillation images (3 passes of 0.5 s per frame) were collected ($\lambda = 0.9464 \text{ \AA}$) (Figure 4.2). The intensities were indexed with MOSFLM [5] and scaled using SCALA [6]. Molecular replacement was performed using AMORE [7], MOLREP [8] and BEAST [9], which are all available in the CCP4 program suite [10].

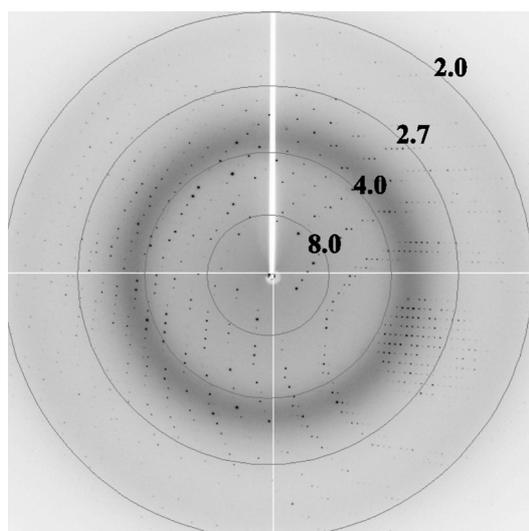


Figure 4.2. X-ray diffraction pattern of a Potato Serine Protease Inhibitor (PSPI 6.1) crystal. The resolution circles are given in \AA .

Results

PSPI 6.1 was purified and crystallised as shown in Figure 4.1 a data set of a single crystal was collected to 1.8 \AA ; because of the use of a square detector the completeness in the highest resolution shell (1.90-1.80 \AA) is only 52.1 %. In Figure 4.2 an X-ray diffraction pattern of a PSPI 6.1 crystal is shown. The data was processed in space-group $P2_1$, with unit cell parameters of $a = 54.82$, $b = 93.92$, $c = 55.44 \text{ \AA}$ and $\beta = 100.7^\circ$. See Table 4.1 for data-collection and processing parameters. The exact Matthews coefficient (V_M) and solvent content of the crystals could not be determined. Neither a self-rotation function nor a native patterson indicated clear rotational or translational symmetry between molecules in the asymmetric unit (A.U.). If there are 2 molecules in the A.U., $V_M = 3.5 \text{ \AA}^3 \text{ Da}^{-1}$, corresponding to a solvent content of 64.2 %, whilst 3 molecules in the A.U. gives a $V_M = 2.3 \text{ \AA}^3 \text{ Da}^{-1}$ and a solvent content of 46.2 %. According to the

Matthews coefficient probabilities [11] the latter seems the more likely. We were not able to solve the structure by molecular replacement using the winged bean chymotrypsin inhibitor [12], which has a 32 % sequence homology or the Kunitz-type soybean trypsin inhibitor as a model. Structure solution by heavy atom methods is in progress.

Table 4.1. Data-collection and processing parameters. Data statistics of the outer resolution shell (1.90-1.80 Å) are given in parentheses, where applicable.

Native crystal PSPI 6.1	
Crystal dimensions (mm)	0.4 x 0.025 x 0.025
Temperature (K)	100
Wavelength (Å)	0.94644
Resolution range (Å)	54.23 - 1.80
Crystal system	Monoclinic
Space-group	P2 ₁
Unit cell parameters (Å)	$a = 54.82, b = 93.92, c = 55.44$ and $\beta = 100.7^\circ$
Total No. reflections	89852
No. unique reflections	45456
Rsym ^a	0.044 (0.13)
Completeness (%)	88.7 (52.1) ^b
Average I/ σ (I)	4.4 (5.6)

^a $R_{sym} = \frac{\sum_h \sum_i |I_{hi} - \langle I_h \rangle|}{\sum_h \sum_i I_{hi}}$, where I_{hi} is the intensity of the i th measurement of the same reflection and $\langle I_h \rangle$ is the mean observed intensity for that reflection

^b data in the higher-resolution shell are less complete because of data collection on a square detector

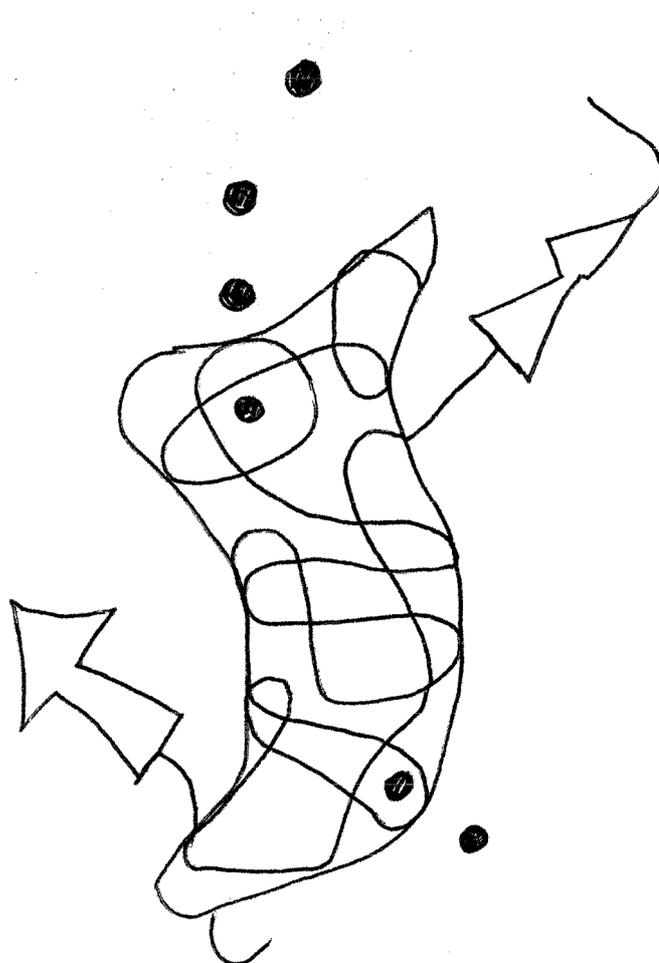
Acknowledgements

We thank M.J. van Raaij for help with the data collection and N.S. Pannu for help with data processing. We are grateful to E. Mitchell and other members of the EMBL-ESRF Joint Structural Biology Group for providing crystallographic data collection facilities and help therewith.

References

1. Richardson, M. (1991). Seed storage proteins: the enzyme inhibitors. *Methods Plant Biochem.* **5**, 259-305.
2. Pouvreau, L., Gruppen, H., Van Koningsveld, G.A., Van Den Broek, L.A. and Voragen, A.G. (2003). The most abundant protease inhibitor in potato tuber (cv. Elkana) is a serine protease inhibitor from the Kunitz family. *J. Agric. Food Chem.* **51**, 5001-5005.
3. Kunitz, M. (1945). Crystallization of a trypsin inhibitor from soybeans. *Science* **101**, 668-669.
4. Pouvreau, L., Gruppen, H., Piersma, S.R., Van Den Broek, L.A., Van Koningsveld, G.A. and Voragen, A.G. (2001). Relative abundance and inhibitory distribution of protease inhibitors in potato juice from cv. Elkana. *J. Agric. Food Chem.* **49**, 2864-2874.
5. Leslie, A.G. (1999). Integration of macromolecular diffraction data. *Acta Cryst.* **D55**, 1696-1702.
6. Evans, P.R. (1993). Data collection and Processing; proceedings of the CCP4 study weekend 1993, edited by Dodson, E., Kraulis, P.J. (1991). *J. Appl. Crystallog.* **24**, 946-950.
7. Navaza, J. (1994). AMoRe: an Automated Package for Molecular Replacement. *Acta Cryst.* **A50**, 157-163.
8. Vagin, A.A. and Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallog.* **30**, 1022-1025.
9. Read, R.J. (2004). Pushing the boundaries of molecular replacement with maximum likelihood. *Acta Cryst.* **D57**, 1373-1382.
10. Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760-763.
11. Kantardjieff, K.A. and Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci.* **12**, 1865-1871.
12. Ravichandran, S., Sen, U., Chakrabarti, C. and Dattagupta, J.K. (1999). Cryocrystallography of a Kunitz-type serine protease inhibitor: the 90 K structure of winged bean chymotrypsin inhibitor (WCI) at 2.13 Å resolution. *Acta Cryst.* **D55**, 1814-1821.

Recombinant Human Lactoferrin



5.1

The protein structures of recombinant human lactoferrin produced in the milk of transgenic cows and human milk-derived lactoferrin are identical

**Ellen A.J. Thomassen[#], Harry A. van Veen[#], Patrick H.C. van Berkel,
Jan H. Nuijens and Jan Pieter Abrahams**

[#]These authors contributed equally to the work and thus should both be regarded as "first author".

Accepted for publication in Transgenic Research (in revised form).

Summary

Human lactoferrin (hLF) is an iron-binding glycoprotein involved in the host defence against infection and excessive inflammation. As the availability of (human milk-derived) natural hLF is limited, alternative means of production of this biopharmaceutical are extensively researched. Here we report the crystal structure of recombinant hLF (rhLF) at a resolution of 2.4 Å as it is expressed in the milk of transgenic cows. To our knowledge, this is the first reported structure of a recombinant protein produced in milk of transgenic livestock. Even though rhLF contains oligomannose- and hybrid-type N-linked glycans next to complex-type glycans, which are the only glycans found on natural hLF, the structures are identical within the experimental error (r.m.s. deviation of only 0.28 Å for the main-chain atoms). Of the differences in polymorphic amino acids between the natural and recombinant hLF variant used, only the side-chain of Asp⁵⁶¹ could be modelled into the rhLF electron density map. Taken together, the results confirm the structural integrity of the rhLF variant used in this study. It also confirms the validity of the transgenic cow mammary gland as a vehicle to produce recombinant human proteins.

Introduction

Lactoferrin (LF) is a metal-binding glycoprotein of 77 kDa belonging to the transferrin family [1]. The molecule is found in milk, tears, saliva, bronchial, intestinal and other secretions, but also in the secondary granules of neutrophils [2]. Based on the many reports of its antimicrobial and anti-inflammatory activity *in vitro*, LF is thought to be involved in the host defence against infection and excessive inflammation, most notably at mucosal surfaces [2]. Antimicrobial activities of LF include bacteriostasis by the sequestration of free iron [3] and bactericidal activity by destabilisation of the cell-wall [4,5]. Anti-inflammatory actions of LF include inhibition of hydroxyl-radical formation [6], of complement activation [7] and of cytokine production [8] as well as neutralisation of lipopolysaccharide (LPS) [9]. Because of these biological activities, a wide variety of applications of LF in human health care has been proposed, such as the treatment of infectious and inflammatory diseases. As a nutraceutical application, the molecule may be used as a component of clinical nutrition products aimed at the prevention and treatment of gastro-intestinal tract infections and inflammations.

The DNA and amino acid sequences of human lactoferrin (hLF) have been determined [10,11]. Human LF consists of a polypeptide chain of 692 amino acids, which is folded into two globular lobes [1]. These lobes, designated the N- and C-lobe, share an internal amino acid homology of about 40% [11], they are connected by a three-turn α -helix. Each lobe folds into α -helix and β -sheet arrays to form domains I and II, respectively, which are connected by a hinge region, creating a deep iron-binding cleft within each lobe. Each cleft binds a single ferric ion with high affinity while simultaneously incorporating a bicarbonate ion [1]. Crystallographic studies of hLF have showed that upon binding of iron, domain I of the N- and C-lobe rotates relative to domain II by $\sim 54^\circ$ and $\sim 20^\circ$, respectively, resulting in a more globular closed, and stable conformation of the entire molecule [12]. Next to the high affinity metal binding in the iron-binding cleft, LF also binds other metals albeit with much lower affinity [13]. This occurs at least in part via surface-exposed histidyl residues [14]. Whereas some of the biological activities of hLF relate to high or low affinity iron-binding, others are mediated by a positively charged domain located in the N-terminus. This domain binds to negatively charged ligands such as the lipid A portion of LPS [15], DNA [16], heparin [17] as well as other proteins such as lysozyme [18] and specific receptors [19,20].

Human LF contains three possible N-glycosylation sites, Asn¹³⁸ in the N-lobe and Asn⁴⁷⁹ as well as Asn⁶²⁴ in the C-lobe [10], which are utilised in about 94%, 100% and 9% of the molecules, respectively [21].

Recently, we reported the production of recombinant hLF (rhLF) in the milk of transgenic cows [22]. Comparative studies between rhLF and hLF from human milk (natural hLF) revealed identical iron-binding and release properties, and despite differences in N-linked glycosylation, equal effectiveness in various infection models [22]. Here, we report the crystallographic structure of rhLF in its iron-saturated conformation. The structure appeared to be identical to the structure reported for iron-saturated natural hLF [23].

Material and Methods

Expression and purification of rhLF

The production and purification of rhLF from the milk of transgenic cows has been described previously [22]. Briefly, a genomic hLF sequence with polymorphic amino acids at position 4 (insertion of Arg), 11 (Ala), 29 (Arg) and 561 (Asp) [24] under control of regulatory elements from the bovine α S₁ casein gene, was introduced into the bovine germline. The resulting transgenic cattle lines showed rhLF expression levels between 0.4 and 2.5 g/L. Purified rhLF was saturated with iron as described [18].

Crystallisation

Crystals were grown by micro-dialysis of rhLF (54 mg ml⁻¹ in 0.9 % NaCl) in a 100 μ l dialysis button against 5 mM sodium phosphate pH 8.5 with 10 % (v/v) ethanol at 4 °C. Deep red crystals appeared after 4 weeks and grew to dimensions of approximately 3 x 2 x 1 mm. The protein crystallised in the orthorhombic space-group P2₁2₁2₁, with cell dimensions $a = 55.94$, $b = 97.38$, and $c = 156.30$ Å with 1 molecule in the asymmetric unit.

Data collection

The dialysis button was transferred to 5 mM sodium phosphate pH 8.5 complemented with 20% (v/v) 2-methyl-2,4-pentanediol (MPD) for a week at 4 °C to stabilise the crystals, as described [1]. Part of a crystal was broken off and mounted in a

quartz capillary. X-ray data were collected at room temperature on a FR591 rotating-anode generator equipped with a MAR345 image-plate detector, using Cu K α radiation ($\lambda = 1.5418 \text{ \AA}$). The crystal-to-detector distance was 220 mm and a 0.5° oscillation angle was used per image for a total of 369 images. The intensities were indexed with MOSFLM [25] and scaled using SCALA [26]. During the scaling process it was observed that a number of the images were not useful, which is most likely caused by a non-uniform quality of the crystal in different directions. These images were left out of the scaling process and not used any further. The resulting data set contained 154 images (77°) and 104905 reflections of which 33492 were unique. The overall completeness of the data set was 96.2 %, with a completeness of 97.9 % in the highest resolution bin (2.53 - 2.40 \AA). The overall redundancy and R_{sym} were 3.3 and 0.057, respectively. Data collection and processing details are summarised in Table 5.1.

Table 5.1. Data-collection and processing parameters

Data collection	rhLF expressed in bovine milk
Crystal dimensions (mm)	3 x 2 x 1
Wavelength (\AA)	1.5418
Resolution range (\AA)	81.65 - 2.40
Crystal system	Orthorhombic
Space group	P2 ₁ 2 ₁ 2 ₁
Unit cell parameters (\AA)	$a = 55.94, b = 97.38, c = 156.30$
Total No. reflections	104905
No. unique reflections	33492
multiplicity	3.3 (3.3) ^a
R_{sym}^b	0.057 (0.33)
Completeness (%)	96.2 (97.9)
Average $I/\sigma(I)$	8.9 (2.2)
solvent content (%)	55.51
$V_M (\text{\AA}^3 / \text{Da})$	2.8

^a Data statistics of the outer resolution shell (2.53 - 2.40 \AA) are given in parentheses, where applicable. ^b $R_{\text{sym}} = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i |I_{hi}|$, where I_{hi} is the intensity of the i th measurement of the same reflection and $\langle I_h \rangle$ is the mean observed intensity for that reflection

Structure elucidation and refinement

The structure was determined by molecular replacement, using MOLREP [27] and the structure of iron-saturated natural hLF [23] as the search model. The search model included two Fe³⁺ and two CO₃²⁻ ions, but all waters and carbohydrates were omitted in the model. The rotation and translation functions were determined, the best solution had an R factor of 0.294 and a correlation coefficient of 0.814. Instead of a full molecular replacement, a rigid body refinement could have been performed, after applying an anti-clockwise rotation parallel to the b-axis on the search model, changing *a* into *c* and *c* into -*a*, which would have given the same result. The model was refined using restrained refinement using REFMAC [28] using 30833 reflections in the resolution range 19.63-2.40 Å (for the refinement statistics see Table 5.2). Water molecules were built using ARP [29]. Electron density maps were calculated and carbohydrates were fitted into the 2F_o - F_c electron density. The final model was checked using PROCHECK [30] and WHATIF [31].

Table 5.2. Refinement statistics

Refinement	rhLF expressed in bovine milk
Resolution range (Å)	19.63 - 2.40
number of reflections	30833 (1640) ^a
Rfactor ^b	0.18 (0.23)
protein atoms / water molecules	5365 / 54
r.m.s. deviations bonds (Å)	0.013
r.m.s. deviations angles (°)	1.40
Average B value protein / solvent (Å ²)	50.24 / 39.60
Average B value sugar / Fe ³⁺ / CO ₃ ²⁻ (Å ²)	86.29 / 31.90 / 26.43
Ramachandran statistics ^c (%)	86.4 / 13.0 / 0.3 / 0.3

^aData statistics of R_{free} are given in parentheses, where applicable

^b $R = \frac{\sum ||F_{obs}(hkl)| - |F_{calc}(hkl)||}{\sum |F_{obs}(hkl)|}$

^cAccording to the program PROCHECK [30]. The percentages are indicated of residues in the most favoured, additionally allowed, generously allowed and disallowed regions of the Ramachandran plot, respectively

Structural comparisons

The structure of rhLF isolated from transgenic bovine milk was compared to the structure of iron-saturated natural hLF (PDB-code 1LFG, [23]), using LSQKAB from the CCP4 program suite [32]. The amino acid residues 1-4 and 1-5 from natural hLF and rhLF, respectively, were disordered and therefore left out of the calculations.

Results and discussion

rhLF model

The final model for iron-saturated rhLF from transgenic bovine milk consists of 692 amino acid residues, 54 water molecules, two Fe^{3+} ions, two CO_3^{2-} ions, and two *N*-acetylglucosamine residues that are N-linked to Asn¹³⁸ and Asn⁴⁷⁹ (numbering according to [10]). The model has an R-factor of 0.18 and an R_{free} of 0.23. The Ramachandran statistics (Table 5.2) show 86.4 % of the residues to be in the "most favoured" regions and 13.3 % in the "allowed" regions; only 0.3 % is in the "disallowed" region, which are residues Leu³⁰⁰ and Leu⁶⁴³. As was described before [23], both residues are located in a γ -turn, which seems to be a conserved property of all transferrin structures determined. The N-terminal amino acid stretch (Gly¹-Arg²-Arg³-Arg⁴-Arg⁵-) was disordered and could not be modelled, which is consistent with the experiences of other authors [23,33]. The flexible surface loop at residues 418-424 showed no ordered density at 1σ , but at 0.7σ the main-chain and most of the side-chain atoms became visible. Many of the amino acid residues that are exposed to solvent have disordered side-chains, especially the long side-chains of arginine and lysine residues, showing high B factors.

Protein structure

The structure of rhLF isolated from the milk of transgenic cows was determined in its closed conformation. The folding of rhLF is the same as that of natural hLF isolated from human milk (PDB code 1LFG). The structure of rhLF was superimposed on natural hLF, the r.m.s. deviation in the atomic positions of the main-chain atoms is 0.28 Å, disregarding the disordered N-terminus. The superimposed structures are shown in Figure 5.1, only the flexible loop of amino acid residues 418-424 and the N-terminus show differences between the C α atoms of the differently expressed hLF proteins, indicating these residues may adopt alternative conformations.

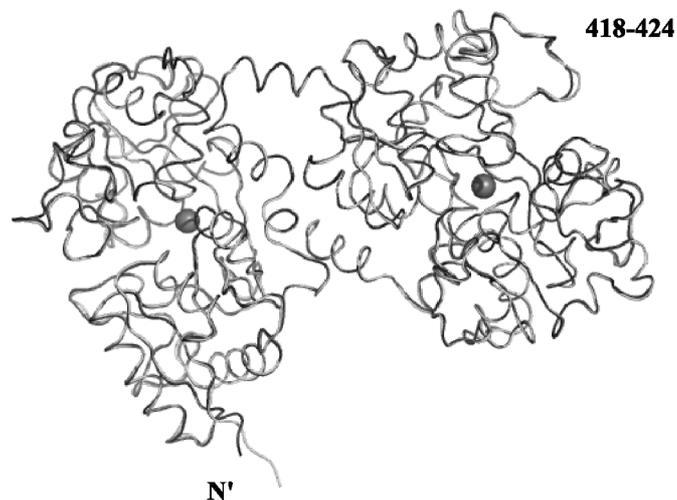


Figure 5.1. α superposition of rhLF and natural hLF. rhLF expressed in bovine milk is shown in yellow and natural hLF isolated from human milk in blue. Grey spheres show the iron atoms. The figure was prepared with PyMOL [34]. For colour picture see Appendix A.

Previously, we reported polymorphic sites in the hLF coding sequence at amino acid position 4 (deletion of Arg), position 11 (Ala or Thr), position 29 (Arg or Lys) and position 561 (Asp or Glu) [24]. Recombinant hLF differs from the natural hLF variant described by Haridas et al. [23] at position 4 (- \rightarrow Arg), 29 (Lys \rightarrow Arg) and 561 (Glu \rightarrow Asp). The Arg⁴ residue of rhLF could not be modelled, the N-terminal stretch Gly¹-Arg²-Arg³-Arg⁴-Arg⁵ may however be modelled upon binding of heparin, as Arg²-Arg³-Arg⁴-Arg⁵ is part of the positively charged domain responsible for binding of the ligand [17]. The side chain of the solvent-exposed Arg²⁹ residue was disordered and could not be modelled. Figure 5.2a shows the polymorphic site Asp⁵⁶¹ in the $2F_o - F_c$ density map. The polymorphic sites did neither change the structure of rhLF nor its biochemical activity, compared to natural hLF [22].

Glycosylation

Human lactoferrin contains three possible N-glycosylation sites, Asn¹³⁸ in the N-lobe and Asn⁴⁷⁹ as well as Asn⁶²⁴ in the C-lobe [10], which are utilised in about 94%, 100% and 9% of the molecules, respectively [21]. The glycans of natural hLF are of the

sialyl-*N*-acetylglucosaminic type [35]. Previously, we reported that rhLF from transgenic bovine milk also contains oligomannose and/or hybrid-type glycans [22].

In the rhLF structure described here, interpretable density for single carbohydrates was present at glycosylation sites Asn¹³⁸ and Asn⁴⁷⁹. The two *N*-acetylglucosamine residues were built into the $2F_o - F_c$ density map contoured at 0.9σ . The *N*-acetylglucosamine at position Asn⁴⁷⁹ is shown in Figure 5.2b. Further carbohydrates were insufficiently ordered to extend the model.

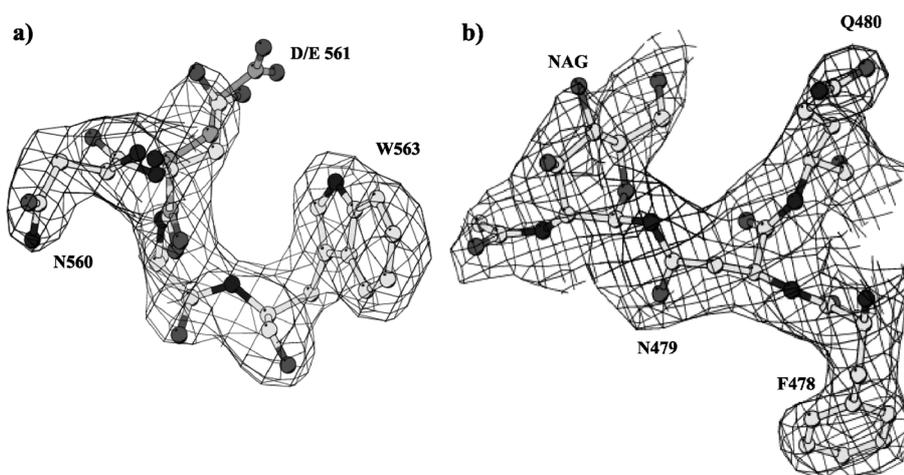


Figure 5.2. a) Crystallographic electron density surrounding Asp⁵⁶¹. rhLF amino acid residues Asn⁵⁶⁰, Asp⁵⁶¹, Ala⁵⁶², and Trp⁵⁶³ are shown in ball-and-stick representation with main chain carbon atoms in light grey, surrounded by the final $2F_o - F_c$ electron density map contoured at 1σ . At the polymorphic position 561 (numbering according to rhLF) the glutamamic acid (E) is shown with main chain carbon atoms in dark grey. b) Crystallographic electron density surrounding Asn⁴⁷⁹. Amino acid residues Phe⁴⁷⁸, Asn⁴⁷⁹, and Gln⁴⁸⁰ and *N*-acetylglucosamine (NAG) are shown in ball-and-stick representation surrounded by the final $2F_o - F_c$ electron density map contoured at 0.9σ . Picture was prepared with BOBSCRIPT [36], which is based on MOLSCRIPT [37].

Conclusions

The application of rhLF from bovine milk in human health care requires a thorough comparison of the physico-chemical and biological characteristics of the recombinant with the natural form. Here we report that despite the presence of polymorphic sites and differences in *N*-linked glycosylation, the three-dimensional structure of recombinant hLF closely matches the structure of natural hLF. This

observation confirms earlier findings of identical biochemical activity of hLF from both sources, paving the way for safe usage of rhLF in humans. In addition, the results illustrate the potential of transgenic cows to produce recombinant human proteins with a virtually identical structure compared to their human counterparts.

Coordinates. Coordinates and structure factors have been deposited in the Protein Data Bank (accession code 2BJJ).

Acknowledgements

We thank R.A.G. de Graaff, N.S. Pannu and M.L. Mannesse for helpful discussions.

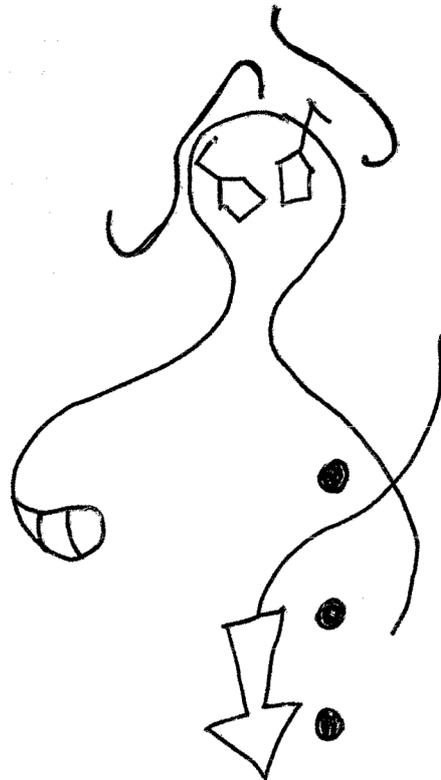
References

1. Anderson, B.F., Baker, H.M., Norris, G.E., Rice, D.W. and Baker, E.N. (1989). Structure of human lactoferrin: crystallographic structure analysis and refinement at 2.8 Å resolution. *J. Mol. Biol.* **209**, 711-734.
2. Nuijens, J.H., van Berkel, P.H. and Schanbacher, F.L. (1996). Structure and biological actions of lactoferrin. *J. Mammary. Gland. Biol. Neoplasia* **1**, 285-295.
3. Reiter, B., Brock, J.H. and Steel, E.D. (1975). Inhibition of *Escherichia coli* by bovine colostrum and post-colostral milk. II. The bacteriostatic effect of lactoferrin on a serum susceptible and serum resistant strain of *E. coli*. *Immunology* **28**, 83-95.
4. Ellison, R.T., III and Giehl, T.J. (1991). Killing of gram-negative bacteria by lactoferrin and lysozyme. *J. Clin. Invest.* **88**, 1080-1091.
5. Ellison, R.T., III, Giehl, T.J. and LaForce, F.M. (1988). Damage of the outer membrane of enteric gram-negative bacteria by lactoferrin and transferrin. *Infect. Immun.* **56**, 2774-2781.
6. Sanchez, L., Calvo, M. and Brock, J.H. (1992). Biological role of lactoferrin. *Arch. Dis. Child.* **67**, 657-661.
7. Kijlstra, A. and Jeurissen, S.H. (1982). Modulation of classical C3 convertase of complement by tear lactoferrin. *Immunology* **47**, 263-270.
8. Zucali, J.R., Broxmeyer, H.E., Levy, D. and Morse, C. (1989). Lactoferrin decreases monocyte-induced fibroblast production of myeloid colony-stimulating activity by suppressing monocyte release of interleukin-1. *Blood* **74**, 1531-1536.
9. Lee, W.J., Farmer, J.L., Hilty, M. and Kim, Y.B. (1998). The protective effects of lactoferrin feeding against endotoxin lethal shock in germfree piglets. *Infect. Immun.* **66**, 1421-1426.
10. Rey, M.W., Woloshuk, S.L., deBoer, H.A. and Pieper, F.R. (1990). Complete nucleotide sequence of human mammary gland lactoferrin. *Nucleic Acids Res.* **18**, 5288

11. Metz-Boutigue, M.H., Jolles, J., Mazurier, J., Schoentgen, F., Legrand, D., Spik, G. *et al* (1984). Human lactotransferrin: amino acid sequence and structural comparisons with other transferrins. *Eur. J. Biochem.* **145**, 659-676.
12. Baker, H.M., Anderson, B.F., Kidd, R.D., Shewry, S.C. and Baker, E.N. (2000). Lactoferrin three-dimensional structure; a framework for interpreting function. *in K. Shimazaki (Ed.), Lactoferrin: Structure, Function and Applications, Elsevier Science B. V. , Amsterdam.* 3-15.
13. Nagasako, Y., Saito, H., Tamura, Y., Shimamura, S. and Tomita, M. (1993). Iron-binding properties of bovine lactoferrin in iron-rich solution. *J. Dairy Sci.* **76**, 1876-1881.
14. Hutchens, T.W. and Yip, T.T. (1991). Metal ligand-induced alterations in the surface structures of lactoferrin and transferrin probed by interaction with immobilized copper(II) ions. *J. Chromatogr.* **536**, 1-15.
15. Appelmelk, B.J., An, Y.Q., Geerts, M., Thijs, B.G., de Boer, H.A., MacLaren, D.M. *et al* (1994). Lactoferrin is a lipid A-binding protein. *Infect. Immun.* **62**, 2628-2632.
16. He, J. and Furmanski, P. (1995). Sequence specificity and transcriptional activation in the binding of lactoferrin to DNA. *Nature* **373**, 721-724.
17. Mann, D.M., Romm, E. and Migliorini, M. (1994). Delineation of the glycosaminoglycan-binding site in the human inflammatory response protein lactoferrin. *J. Biol. Chem.* **269**, 23661-23667.
18. van Berkel, P.H., Geerts, M.E., van Veen, H.A., Kooiman, P.M., Pieper, F.R., de Boer, H.A. *et al* (1995). Glycosylated and unglycosylated human lactoferrins both bind iron and show identical affinities towards human lysozyme and bacterial lipopolysaccharide, but differ in their susceptibilities towards tryptic proteolysis. *Biochem J.* **312 (Pt 1)**, 107-114.
19. Legrand, D., van Berkel, P.H., Salmon, V., van Veen, H.A., Slomianny, M.C., Nuijens, J.H. *et al* (1997). The N-terminal Arg2, Arg3 and Arg4 of human lactoferrin interact with sulphated molecules but not with the receptor present on Jurkat human lymphoblastic T-cells. *Biochem J.* **327 (Pt 3)**, 841-846.
20. Ziere, G.J., Bijsterbosch, M.K. and van Berkel, T.J. (1993). Removal of 14 N-terminal amino acids of lactoferrin enhances its affinity for parenchymal liver cells and potentiates the inhibition of beta- very low density lipoprotein binding. *J. Biol. Chem.* **268**, 27069-27075.
21. van Berkel, P.H., van Veen, H.A., Geerts, M.E., de Boer, H.A. and Nuijens, J.H. (1996). Heterogeneity in utilization of N-glycosylation sites Asn624 and Asn138 in human lactoferrin: a study with glycosylation-site mutants. *Biochem J.* **319 (Pt 1)**, 117-122.
22. van Berkel, P.H., Welling, M.M., Geerts, M., van Veen, H.A., Ravensbergen, B., Salaheddine, M. *et al* (2002). Large scale production of recombinant human lactoferrin in the milk of transgenic cows. *Nat. Biotechnol.* **20**, 484-487.
23. Haridas, M., Anderson, B.F. and Baker, E.N. (1995). Structure of human diferric lactoferrin refined at 2.2 Å resolution. *Acta Cryst.* **D51**, 629-646.
24. van Veen, H.A., Geerts, M.E., van Berkel, P.H. and Nuijens, J.H. (2004). The role of N-linked glycosylation in the protection of human and bovine lactoferrin against tryptic proteolysis. *Eur. J. Biochem.* **271**, 678-684.
25. Leslie, A.G. (1999). Integration of macromolecular diffraction data. *Acta Cryst.* **D55**, 1696-1702.

26. Evans, P.R. (1993). Data collection and Processing; proceedings of the CCP4 study weekend 1993, edited by Dodson, E., Kraulis, P.J. (1991). *J. Appl. Crystallog.* **24**, 946-950.
27. Vagin, A.A. and Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallog.* **30**, 1022-1025.
28. Murschudov, G.N., Vagin, A.A. and Dodson, E.J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelyhood Method. *Acta Cryst.* **D53**, 240-255.
29. Lamzin, V.S. and Wilson, K.S. (1997). Automated refinement for protein crystallography. *Methods In Enzymology* **277**, 269-305.
30. Laskowski, R.A. (1993). PROCHECK; a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.
31. Vriend, G. (1990). WHAT IF: A molecular modelling and drug design program. *J. Mol. Graph.* **8**, 52-56.
32. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst.* **D50**, 760-763.
33. Sun, X.L., Baker, H.M., Shewry, S.C., Jameson, G.B. and Baker, E.N. (1999). Structure of recombinant human lactoferrin expressed in *Aspergillus awamori*. *Acta Cryst.* **D55 (Pt 2)**, 403-407.
34. DeLano, W.L. (2002). The PyMOL Molecular Graphics System on World Wide Web <http://www.pymol.org>.
35. Spik, G., Strecker, G., Fournet, B., Bouquelet, S., Montreuil, J., Dorland, L. *et al* (1982). Primary structure of the glycans from human lactotransferrin. *Eur. J. Biochem.* **121**, 413-419.
36. Esnouf, R.M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Cryst.* **D55 (Pt 4)**, 938-940.
37. Kraulis, P.J. (1991). MOLSCRIPT; a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

Summary and Conclusions



Summary and Conclusions

The three-dimensional structure of proteins is important for understanding the molecular basis for their specific activity. Several methods exist for investigating the three-dimensional structure, each with its own advantages and disadvantages; nuclear magnetic resonance (NMR) which has a restriction in the size of the protein, cryo-electron microscopy (cryo-EM) which produces structures of limited resolution, and X-ray crystallography. The main problem to overcome in X-ray crystallography is the requirement of three-dimensional, well diffracting crystals. The determination of the crystallisation conditions is still a daunting task for most proteins and essentially a trial and error procedure. As this is a very time- and protein-consuming task, many groups are developing or using robots to perform crystallisation trials in the nano-litre volume range. After obtaining crystals, the resolution that can be achieved in X-ray crystallography depends on the quality of the crystals. Through the application of synchrotron beam-lines, however, a resolution of 1.5 Å or even better is not an unrealistic goal anymore.

This thesis describes the purification of several different proteins as well as crystallisation experiments, in order to determine their structure. Unfortunately, problems in obtaining correctly folded protein and diffracting crystals in two cases prevented the elucidation of the full 3D atomic structure. Therefore other methods had to be used to get a glimpse of the structural basis of some of these proteins.

The human T cell receptor-CD3 complex consists of at least eight polypeptide chains: CD3 γ ϵ - and δ ϵ -dimers associate with the disulphide linked $\alpha\beta$ - and $\zeta\zeta$ -dimers to form a functional receptor complex. The exact interactions within this complex are still unknown, therefore the impact of single amino acid mutations in the CD3 γ chain on the CD3 γ ϵ heterodimer interaction was investigated. Four amino acid residues were mutated in the CD3 γ chain to different conserved or non-conserved residues and introduced in human T cells defective in CD3 γ -expression. Intracellular as well as cell surface expression of the introduced CD3 γ chains was determined as well as the association with CD3 δ , CD3 ϵ and the T cell receptor. Because cells from a CD3 γ deficient patient were

available, a comparison to this negative control could be made. The results of these experiments are described in chapter 2.1, showing the three phenotypes that were observed: i) the introduction of wild type CD3 γ and CD3 γ (78Y-F) fully restored the T cell receptor assembly and expression; ii) the introduction of CD3 γ (82C-S), CD3 γ (85C-S), and CD3 γ (76Q-E) all resulted in an impaired association between CD3 γ and CD3 ϵ and a lack of cell surface expressed CD3 γ ; iii) the introduction of CD3 γ (76Q-L) and CD3 γ (78Y-A) restored the expression of TCR- $\delta\epsilon\gamma\epsilon\zeta_2$ complexes although the association between CD3 γ and CD3 ϵ was impaired. These results show that particular amino acids in the CD3 γ chain are very important for optimal association between CD3 γ and CD3 ϵ and the assembly of a cell surface expressed $\alpha\beta$ TCR- $\delta\epsilon\gamma\epsilon\zeta_2$ complex.

In order to determine the structure of the intracellular domains of the human CD3 δ and CD3 ϵ that are thought to be important for the intracellular dimer formation, synthetic peptides corresponding to these domains have been synthesised. CD3 δ is a 44-mer, molecular weight and calculated pI of 4995.38 Da and 6.93, respectively. CD3 ϵ is a 55-mer, with molecular weight and pI of 6179.03 Da and 11.52, respectively. Many different crystallisation trials were performed, but for CD3 ϵ this did not result in even the formation of micro-crystals. The crystallisation trials for CD3 δ resulted in "sea urchin" type crystals of poor quality. Several rounds of micro-seeding the needles of the "sea urchin" did not result in better crystals. NMR spectra of the peptides showed poor dispersion and narrow line width, therefore it was concluded that the peptides are primarily random coil. After mixing of the peptides a 1D proton NMR was recorded, the spectrum of the combined peptides was a superimposition of the 2 spectra of the single peptides, leading to the conclusion that no interaction or folding is indicated by this NMR experiment. These experiments are described in chapter 2.2.

The bacteriophage T4, a bacterial virus which infects *Escherichia coli*, looks like a sort of moon-lander. It has a DNA containing head with a contractile tail and short and long tail fibres attached to a so-called baseplate. It adsorbs to the host's cell receptor via its long tail fibres and next the short tail fibres fold out from under the baseplate and also bind to the host's cell receptor. With a needle-like puncturing device, a hole is drilled into the host cell, providing the entry for DNA injection. After the phage particles are formed in the host cell, the host will burst and the new phages are released.

In chapter 3.1, the structure of the receptor-binding domain of the short tail fibre is described based on a 1.5 Å resolution X-ray structure determination. The short tail fibres are formed by trimers of the gene product 12 (gp12) and are attached to the baseplate via gp11. They contain 527 amino acids and are very stable, the protein needs to be heated to 56 °C before partial unfolding occurs. It was known that the full-length gp12 could not be crystallised due to aggregation. To overcome this problem, the protein was mildly proteolysed after partially unfolding in the hope of getting a shorter fragment of gp12 that would be able to crystallise. In the presence of zinc during the proteolysis, a 45 kDa fragment of gp12 was obtained, containing amino acids 85-527 according to N-terminal sequencing. The electron density map, however, only showed good electron-density for residues 330-527. The residues 85-329 seemed to be disordered in the crystal. Bacterial cell binding experiments showed that the 45 kDa fragment contains the receptor-binding domain, which was lacking in an earlier elucidated structure. The structure of the receptor-binding domain reveals a knitted trimeric metal-binding fold, containing one zinc ion octahedrally coordinated by the Nε2 of six histidine residues. It was reported earlier that gp12 was a zinc metallo-protein, assumed to contain one zinc atom per monomer. In the structure described in chapter 3.1, one zinc atom per trimer was found (although it was not possible to rule out that there are other zinc binding sites in the parts missing in the structure). The structure looked as if the protein was knotted, but after having a closer look the protein turned out to be knitted, i.e. it does not produce a topological knot upon "pulling at the ends". The three strands, with hardly any regular secondary structure elements, are intensively intertwined and together with the octahedrally coordinated zinc ion this gives the trimeric protein a very stable conformation. No related structure in the protein structure database could be found. The collar domain has some limited structural homology with the central domain of gp11, which is the connector protein between the baseplate and the short tail fibre (gp12).

Some suggestions have been made for binding of the LPS to the receptor-binding domain, based on the nature of the LPS core. It was assumed that phosphate groups bind to positive charges on the protein surface while the sugars bind to the aromatic side-chains. The trimer in general, and its top in particular, is positively charged, which could contribute to its affinity for the negatively charged *E.coli* membrane. Along the sides of the trimer there are three grooves which are flanked by or contain positively charged residues and aromatic amino acids. Each of these grooves may bind an LPS molecule.

This means that there is very tight binding between T4 and the host cell, if it is considered that the bacteriophage T4 has six short tail fibres which can bind simultaneously to the *E.coli* membrane, each through three receptor-binding sides.

In order to determine whether the receptor-binding domain is capable of binding to the host cell's lipopolysaccharides (LPS), crystallisation experiments were performed with the 45 kDa fragment in the presence of LPS. The results described in chapter 3.2 show that the fragment did not crystallise with LPS bound to it, but surprisingly the crystals had a different space-group and showed more ordered density. Electron density was now present for amino acid residues 246-527. The overall structures of the visible parts for the two gp12 trimers in space group R32 and P321 are the same. After superposition of the two trimers, amino acid residues 465-479 in the "bonnet" (the top of the trimeric protein containing the receptor-binding domain) show differences in the main-chain positions. Whether these changes are caused by differences in crystal contacts or by the LPS present in the crystallisation trials is unknown. Low-resolution density was present for some of the residues between 85-245. Attempts to improve the map by density modification using a solvent mask were unsuccessful. Also searching for repeats in the N-terminal electron density knowing that the N-terminal domain contains six repeats of which one is visible in the 45 kDa structure, did not result in better maps and Rfactors. Regrettably, it was not possible to interpret the low-resolution density.

In collaboration with the university of Wageningen a potato serine protease inhibitor (PSPI) was crystallised. Serine protease inhibitors are important due to their potent activity in preventing carcinogenesis in a wide variety of *in vivo* and *in vitro* model systems. Potato tubers contain a wide range of such protease inhibitors. In cv. *Elkana* potato tubers, protease inhibitors represent about 50 % of the total amount of soluble protein. One of the isoforms of the most abundant protease inhibitor, a dimeric double-headed Kunitz-type inhibitor, was crystallised. No high-resolution structural information on this type of inhibitors had so far been obtained, as all structures known at that time were of the monomeric single-headed or monomeric double-headed types. Crystals were grown and a native data set was collected to 1.8 Å resolution. Due to the lack of a dimeric double-headed inhibitor model in the PDB, so far the structure could not be solved by molecular replacement. Using monomeric double-headed and dimeric single-headed models did not result in the elucidation of the structure. Unfortunately these models were not homologous enough and therefore did not result in a solution. Heavy atom derivative

trials were performed, but resulted in the disruption of the crystal lattice or to low occupancies for the heavy metal positions. The crystallisation conditions are described in chapter 4. As soon as a suitable model for molecular replacement becomes available, the structure will be determined.

Human lactoferrin (hLF) is an iron-binding glycoprotein involved in the host defence against infection and excessive inflammation. Together with Pharming in Leiden the crystal structure of recombinant human lactoferrin (rhLF) expressed in the milk of transgenic cows was reported at a resolution of 2.4 Å. As far as is known the first structure of a recombinant protein produced in milk of transgenic livestock. The results described in chapter 5 show that the structures of rhLF and natural hLF (derived from human milk) are identical despite differences in N-linked glycosylation, confirming the validity of the transgenic cow mammary gland as a vehicle to produce recombinant human proteins.

Nederlandse samenvatting

De kristalstructuurbepaling van eiwitten door middel van Röntgen-diffractie is één manier waarmee de driedimensionale structuur van een eiwit bepaald kan worden. Aan de hand van de verkregen informatie kan vaak iets gezegd worden over de functie en de werking van het eiwit. Röntgen-diffractie wordt ook gebruikt in de ontwikkeling van medicijnen. Om te beginnen wordt de structuur van het ziekteverwekkende eiwit, het zogenaamde “target eiwit”, bepaald. De kennis van deze structuur wordt vervolgens gebruikt om geschikte medicijnen te ontwikkelen. Deze medicijnen moeten in staat zijn het actief gedeelte van het “target eiwit” te binden, om daardoor de activiteit van het eiwit uit te schakelen. Aan de hand van de kristalstructuur van “target eiwit” met het daaraan gebonden medicijn kan gekeken worden of het mogelijk is de eigenschappen van het medicijn te verbeteren. Dit kan door bijvoorbeeld het medicijn zo te veranderen dat de binding van het medicijn aan het “target eiwit” versterkt wordt waardoor een lagere dosering van het desbetreffende medicijn mogelijk is. De eiwitkristallografie is in de laatste jaren enorm verbeterd, door o.a. sterkere Röntgenbronnen, waardoor een resolutie van 1.5 Å of hoger niet langer een onrealistisch doel is. In de eiwitkristallografie wordt gewerkt met de eenheid Angstrom (Å), $1 \text{ Å} = 1.0 \times 10^{-10} \text{ m}$, en komt ongeveer overeen met de afstand tussen twee atomen in de materie.

Dit proefschrift beschrijft de zuivering en kristallisatie van verschillende eiwitten. In **hoofdstuk 1** worden twee van de bestudeerde eiwitten besproken. Een van deze eiwitten betreft het menselijk T cel receptor CD3 complex (TCR-CD3), dat tenminste uit acht eiwitketens bestaat, namelijk de $\alpha\beta$ -keten, het $\delta\epsilon$ - en $\epsilon\gamma$ -dimeer en het $\zeta\zeta$ -dimeer. Er zijn nog steeds discussies gaande of dit werkelijk de samenstelling is van het TCR-CD3 complex. In hoofdstuk 1 wordt verder een introductie gegeven over de bacteriofaag T4, een virus dat in staat is bacteriën te infecteren. Het virus heeft de uiterlijke kenmerken van een maanlander. Het virus heeft 6 lange en 6 korte pootjes die aan de gastheercel binden voor en tijdens de infectie. Deze pootjes zorgen ervoor dat tijdens de infectie de bacteriofaag zijn DNA via een zogenaamde injectiespuit kan injecteren in de gastheercel. Er was nog weinig structurele informatie beschikbaar over deze pootjes. Na infectie vermenigvuldigt de bacteriofaag zich in de gastheer totdat deze openscheurt en de nieuwe bacteriofagen vrij komen.

De overproductie en zuivering van verschillende CD3 ketens leverde geen gevouwen eiwitten op, wat de poging tot kristallisatie deed mislukken. In diezelfde periode werd de NMR structuur beschreven van het CD3 ϵ γ heterodimeer, dit deed mij besluiten de hoofdlijn in het onderzoek te verschuiven. In plaats van op de kristallisatie werd het onderzoek gericht op de associatie tussen de verschillende heterodimeren *in vivo*. In **hoofdstuk 2** is de invloed beschreven van mutaties in de CD3 γ keten op de associatie *in vivo* van het CD3 ϵ γ dimeer en de daaropvolgende T cel receptor CD3 complexvorming. De resultaten laten zien dat bepaalde aminozuren in de CD3 γ keten erg belangrijk zijn voor een goede associatie van het CD3 ϵ γ heterodimeer. Sommige mutaties verhinderen een goede vorming van het $\alpha\beta\delta\epsilon\gamma\zeta\zeta$ complex op het celmembraan, ondanks het feit dat FACS analyse aantoonde dat de cellen CD3 positief zijn. Mogelijkerwijs is in deze cellen de samenstelling van het TCR-CD3 complex op het cel-oppervlak veranderd in $\alpha\beta\delta\epsilon\delta\epsilon\gamma\zeta\zeta$. Aangezien de CD3 δ keten grote overeenkomsten vertoont met de CD3 γ keten zou dit een mogelijkheid kunnen zijn. Tevens werden gesynthetiseerde peptiden van de intracellulaire domeinen van CD3 δ en CD3 ϵ gebruikt voor kristallisatie en NMR experimenten. Uit deze experimenten blijkt dat de beide peptiden onder de heersende omstandigheden zich niet in een gevouwen conformatie bevinden en daardoor structuurbepaling uitgesloten is.

In **hoofdstuk 3** is de kristalstructuur van het receptorbindende domein van de korte pootjes van de bacteriofaag T4 beschreven. De structuur bestaat uit drie identieke ketens die als in een knotje in elkaar geweven zijn. In het midden van deze drie ketens is een zink ion aanwezig. Alle drie de ketens binden aan dit ion door middel van twee histidine aminozuren. Dit zink-ion blijkt voor een grote stabilisatie van het trimeer te zorgen, want wanneer het verwijderd wordt door middel van EDTA blijkt het receptor bindende domein bij verhitting uit elkaar te vallen. Terwijl bij aanwezigheid van zink het receptorbindende domein bij verhitting niet ontvouwt. Tevens worden in hoofdstuk 3 resultaten beschreven betreffende hetzelfde eiwit maar dan gekristalliseerd in een andere ruimtgroep. In deze kristalvorm blijkt een groter deel van de elektronendichtheid geordend te zijn in vergelijking met de eerder verkregen kristallen. In deze extra elektronendichtheid kon de structuur van een fragment van de korte pootjes (het 33 kDa fragment) worden ingebouwd. Echter tot op heden was het niet mogelijk om overal in deze extra elektronendichtheid aminozuren te bouwen.

In samenwerking met de universiteit van Wageningen kristalliseerde ik een serine protease inhibitor (PSPI, potato serine protease inhibitor) uit aardappel en kon ik een natieve dataset tot 1.8 Å meten. Deze resultaten staan beschreven in **hoofdstuk 4**. Deze protease inhibitors zijn onlangs in de belangstelling gekomen vanwege hun mogelijke anti-kanker activiteit. De PSPI is het eerste “double-headed” dimeer waarvan de kristallisatiecondities zijn beschreven. Andere protease inhibitors waarvan de driedimensionale structuur bekend is, zijn of een “double-headed” monomeer dan wel een “single-headed” dimeer. Helaas was geen van de aanwezige structuren in de PDB geschikt als zoekmodel in de Molecular Replacement. De gemaakte derivaten vertoonden een te lage bezetting van de zware atomen of verstoorden het kristalrooster, waardoor de structuur nog niet bepaald kon worden.

In **hoofdstuk 5** worden de resultaten beschreven van de structuurbepaling van menselijk lactoferrine (hLF) geproduceerd in melk van transgene koeien. Dit onderzoek werd uitgevoerd in samenwerking met Pharming te Leiden. De resultaten laten zien dat ondanks de verschillen in N-linked glycosylatie, de kristalstructuur van hLF uit transgene koeienmelk overeenkomt met die van hLF geïsoleerd uit menselijke melk. Er was door Pharming aangetoond dat er geen verschil is in beide lactoferrines, getest in diverse biochemische experimenten en deze structuurbepaling bevestigt de mogelijkheid transgene koeien te gebruiken voor de productie van menselijke recombinante eiwitten in melk.

Hoofdstuk 6 is een samenvatting met de daarbijbehorende conclusies.

List of abbreviations

APC: antigen-presenting cell
BCR: B cell receptor
BLAST: basic local alignment search tool
BSA: bovine serum albumin
CD: clusters of differentiation
cv: cultivar
dCTP: deoxycytidine 5'-triphosphate
DNA: deoxyribonucleic acid
DOC: sodium deoxycholate
E. coli: *Escherichia coli*
EDTA: ethylenediaminetetra-acetic acid
ER: endoplasmic reticulum
FACS: fluorescence activated cell sorter
FBS: fetal bovine serum
FCS: fetal calf serum
FITC: fluorescein isothiocyanate
GAM-FITC: goat anti-mouse - fluorescein isothiocyanate
GFP: green fluorescent protein
gp: gene product
h: hour
HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hLF: human LF
Hm-dCTP: 5-hydroxymethyl dCTP
hoc: highly antigenic outer capsid
HSA: human serum albumin
Ig: immunoglobulin
IMDM: Iscove's modified Dulbecco's medium
IPTG: Isopropyl β -D-thiogalactoside
iron-saturated rhLF: rhLF that has completely been saturated with iron *in vitro*
ITAM: immunoreceptor tyrosine-based activation motif

kDa: kilodalton
KDO: 2-keto-3-deoxyoctonic acid
LF: lactoferrin
LPS: lipo-polysaccharide
LTF: long tail fibre
mAb: monoclonal antibody
mCi: milli-currie
MHC: major histocompatiblity complex
min: minute(s)
natural hLF: hLF purified from human milk
NHS: normal human serum
nm: nano-meter
NMR: nuclear magnetic resonance
NP40: Nonident-40
NRS: normal rabbit serum
OmpC: outer membrane protein C
PAGE: poly-acrylamide gel electrophoresis
PAS: protein A sepharose
PBMC: Peripheral blood mononuclear cells
PBS: phosphate-buffered saline
PE: phycoerythrin
PEG: polyethylene glycol
PHA: phytohemagglutinin
pI: iso-electric point
PSPI: potato serine protease inhibitor
PTK: protein tyrosine kinases
rhLF: recombinant hLF
rIL: recombinant interleukin
SDS: sodium dodecylsulphate
soc: small outer capsid
STF: short tail fibre
TCR: T cell receptor
TM: transmembrane

List of publications

"The structure of the receptor-binding domain of the bacteriophage T4 short tail fibre reveals a knitted trimeric metal-binding fold"

E. Thomassen, G. Gielen, M. Schütz, G. Schoehn, J.P. Abrahams, S. Miller and M.J. van Raaij. (2003). *J. Mol. Biol.* **331**, 361-373.

"Area detectors in structural biology"

Plaisier JR, Koning RI, Koerten HK, van Roon AM, **Thomassen EAJ**, Kuil ME, Hendrix J, Broennimann C, Pannu NS, Abrahams JP. (2003). *Nuclear Instruments & Methods in Physics Research, section A.* **509**, 274-282.

"Mechanism of thrombin's enigmatic sodium switch revealed"

Abrahams JP, **Thomassen EAJ**. (2003). *Structure.* **11** (4), 363-364.

"Crystallization and preliminary X-ray crystallographic studies on a Kunitz-type Potato Serine Protease Inhibitor"

E.A.J. Thomassen, L. Pouvreau, H. Gruppen, and J.P. Abrahams. (2004). *Acta. Cryst.* **D60**, 1464-1466.

"The impact of single amino acid substitutions in CD3 γ on the CD3 $\epsilon\gamma$ interaction and T cell receptor-CD3 complex formation"

E.A.J. Thomassen, E.H.A. Spaenij-Dekking, A. Thompson, K.L. Franken, Ö. Sanal, J.P. Abrahams, M.J.D. van Tol, and F. Koning.
submitted

"The protein structures of recombinant human lactoferrin produced in the milk of transgenic cows and human milk-derived lactoferrin are identical"

E.A.J Thomassen, H.A. van Veen, P.H.C. van Berkel, J.H. Nuijens and J.P Abrahams,
Accepted for publication in Transgenic Research (in revised form).

Curriculum Vitae

Ellen Thomassen werd op 6 september 1974 geboren te Meerssen. Na het behalen van het MAVO diploma aan de Koningin Gerberga MAVO te Meerssen begon zij in 1990 aan het MLO in Sittard, studierichting analytische chemie. In 1993 startte zij met de studie Biochemie aan het HLO eveneens gelegen te Sittard. In 1996 werd stage gelopen bij Metrohm AG in Herisau Zwitserland in samenwerking met het Technikum Winterthur onder begeleiding van Dr. A. Amrein, Ing. D. Strohm en Ir. B. Wauben. Tijdens deze stage werd onderzoek gedaan naar de ontwikkeling en karakterisatie van een nieuw prototype voor de coulometrische waterbepaling volgens Karl Fisher. Mede door de resultaten van dit onderzoek kon enige tijd later de nieuwe KF756 coulometer op de markt gebracht worden. Tevens werd in Zwitserland het ingenieurs diploma verworven aan het Technikum Winterthur. Na terugkomst in Nederland werd er in 1997 begonnen met de studie Biochemie aan de Universiteit Leiden. Hier werd stage gelopen in de vakgroep Biofysische Structuurchemie (BFSC) onder begeleiding van Prof. Dr J.P. Abrahams. Tijdens deze stage werd er onderzoek gedaan naar het vetzuursynthase uit *Saccharomyces cerevisiae*. Het diploma van de Universiteit Leiden werd in 1999 verkregen.

Van 21 september 1999 tot 21 september 2004 was zij werkzaam als assistent in opleiding in de groep biofysische structuurchemie van de Universiteit Leiden. De resultaten van het onderzoek staan beschreven in dit proefschrift.

Sinds 1 november 2004 is zij werkzaam als postdoc in de groep biofysische structuurchemie van de Universiteit Leiden.

Nawoord

Het schrijven van een proefschrift in je eentje is onmogelijk. De samenwerking met anderen is van cruciaal belang. Daarom wil ik graag enkele woorden richten tot de personen die mij op één of andere manier geholpen hebben deze aio-periode onvergetelijk te maken. Aan het begin van dit project werd gestart in de groep Koning. Ook al waren mijn voordrachten over een totaal ander onderwerp dan jullie gewend waren, jullie waren altijd enthousiast en bereid iets nieuws te leren. Hier wil ik graag enkele mensen noemen. Yvonne, je hebt me iets laten zien van de wereld die celkweek heet. Liesbeth, het was plezierig met je samen te werken in het "CD3 γ -mutanten" project, het duurde uiteindelijk iets langer dan verwacht. Allan, het "namen-probleem" is waarschijnlijk verholpen sinds ik in het Gorlaeus zit. Willemijn, ook al ben je enige tijd geleden al vertrokken uit L320, ik ben erg blij dat je altijd bereikbaar bent voor een goed gesprek of een ontspannend dagje in de sauna (het gaat ons misschien nog eens lukken).

Na een verhuizing richting het Gorlaeus nam ik mijn intrek in L005 bij de groep biofysische structuurchemie. Hier was er altijd plaats voor een lach en een liedje naast de wetenschap (Stil!! RAG). Collega's en oud-collega's, jullie waren altijd in voor een bezoekje aan de borrelbunker of een hapje en drankje in de stad waar de nodige discussies plaatsvonden over het werk. De studenten Gerrit en Rinus, door jullie enthousiaste medewerking tijdens jullie stage bij BFSC is hoofdstuk 3 tot stand gekomen. Jasper, door jouw toedoen ben ik waarschijnlijk een van de weinige Limburgers die de liedjes van Rubberen Robbie kan meezingen. Marike, na de nodige stress op het werk was het altijd heerlijk om mijn agressie kwijt te kunnen door op zo'n klein balletje te slaan (sorry, voor de momenten dat het balletje jou een mooie blauwe plek bezorgde). Alle vrienden, volleyballers en "Nare-geidten", jullie waren onmisbaar voor de nodige ontspanning en inspanning gedurende mijn aio-periode. Hopelijk mag ik hier nog lang van genieten.

En niet in de laatste plaats mijn ouders, jullie stonden altijd voor me klaar ook al had ik soms rare plannen tijdens mijn studie. Zoals jullie zien is er toch nog wat van mij terechtgekomen zonder HAVO-diploma en MLO-stage. Ik ben heel blij met jullie onvoorwaardelijke steun en vertrouwen. Ik hoop dat jullie na het lezen van dit proefschrift iets meer begrijpen waar ik mee bezig ben geweest de laatste jaren.

Ellen

