THE MHC-GLYCOPEPTIDE-T CELL INTERACTION IN COLLAGEN INDUCED ARTHRITIS

- a study using glycopeptides, isosteres and statistical molecular design in a mouse model for rheumatoid arthritis

by

LOTTA HOLM

Akademisk avhandling

Som med tillstånd av rektorsämbetet vid Umeå universitet för erhållande av Filosofie Doktorsexamen vid Teknisk–Naturvetenskapliga fakulteten, framlägges till offentlig granskning vid Kemiska institutionen, Umeå universitet, sal KB3B1, KBC, lördagen den 11 november 2006, kl. 10.00. Fakultetsopponent: Doc. Morten Grötli, Läkemedelskemi, Kemiska institutionen, Göteborgs universitet, Göteborg, Sverige
Title
The MHC-glycopeptide-T cell interaction in collagen induced arthritis - a study using glycopeptides, isosteres and statistical molecular design in a mouse model for rheumatoid arthritis

Author
Lotta Holm, Department of Chemistry, Organic Chemistry, Umeå University, SE – 901 87, Umeå, Sweden.

Abstract
Rheumatoid arthritis (RA) is an autoimmune disease affecting approximately 1% of the population in the western world. It is characterised by a tissue specific attack of cartilage in peripheral joints. Collagen induced arthritis (CIA) is one of the most commonly used animal models for (RA), with similar symptoms and histopathology. CIA is induced by immunisation of mice with type II collagen (CII), and the immunodominant part was previously found to be located between residues 256-270. This thesis describes the interaction between the MHC molecule, glycopeptide antigens from CII and the T cells that is essential in development of CIA. The glycopeptide properties for binding to the mouse MHC molecule A\textsuperscript{q} have been studied, as well as interaction points in the glycopeptide that are critical for stimulation of a T-cell response.

The thesis is based on five studies. In the first paper the minimal glycopeptide core, that is required for binding to the A\textsuperscript{q} molecule while still giving a full T cell response was determined. The second paper studied the roles of amino acid side-chains and a backbone amide bond as T-cell contact points. In the third paper the hydrogen bond donor-acceptor characteristics of the 4-OH galactose hydroxyl group of the glycopeptide was studied in detail. In the fourth paper we established a structure activity relationship (QSAR model) for (glyco)peptide binding to the A\textsuperscript{q} molecule. Finally, the stereochemical requirements for glycopeptide binding to the A\textsuperscript{q} molecule and for T-cell recognition was studied in the fifth paper.

The study was performed using collagen glycopeptide analogues, which were synthesised on solid phase. Amide bond and hydroxyl group isosteres were introduced for study of hydrogen bond donor-acceptor characteristics. Statistical methods were used to design a representative peptide test set and in establishing a QSAR model.

The results give a deeper understanding of the interactions involved in the ternary MHC-glycopeptide-T cell complex. This information contributes to research directed towards finding new treatments for RA.

Keywords
Rheumatoid arthritis, collagen, T-cell, class II MHC, solid phase peptide synthesis, glycopeptide, peptide isostere, PCA, statistical molecular design, PLS, QSAR, sequence binding motif

## Contents

1. List of Papers ......................................................... 7
2. List of Abbreviations .................................................. 9
3. Introduction .......................................................... 11
   3.1. Immunological response towards antigens ................. 11
   3.2. Rheumatoid arthritis ........................................... 15
   3.3. Treatments for RA ................................................ 18
   3.4. Collagen induced arthritis ..................................... 20
4. Objectives of the Thesis ............................................. 24
5. Synthetic tools ...................................................... 26
   5.1. Solid-phase peptide synthesis ................................. 26
   5.2. Peptide isosteres ................................................ 29
   5.3. Glycosylated building blocks in SPPS ......................... 30
   5.4. 4-F- and 4-Me-O-galactose .................................... 32
6. Biological tools ...................................................... 35
   6.1. Peptide binding assay(s) ......................................... 35
   6.2. T cell stimulation assay ........................................ 37
7. Computational tools .................................................. 39
   7.1. QSAR .............................................................. 39
   7.2. Molecular descriptors in QSAR ................................. 40
   7.3. Principal component analysis ................................... 42
   7.4. Selection by D-optimal design .................................. 44
   7.5. Regression methods and PLS .................................... 46
   7.6. Model validation ................................................ 47
8. Results and discussion ............................................... 50
   8.1. The minimal peptide core based on the CII glycopeptide, capable of binding to the Aq molecule as well as stimulating a full T-cell response .................................. 50
8.2. Side-chain T cell contact points of the collagen glycopeptide that are required for T-cell stimulation 52
8.3. Interactions of the backbone amide bonds 54
8.4. The importance of hydroxyl group 4 of the galactose moiety for T-cell recognition 56
8.5. A QSAR model for peptide binding to the A8 molecule 57
8.6. The stereochemical requirements for glycopeptide binding to the A8 molecule and for T-cell recognition 59

9. Concluding remarks and future prospects 62
10. Acknowledgements 65
11. References 67
1. **List of Papers**

This thesis is based on the following papers, which will be referred to in the text by the corresponding roman numerals (I-V).


IV  **L. Holm**, K. Frech, B. Dzhambazov, R. Holmdahl, J. Kihlberg and A. Linusson, Multivariate design, synthesis and biological evaluation of peptides binding to A<sup>q</sup> class II MHC molecules -a QSAR model for peptide binding in a mouse model for rheumatoid arthritis, *Submitted*

V  **L. Holm**, I. Fredriksson, B. Dzhambazov, R. Holmdahl and J. Kihlberg, Stereo-chemical requirements for collagen glycopeptide binding to A<sup>q</sup> class II MHC molecules and stimulation of T-cells obtained in a mouse model for rheumatoid arthritis, *Manuscript*

Papers I-III reprinted with kind permission from Elsevier Science.
Related papers (i-ii)

i  J. Bäcklund, A. Trescow, R. Bockermann, B. Holm, L. Holm, S. Issazadeh-Navikas, J. Kihlberg and R. Holmdahl, Glycosylation of type II collagen is of major importance for T cell tolerance and pathology in collagen-induced arthritis, European Journal of Immunology, 2002, 32, 3776-3784

ii  M. Hedenström, L. Holm, Z-Q. Yuan, H. Emténäs and J. Kihlberg, Stereoselective synthesis of \(\text{[CH}_2\text{O]}\)Ala pseudopeptides and conformational analysis of a Phe \(\text{[CH}_2\text{O]}\)Ala containing analogue of the drug desmopressin, Bioorganic & Medicinal Chemistry Letters, 2002, 12, 841-844
## 2. List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Number of components in a PCA or PLS model</td>
</tr>
<tr>
<td>APC</td>
<td>Antigen Presenting Cell</td>
</tr>
<tr>
<td>APL</td>
<td>Altered Peptide Ligand</td>
</tr>
<tr>
<td>CIA</td>
<td>Collagen Induced Arthritis</td>
</tr>
<tr>
<td>CII</td>
<td>Type II Collagen</td>
</tr>
<tr>
<td>CTL</td>
<td>cytotoxic T lymphocytes/ T killer</td>
</tr>
<tr>
<td>DIC</td>
<td>diisopropylcarbodiimide</td>
</tr>
<tr>
<td>DModX</td>
<td>Distance to the Model in X</td>
</tr>
<tr>
<td>HO-4</td>
<td>Hydroxyl group 4 of the galactose moiety</td>
</tr>
<tr>
<td>HOAt</td>
<td>7-aza-1-hydroxybenzotriazole</td>
</tr>
<tr>
<td>HOBt</td>
<td>1-hydroxy-benzotriazole</td>
</tr>
<tr>
<td>MHC</td>
<td>Major Histocompatibility Complex</td>
</tr>
<tr>
<td>PC</td>
<td>Principal Component</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
<tr>
<td>PLS</td>
<td>Projections to Latent Structures by means of Partial Least Squares</td>
</tr>
<tr>
<td>Q^2</td>
<td>goodness of prediction</td>
</tr>
<tr>
<td>QSAR</td>
<td>Quantitative Structure-Activity Relationship</td>
</tr>
<tr>
<td>R^2</td>
<td>goodness of fit</td>
</tr>
<tr>
<td>RA</td>
<td>Rheumatoid Arthritis</td>
</tr>
<tr>
<td>SAR</td>
<td>Structure-Activity Relationship</td>
</tr>
<tr>
<td>SPPS</td>
<td>Solid-Phase Peptide Synthesis</td>
</tr>
<tr>
<td>T reg</td>
<td>T regulatory cells/T suppressor</td>
</tr>
<tr>
<td>TCR</td>
<td>T cell Receptor</td>
</tr>
<tr>
<td>T_{H1}</td>
<td>Helper T cells type 1</td>
</tr>
<tr>
<td>T_{H2}</td>
<td>Helper T cells type 2</td>
</tr>
<tr>
<td>X</td>
<td>Data matrix of descriptive variables</td>
</tr>
<tr>
<td>Y</td>
<td>Data matrix of measured responses</td>
</tr>
</tbody>
</table>

Amino acids are referred to by their conventional one or three letter codes.
3. Introduction

3.1. Immunological response towards antigens

The immune system of the human body is an advanced defence system against foreign microorganisms such as bacteria, viruses, fungi and parasites. Its roles include distinguishing between the body’s own structures (self) and foreign (nonself) structures and to subsequently start and maintain an inflammatory response, resulting in elimination of the foreign microorganism. To accomplish this task, the immune system utilises two routes of defence: the innate immunity and the adaptive immunity.

The innate immunity involves the recognition of features common to many pathogens by surface receptors on tissue residing macrophages or dendritic cells. Such event further triggers the cells to engulf the microorganism (fagocytosis) and secrete cytokines and chemokines i.e. proteins with effect on other cells. This type of interaction, involving a ready set of encoded receptors for common structures of pathogens, is unspecific but allow a very rapid first line of defence, which occurs within minutes. The adaptive immune system, on the other hand, is very specific in its recognition of structures, so called antigens, by lymphocytes. There are two major types of lymphocytes, T lymphocytes (T cells) or B lymphocytes (B cells), named after their maturation in the thymus and bone marrow respectively. The initial activation of these cells on the first encounter with a specific antigen is a slow process which takes days to develope. It does, however, permit the establishment of long-lived resting memory cells, which facilitate a quick response upon reinfection, and thereby a sometimes lifelong protective immunity. Although superior in antigen recognition, the adaptive immunity relies on the non-specific innate immunity for its initiation.
Following the initial activation of macrophages and dendritic cells by the innate immune system the engulfed microorganism is processed, i.e. proteins from the microorganism are degraded into peptides, some of which are loaded on class II major histocompatibility complex (MHC II) molecules (Figure 3.1.1). The MHC II molecule is a transmembrane protein consisting of two heterodimer chains (α and β) which are non-covalently associated. The two outermost α1 and β1 immunoglobulin domains form an antigen binding cleft, which consists of a β-sheet with α-helixes on the edges. When bound the peptide adopts an extended conformation in the binding cleft, and the open ends allow peptides of varying length to bind. Side chains of the peptide, often referred to as anchor residues, which are often hydrophobic contribute to high affinity binding by fitting into complementary pockets of the binding site. Once formed the MHC-peptide complex is transported to the surface of the antigen presenting cell (APC). In addition, the cell begins to express co-stimulatory molecules (B7) following the stimulation by cytokines secreted by macrophages upon recognition of the invading pathogen. Presentation of the processed antigens by APCs in the draining lymph node to circulating naive helper T cells direct the type of inflammatory response, and which type of effector T cells that are produced. Only T cells which specifically recognise the presented antigen through its T cell receptor (TCR), and at the same time recognise the co-stimulatory molecule B7 through the CD28 molecule, will divide and mature into effector cells which most of them re-enter the circulation. Local release of cytokines and chemokines at the site of infection in the body direct the effector T lymphocytes to enter
Once encounter its highly specific antigen, the lymphocyte aid in the elimination process of the microorganism by their respective effector actions (Figure 3.1.2).

There are three main classes of effector T cells: the helper T cells 1 and 2 (T\(_{H1}\) and T\(_{H2}\)) and the cytotoxic T lymphocytes (CTL/ T killer), but a population of regulatory T cells (T\(_{H3}\) and T\(_{R1}\)) also exists. The cytokine environment experienced by the naive T cell upon antigen presentation, as well as the degree (high/low density) of antigen presentation, the peptide-MHC affinity and the strength of the antigen-TCR interaction influences the outcome to give effector T\(_{H1}\) or T\(_{H2}\) cells. The T\(_{H1}\) are promoting pro-inflammatory cell mediated responses, with release of macrophage activating cytokines (e.g. IFN-\(\gamma\), TNF-\(\alpha\)) and B cell inhibiting and regulating cytokines (e.g. TNF-\(\beta\)) upon interaction with its presented antigen, i.e. cell mediated immunity predominates. In contrast T\(_{H2}\) cells, which have a greater anti-inflammatory potential, secret cytokines (IL-4 and IL-5) which activate B cells and others (IL-10) which mostly inhibit macrophage activation, i.e. cell humoral immunity predominates. In summary, T helper cells help and regulate the action of other cells of the immune system. Stimulated B cells, for example, mature into plasma cell capable of secreting antibodies. The rediscovered T regulatory cells, also called “T suppressor” or “Treg” cells, differ in their set of secreted cytokines (such as TGF\(\beta\)), which all suppresses immune responses. When a cytokines binds to its receptor this could stimulate a signal transduction cascade that results in activation of transcription factors, subsequent gene expression and synthesis of proteins that serve as
mediators for inflammation and tissue degradation (e.g. chemokines, tissue degrading enzymes, adhesion molecules, co-stimulatory molecules etc.) The cytokine signalling from different effector cells could also inhibit each other action, and the combined responses ideally provide a balanced/adequate immuno reaction.

In cases when the pathogens, such as viruses, reside inside a cell the cell has the possibility to process and present viral antigens on class I MHC molecule (similar to MHCII) on its surface (Figure 3.1.2). When a cytotoxic T recognises its specific antigen, it responds by secreting cytotoxic chemicals that subsequently kill the infected cell. In this way surrounding cells may be rescued from subsequent infection.

Despite the diverse repertoire of T cells produced, having TCRs capable of recognising an infinite number of antigens a central and a peripheral tolerance mechanisms normally prevents attacks on self-tissue. Central tolerance is achieved in the thymus upon maturation of T cells. In a first selection only T cells with ability of recognising antigens presented by the particular set of MHC molecules expressed by the individual are retained (positive selection). In a second step lymphocytes with TCR that bind with high affinity to presented auto-antigens, thereby having potential to react with self, are eliminated (negative selection). In this way only T cells within a set affinity window are allowed to reach circulation. In addition, a peripheral tolerance is achieved in the tissues through the processes of clonal anergy, supressor T cell regulation and regulatory networks as well as by the exclusive location of the antigen. Production of co-stimulatory molecules are initially stimulated by the innate immune system upon recognition with pathogens. When a naive T cell recognises a self antigen in the absence of co-stimulatory molecules, the cell gets into anergy. Anergy is a state of unresponsiveness in which the T cell could not be re-activated, not even upon restimulation in presence of co-stimulation. Suppressor cells (Tregs), together with cytokine regulatory nets, aids in restricting the immune response. Moreover, some antigens are only seen in a restricted locations, i.e. organs such as the eye, uterus and brain that are out of reach from the immune system.

Breakage of the self-tolerance, either through lymphocyte escape of central tolerance or breakage of peripheral tolerance, might lead to an autoimmune
disease that eventually becomes chronic. A single tissue or multiple organs (systemic inflammation) could be affected. Common autoimmune diseases are for example multiple sclerosis (MS), diabetes and rheumatoid arthritis (RA).

### 3.2. Rheumatoid arthritis

**RHEUMATOID ARTHRITIS** (RA) is one of the most common autoimmune diseases affecting approximately 1% of the population. It is characterised by an autoimmune inflammation of peripheral cartilagous joints. The symptoms including swelling, stiffness and pain in the joints, and subsequent erosion of underlying bone, which might lead to deformity resulting in malfunction of the joints. RA is about two to three times as common in women as in men, with the peak incidence between the age 40 and 60. With time RA could result in substantial functional disability and a shortening of the life expectancy by 5-10 years. For RA diagnosis, at least four out of the following seven phenotypic criteria should be fulfilled:

1. Morning stiffness
2. Arthritis of three or more joints
3. Arthritis of hand joints
4. Symmetric arthritis
5. Rheumatoid noduli, i.e. subcutaneous nodule, affecting approx. 25% of the cases.
6. Serum rheumatoid factor. Auto-antibodies of IgM or IgG type, reactive against the Fc part of IgG antibodies. The rheumatoid factor is found in approx. 70% of RA patients, but also exists to some extent in other diseases and even in 10% of the healthy population.
7. Radiographic changes, i.e. bone erosion detected by X-ray.

The most prominent effect in RA is inflammation in the cartilage of joints (articular chondritis) which generally starts with arthritis of fingers (Figure 3.2.1), but feet and knees are also commonly affected. An extraarticular chondritis could effect the larynx and the nasal septum. Other tissues that also could be effected in RA are skin, blood vessels, heart, and lungs.
The pathological changes of the collagenous joint proceed through several steps over time (Figure 3.2.2).\textsuperscript{5,6} Initially the synovium, responsible for production of the synovial fluid, thickens and infiltration by inflammatory cells (mainly T cells, macrophages and plasma cells) is greatly increased. Furthermore the inflamed synovium is rich in new blood vessels, and fibrin deposits on its surface. These changes together result in soft tissue swelling. Second, the articular cartilage is destroyed and replaced by growth of pannus, i.e. a highly vascularized and granulated tissue formed by the destructive part of the synovial membrane. The degradation of cartilage is mainly mediated by enzymes called metalloproteinases. These are secreted from chondrocytes and synovial fibroblasts, which have been activated by inflammatory cytokines. At the final stage, the destruction of bone starts by activation of osteoclasts (a large bone degrading cell) thereby disrupting the balance with the osteoblasts (the bone repair cells). This stage is irreversible and if it is not halted it results in bone erosion and joint deformity.

The possible auto-antigen responsible for the initial inflammatory response in RA is unknown. Several possible auto-antigens have however been suggested, including some collagenous cartilage proteins. Both auto-antibodies\textsuperscript{7} and T cells\textsuperscript{8,9} reactive towards type II collagen (CII) have been found in RA patients, but also antibodies to other collagens CIX, CXI\textsuperscript{10} of the cartilage fibril. Moreover, elevated antibody levels have also been detected towards some non-collagenous cartilage proteins, such as aggrecan,\textsuperscript{11} and also towards non-cartilage proteins.
The genetic background of individuals influences their susceptibility to RA, with the strongest linkage being found to the MHC II region. Individual carriers of the HLA-DR4 or DR1 genes have an increased susceptibility to RA, which further implies T cell involvement in the disease. The susceptible MHC molecules share the same epitope in the otherwise hypervariable region of the binding cleft, which has led to the shared epitope hypothesis for susceptibility to arthritis. Besides the genetic linkage environmental factors such as physical and emotional stress can effect incidence.

The reason for breakage of tolerance towards self structures in RA is unknown. One common model suggests induction of inflammation through molecular mimicry. A response starts towards a bacterial or virus antigen with resemblance to an auto-antigen, and subsequently leads to cross-reactivity with structures preliminary present in the joints. The Epstein-Barr virus has been suggested as a candidate which could cause cross-reactivity to collagen. Another model is bystander activation, in which an infection, and the following inflammation, causes a destruction of the target organ, i.e. cartilage in RA, which then by chance activates autoreactive T cells.

In conclusion RA is believed to be caused by a combination of genetic and environmental factors. As Smollen states: “in a genetically predisposed host T_{H1} cells become activated by arthritogenic antigen(s) in conjugation with co-stimulatory signals and an appropriate cytokine environment.”
3.3. Treatments for RA

Although several treatments for RA have been developed, with varying degrees of effectiveness for relieving the symptoms, there is no cure as yet. Physiotherapy and associated treatments can help prevent losses of strength and function of joints, and several drug therapies are available, involving the use of non-steroidal anti-inflammatory drugs (NSAIDs) or disease-modifying antirheumatic drugs (DMARDs). NSAIDs interfere with a specific part of the pro-inflammatory cascade (the generation of prostaglandin), thereby reducing the inflammation and stiffness of the affected joint(s) and associated pain, but they do not address the underlying cause of the disease or retard joint destruction. The more effective DMARD treatments could interfere with both the inflammatory and destructive processes of RA, but usually take longer to act. DMARD treatments include use of monoclonal antibodies, soluble receptors, receptor antagonists and anti-inflammatory cytokines. New drugs are under development and new potential targets are being investigated, for example inhibition of matrix metalloproteinases, including collagenases responsible for collagen degradation, which are present at elevated levels in inflamed joints. For example, collagen peptidomimetics have been used as inhibitors of collagenases.

Future strategies using T-cell epitope-specific therapies, which avoid systemic unspecific immune suppression, have aroused great interest in recent years. Such strategies include induction of T cell tolerance to auto-antigens, elimination of auto-reactive T cells and alteration of T cell responses. Attempts are being made to induce T cell tolerance to auto-antigens by administering the antigens in a way that exploits natural peripheral self-tolerance mechanisms (e.g. orally or nasally), and the outcome is dose-dependent. If large doses are administered, apoptosis (cell death) could be induced in the antigen-specific T cells. If, on the other hand, low doses are administered, the potentially auto-reactive T cells may become anergic. No activation will then occur since co-stimulatory molecules are not normally present in peripheral tissues. Anergy is a state in which the T cell is insensitive to further stimulation, and even upon contact with a professional APC, presenting its specific antigen as well as co-stimulatory molecules, they remain unresponsive. This effect not only disarms the specific T cell, but may also block the action of other potentially auto-
reactive T cells by competing for binding to the presented antigen. The potential effectiveness of the strategy for treating auto-immune diseases has been demonstrated in mice; e.g. mice fed with insulin have been shown to have increased resistance to diabetes, while feeding mice with collagen has reduced the severity of arthritis.\textsuperscript{15, 16} However, this approach has not always had positive outcomes in clinical trials, positive effects have been found in some cases, but not others, and in some instances there have even been increases in disease severity. Furthermore, peptide drugs have several drawbacks, including their generally rapid degradation by proteases, high conformational flexibility, and low lipophilicity. A strategy that attempts to overcome these problems is to induce T cell anergy by artificial antigen-presenting cells (liposomes) without co-stimulatory molecules that can also act as drug delivery systems.\textsuperscript{17}

Another strategy to induce tolerance to auto-antigens, or at least reduce the severity in RA, is to alter T cell responses by shifting the proportions of the T cell populations. T\textsubscript{\text{H}}1 effector cells predominate in arthritic tissues, but it may be possible to shift/reduce their dominance by manipulating the cytokine balance or injecting altered peptide ligands. Since the cytokine balance at the site of presentation affects the subsequent effector function, T\textsubscript{\text{H}}2 effector function could be altered by manipulating the cytokines (by increasing levels of IL4 and IL10 for instance, which should increase the T\textsubscript{\text{H}}2-type cytokine profile.) Altered peptide ligands (APL’s) are peptides that are related, but not identical, to the immunodominant peptide. Due to the altered MHC binding properties and interaction points with the TCR of APL’s, which act as antagonists or partial agonists, the signal from the TCR could be weakened or eliminated in their presence, thereby modifying the effector outcome. For example has an APL, where a critical TCR contacting residue was substituted, shown to shift the normal pro-inflammatory T\textsubscript{\text{H}}1 dominating response to a dominating T\textsubscript{\text{H}}2 response for a peptide that normally induces an autoimmune disease (EAE, model for MS) in mice.\textsuperscript{18, 19} In another example, substitution of amino acids in an immunodominant collagen peptide resulted in decreased MHC binding and elicited an T\textsubscript{\text{H}}2 cytokine profile.\textsuperscript{20} It may also be possible to induce T regs, which could then effectively suppress the inflammatory signals of other cells. The T regs could thus also protect against other tissue-specific auto-antigens. The APL approach could also induce anergy in the T cells.\textsuperscript{21} The effector function is ultimately determined by the combined affinity/on-off rates of the TCR upon recognition of the antigen (similar to the process of thymic selection).
Finally, two other strategies that could potentially block antigen interactions and provide short-term protection could be worth mentioning. Auto-reactive T cells could be disarmed by injecting monoclonal antibodies specific for their TCRs. Alternatively, “superbinders” with very high affinity for disease-associated MHC molecules could be introduced, although this would (unfortunately) also prevent presentation of pathogenic antigens.

Due to the complexities involved (as briefly outlined above), the development of successful T cell epitope-specific therapies generally requires substantial knowledge about the auto-antigen(s) involved, including their interaction points with the respective MHC molecules and reactive T cell receptors.

### 3.4. Collagen induced arthritis

The availability of animal models is vital in research and development of drugs for complex diseases such as RA. Several animal models have been established with different levels of similarity in symptoms and histopathology as those found in RA. Arthritis in peripheral joints can, for example, be induced by immunizing mice with cartilage specific proteins (CII, CXI, aggrecan), adjuvants (CFA) or mineral oil (Pristane), and even develops spontaneously in some mouse strains.

By injection of type II collagen (CII) in rat, mouse and chicken, the animal develops collagen induced arthritis (CIA), which has become the most commonly used animal model for RA. Figure 3.4.1 shows a mouse with CIA, to the left, with characteristic swollen joints as compared to the reference mouse to the right.

Figure 3.4.1 Adapted from group Holmdahl\textsuperscript{4}
CII is the main constituent in the cartilage fibril and has a role as a structural protein, primarily providing for tensional stabilization. It consists of three identical $\alpha$-chains of about 1000 amino acids which are twisted around each other to form an extended triple helix. The repeated Gly-Xaa-Yaa motif of CII, where every third amino acid is a glycine and X and Y often are proline and hydroxyproline, respectively, strongly promotes the triple helical structure. Another common amino acid in CII is lysine, which can undergo posttranslational modifications resulting in hydroxylation of the side chain, and subsequently glycosylation with a $\beta$-D-galactopyranosyl- or an $\alpha$-D-glucopyranosyl-(1→2)-$\beta$-D-galactopyranosyl unit.

Both homologous CII, i.e. from the mouse, and heterologous CII, e.g from rat, together with Complete Freunds Adjuvant can induce CIA in mice. While heterologous CII induces a more severe arthritis with faster onset and higher antibody titers, immunisation with homologous CII induces a more chronic arthritis. Susceptibility to CIA in mice has been linked to the MHC molecules $A^q$ and $A^r$, while the $A^p$ molecule which only differs in four amino acids as compared to $A^q$ does not confer susceptibility. The MHCII genetic linkage and the fact that mice lacking $\nu\nu$ T cells are resistant to CIA strongly implies the importance of T cells for initiation of CIA. After immunisation the CII is processed and presented by $A^q$ molecules on APCs to auto-reactive T cells, which thereby start the inflammatory response, (Figure 3.4.2) with B cell stimulation and subsequent antibody production.

By cyanogen-bromide cleavage of CII the immunodominant part was restricted to be within the CB11 fragment, corresponding to 1/3 (274 aa) of the CII molecule. By comparing the sequences of chick CII which was used for introduction of CIA with the human and mouse CII sequence a set of synthetic peptides were selected and tested for T cell recognition. This further limited the region of CII that is essential for eliciting CIA to amino acids 256-270. Rat was later shown to have the same amino acid sequence as chick within this region, and the only amino acid that differed from mouse CII was position 266, with a conserved amino acid exchange from glutamic acid in rat and chick to aspartic acid in mice. Due to the different reactivities for mice and rat derived CII, position 266 was suggested to be an important residue for T cell recognition. The amino acid exchange was also found to influence the binding
THE MHC/GLYCOPEPTIDE-T CELL INTERACTION IN CIA

affinity of the peptides for the A4 molecule, which subsequently correlated to severity of disease.36 Brand and colleagues has further suggested the immunodominant core of CII to be located between 260-267 using overlapping peptides.

![Diagram of MHC class II molecule with CD4 T cell and TCR](image)

Figure 3.4.2

Earlier a panel of 29 T cell hybridomas was established after immunisation of mice with rat CII.32, 35, 37, 38 Out of the 29 studied hybridomas only 6 recognised the syntetically produced non-modified peptides containing the immunodominant core. Eventually this resulted in discovery of the importance of posttranslational modifications of lysine at position 264 for T cell stimulation (Figure 3.4.2).38 While two T cell hybridomas recognised the disaccharide (an D-glucopyranosyl-(1->2)- D-galactopyranosyl unit) located on hydroxylysine 264 in the CII peptide, the majority i.e. 20 out of 29 responded specifically to the monosaccharide (a D-galactopyranosyl unit) when attached to hydroxylysine 264.32, 39 By using a set glycopeptide analogues, deoxygenated in the D-galactose moiety the T cell hybridomas could be further sub-divided based on their fine-specificity for the different carbohydrate hydroxyl groups.40 Interestingly, this subdivision was found to correlate well with the sequences of the TCRs in the different subgroups.32 Most of the 20 hybridomas interact with the hydroxyl groups at positions 2 and 4 of the galactose moiety. Some hybridomas are also dependent on the hydroxylysine amino group, as well as the O-glycosidic bound of the galactose moiety as revealed by synthesis of C-
glycoside analogues.\textsuperscript{41, 42} An alanine scan throughout the peptide confirmed the existence of two major anchor residues to the A\textsuperscript{8} molecule within the 256-270 region; isoleucine 260 and phenylalanine 264, respectively.\textsuperscript{36}

A recent transgenic mice expressing the heterologous CII epitope with a glutamic acid in position 266 in cartilage (i.e. same as in human) has been established, which also showed susceptibility to CIA. However, once the rat CII galactosylated peptide was injected as a intraperitoneally tolerogen in newborn transgenic mice, it showed reduced incidence from subsequent arthritis induction.\textsuperscript{16} Moreover, the rat peptide in complex with the A\textsuperscript{8} molecule significantly reduced both disease progression and severity in mice with ongoing chronic relapsing arthritis.\textsuperscript{43} The immunodominant rat collagen glycopeptide i.e. shows some very promising vaccination potential.

While a detailed knowledge has been obtained of the contact points for T cell recognition of the galactose residue in the CII glycopeptides, less is known about the size and contact points of the peptide part. What is, for instance, the contribution of individual amino acids in the peptide for A\textsuperscript{8} binding and T cell response? Which modifications would be tolerable in the peptide part and in carbohydrate hydroxyl groups for T cell stimulation, and which could not be altered? Which part of the glycopeptide could be altered in attempts to make the peptide more drug-like? Answering these and related questions would provide a closer understanding of the MHC–glycopeptide-T cell interaction, which would be beneficial in understanding the response to auto-antigens in RA and could, in the long term, lead to improved treatments.
4. Objectives of the Thesis

The goal of this thesis is to study the MHC-glycopeptide-T cell three-component system which is associated with development of CIA in mice, a common animal model for RA. A closer understanding and knowledge about the interactions involved in the glycopeptide-MHC binding, as well as the important T-cell contact points in the glycopeptide, would provide a basis for future drug-design strategies in T cell epitope-specific immune therapy. The thesis elucidates the binding properties to \( \text{A}^\text{q} \) class II MHC molecules and the T-cell interactions points of the \( \text{A}^\text{q} \)-bound immunodominant glycopeptide found within residues 256-270 of type II collagen.

The specific aims include (Figure 4.1):

- Finding the shortest peptide within the CII256-270 glycopeptide capable of binding to the \( \text{A}^\text{q} \) molecule as well as stimulating a full T-cell response. (I)
- Evaluating the side-chain T-cell contact points of the collagen glycopeptide that are required for T-cell stimulation. (II)
- Studying the interactions of the isoleucine 260 alanine 261 backbone amide bond of the glycopeptide with the \( \text{A}^\text{q} \) molecule and T-cell receptor, by introduction of a peptide isostere. (II)
- Studying the importance, in terms of hydrogen bond donor-acceptor characteristics, of hydroxyl group 4 of the galactose moiety for T-cell recognition. (III)
- Establishing a QSAR model for (glyco)peptide binding to the \( \text{A}^\text{q} \) molecule. (IV)
- Studying the stereo-chemical requirements for glycopeptide binding to the \( \text{A}^\text{q} \) molecule and for T-cell recognition. (V)
THE MHC-GLYCOPEPTIDE-T CELL INTERACTION IN CIA

Figure 4.1
5. Synthetic tools

Invaluable information can be obtained by investigating biological hypotheses using tools provided by synthetic organic chemistry. As a result, old borders between traditional disciplines such as chemistry and biology disappear as research becomes more interdisciplinary. This chapter provides a brief description of the synthetic tools used to probe the main biological goals of this thesis; the peptide interaction and the T cell response to that complex.

5.1. Solid-phase peptide synthesis

Peptide synthesis in solution was first described by Emil Fisher in the early 20th century. It took another sixty years before the first peptide was synthesised on solid-phase, a development which eventually enabled faster synthesis, preparation of peptides in high yields and purity and also allowed atomisation. Solid-Phase Peptide Synthesis (SPPS) has now become a widely used method, which is described extensively in the literature.

In SPPS peptides are synthesised in a linear and stepwise way from the C-terminus, by attaching ready made amino acid building blocks to a growing peptide chain anchored to a solid support (Figure 5.1.1). This offers great advantages as compared to synthesis in solution. Reactions can easily be performed with excess reagents, driving couplings to completion and shortening reaction times. Excesses of reagents are then removed by a simple filtration after each coupling thereby avoiding difficult and time-consuming purification steps. The solid supports (resins) are insoluble polymeres that form a network and carry handles, called linkers, for peptide attachment. The exact chemical composition of the resin influences loading capacities and swelling properties in different solvents used in the reactions. A linker is used to attach the first amino acid to the resin and later allows the final peptide to be cleaved from the resin. While some linkers, e.g. the Wang linker, renders a peptide having a C-
terminal carboxylic acid after cleavage, others such as the RINK linker are cleaved to give a C-terminal amide. In a repeated cycle, a large access of the activated N\textsuperscript{\text{\text{\O}}}protected amino acid is added to the resin carrying the growing peptide chain.

![Figure 5.1.1](image)

Amino acid side chain functionalities are protected were appropriate to avoid unwanted side reactions. Coupling reagents are used to activate the carboxylic acid in the amide formation reaction. For standard conditions, couplings are performed with 1-hydroxy-benzotriazole (HO\text{Bt})\textsuperscript{50} or 7-aza-1-hydroxybenzotriazole (HO\text{At})\textsuperscript{51} which are used together with carbodiimides (DCC or DIC), or incorporated in uronium salts (HBTU, HATU) (Figure 5.1.2).

It should be pointed out that a broad range of coupling reagents exists that have different efficiency and provide possibilities for handling difficult couplings. Couplings generally get more difficult upon increasing peptide length, especially for coupling of hydrophobic and bulky amino acids. This is assumed to be due to that secondary structures start forming and also because of aggregation between chains. Coupling reaction could be monitored visually for example by addition of the indicator bromophenolblue which detects free amines,\textsuperscript{52} or by \textsuperscript{19}F-NMR spectroscopy using fluorinated linkers and building blocks. The terminal N\textsuperscript{\text{\text{\O}}}
protecting group is removed after completion of each coupling cycle, leaving the side chain protecting groups unaffected (Figure 5.1.1).

Two main protection groups protocols have been used in SPPS; Merrifield’s “Boc-strategy” where the α-amino group is protected by a tert-butyloxycarbonyl group, or the “Fmoc-strategy” based on the 9-fluorenylmethoxycarbonyl group. The Boc group is cleaved with the moderately strong acid TFA, while the side chain protective groups and the linker are cleaved with the stronger HF. Since both types of protecting groups are cleaved under acidic conditions the repeated TFA treatments, required in synthesis of longer peptides, could result in significant, accumulated cleavage from the resin. The more popular Fmoc-strategy uses orthogonal protective groups, where cleavage of the base labile Fmoc group leaves the acid labile side chain protecting groups and linker intact. Bases such as piperidine or morpholine are used to remove the Fmoc group, whereas simultaneous side-chain deprotection and cleavage from the resin are performed with TFA after completion of all coupling cycles. The crude peptides are in general purified by reversed phase high performance liquid chromatography (HPLC).

Figure 5.1.2
In Papers I-V all peptides were prepared using the Fmoc-strategy in a manually operated reactor, or in automated peptide synthesizer. The DIC/HOBt combination or HBTU was used for standard couplings and HOAt/DIC or HATU for difficult couplings. Synthesis was performed by the stepwise approach, which facilitated substitution of one amino acid by another in synthesis of peptide analogues. Peptides with single amino acid replacements as compared to the wt collagen peptide, containing non-coded D-amino acids, pre-made isosteric building blocks (section 5.2) or glycosylated amino acids (section 5.3, 5.4) was synthesised.

5.2. Peptide isosteres

Peptides are generally considered as poor drug candidates due to their metabolic instability, poor absorption through membranes and high flexibility. These drawbacks may be overcome by conversion of peptides into peptidomimetics, i.e. “chemical structures derived from peptides that imitate desired properties of the natural molecules.”\textsuperscript{53} The mimetic could for example be designed to prevent degradation by proteolytic enzymes, to increase the bioavailability or to reduce the flexibility to the desirable active conformation,\textsuperscript{54} thereby preventing multiple biological responses and decreasing the entropy-cost upon receptor binding. Peptidomimetics may also be used to study interactions of the original peptide, for example the involvement of an amide bond in hydrogen bonding to a receptor and the biological response elicited through that interaction. Several modifications of the peptide backbone have been reported, including isosteric exchange of atoms or units, or the introduction of additional fragments. The exchange could include the NH, CO or C\textsuperscript{a} groups of the peptide chain, or a whole fragment of the peptide backbone. Some examples are the azapeptides (exchange of C\textsuperscript{a} for a nitrogen atom), N-metylation of the amide NH or peptides in which the amide CO has been reduced- or converted into a sulfino- or phosphinogroup.\textsuperscript{54} An excellent review on amide bound isosters are provided by Venkatesan and Kim 2002, and some of the presented and most common amide bond isosters are shown in figure 5.2.1. Examples of other peptidomimetics developed in our group include mimetics of [\textsuperscript{55}, 56] and [\textsuperscript{57}] turns,\textsuperscript{57} and also [\textsuperscript{58}, 59] strand mimetics based on a substituted pyridine scaffold.
In Paper III the methylene ether isoster (Figure 5.2.1) was used to study hydrogen bonding between the peptide backbone in CII 260-267 and Aβ, while at the same time providing stabilisation towards enzymatic degradation. The methylene ether isostere has a close geometrical resemblance to the amide bond in the extended conformation, and it may be used for replacement of amide bonds without having a major effect on the structure of the substituted peptide.

The methylene ether isostere has been incorporated in several bioactive peptides (substance P, enkephaline, desmopressin). None of the normal hydrogen bond donor and acceptor characteristics are retained by the methylene ether isostere as the carbonyl group (acceptor) is replaced by a methylene group, and the amine (donor) is replaced by the oxygen atom (acceptor if not to sterically hindered). The methylene ether isoster may be synthesised as a dipeptide building block by a simple Williamson’s ether synthesis, provided that the C-terminal amino acid is restricted to glycine and alanine equivalents.

![Diagram of amide, N-alkyl, ketomethylene, "reduced amide", retro inverso, carba, E-alkene, methylene ether]

Figure 5.2.1

5.3. Glycosylated building blocks in SPPS

Posttranslational modifications, such as hydroxylation and subsequent glycosylations, are common in collagen and are accomplished by enzymatic glycosyl transferases in vivo. While N-linked glycans are linked to proteins via the N of asparagine, O-linked glycans are connected via the hydroxyl group in serine, threonine or hydroxylysine. In vitro, enzymatic glycosylations may only be accomplished where appropriate glycosyltransferases and glycosyl donors are available. A more common
approach is to prepare glycopeptides by organic synthesis, but a combined chemical/enzymatic approach may be used for attachment of larger carbohydrate. SPPS is not only excellent for introduction of non-natural amino acids and modifications, such as isosteric fragments, into peptides but also for introduction of carbohydrate moieties thereby forming glycopeptides. This may be accomplished by either of two strategies; glycosylation of the completed peptide, performed before cleavage from the resin or in solution after cleavage (the direct condensation strategy), or by using a glycosylated amino acid in peptide synthesis (the building block strategy). The building block strategy has been used more commonly, and generally simplifies synthesis with for example reduction of protecting groups manipulations, avoidance of secondary and tertiary structures that prevent glycosylation, and eliminating the need for difficult separations of anomic mixtures of glycosylated peptides.

The Fmoc strategy is preferable in glycopeptide synthesis because the continuous (TFA) and harsh (HF) acidic treatments used in the Boc strategy may result in cleavage of the carbohydrate moiety from the glycopeptide. The major problem in chemical glycoside synthesis is to generate the correct stereochemistry for the glycosidic linkage. In the building block approach this difficult linkage is created before insertion of the glycosylated amino acid into the peptide. In the synthesis of O-linked glycosides a glycosyl donor with a leaving group at the anomic position is converted to an electrophilic oxocarbenium ion by a promotor. The glycosyl acceptor, i.e. the hydroxylated amino acid, attacks from either side of the donor thereby forming an or -glycosidic bond. Some general principles may be used to affect the or -outcome, for example use of participating group at C-2 of the donor, the anomic effect, or solvent effects. The glycosyl donors are most commonly activated by Lewis acids or via heterogenous catalysis. Glycosyl bromides may be activated using an insoluble silver silicate promotor because of the high affinity of silver for bromine. This approach was further developed by B. Holm, et al. for introduction of the -O-glycosidic linkage in galactosylated hydroxylysine building blocks.

In short, acetobromogalactose was reacted with protected hydroxylysine ((5R)-N-(Fluoren-9-ylmethoxycarbonyl)-N-benzyloxy-carbonyl-5-hydroxy-L-lysine Allyl Ester) under promotion by silver silicate promotor to give the desired -O-glycoside in high yield. (Figure 5.3.1) The acetyl protective groups of the
THE MHC-GLYCOPEPTIDE-T CELL INTERACTION IN CIA

galactose moiety not only function as participating group that promote \(\beta\)-glycoside formation, but also stabilise the glycosidic linkage towards acids.\(^{70}\) The reaction could take place through a \(S_N2\) or \(S_N1\) pathway, where reaction through the latter might permit some side reaction resulting in orthoester formation. The orthoester formation are luckily reversible and could often be rearranged under acidic conditions. After deallylation, the glycosylated building block was ready for insertion into collagen glycopeptides by the SPPS Fmoc strategy. A final base catalysed deprotection of the carbohydrate acetyl protecting groups is carried out with diluted sodium methoxide under conditions which do not result in epimerization of peptide stereocenters.\(^{71}\)

In CIA, most of the carbohydrate specific T cell hybridomas, recognise the posttranslationally modified collagen peptide having a galactosyl moiety on the central hydroxylysine 264.\(^{38}\) In Papers I-III and V silver silicate promoted glycosylation was used to synthesise the \(\beta\)-O-linked galactosylated hydroxylysine building blocks that were then introduced in the glycopeptides using SPPS.

![Figure 5.3.1](image-url)

**5.4. 4-F- and 4-Me-O-galactose**

The hydrogen bond interaction is very common in biological systems due to the large number of possible hydrogen bond donors and acceptors found in biomacromolecules. This special case of the dipole/dipole attraction, observed between small electronegative atoms and hydrogen atoms attached to electronegative atoms has a bond strength.
corresponding to 7-40 kJ/mol. Hydrogen atoms attached to oxygen and nitrogen may form hydrogens bond with oxygen, nitrogen and even flavor atoms functioning as hydrogen bond acceptors. It should, however, be pointed out that flavor seldom is involved in hydrogen bonding. The hydroxyl group, which is a dominant feature of carbohydrates, acts both as a hydrogen bond donor and acceptor. Its role in a specific interaction, e.g. in receptor binding, can be studied by remove, i.e. deoxygenation, or further transformation. Common transformations to hydroxyl group isosteres, include esters, ethers or conversion to fluorides.

In **Paper III** the 4-hydroxyl group of the galactose moiety in immunodominant glycopeptide CII259-273 was altered to 4-flouro and 4-methoxy groups, to investigate hydrogen bond donor-acceptor properties of importance for T cell stimulation (Figure 5.4.1). The O-methylated isostere has lost its hydrogen donor bond possibility, introduces some steric hindrance but may still function as a hydrogen bond acceptor. The flourine isostere mimics the hydroxyl group in size and polarization of the bond to C-4. It has lost its hydrogen bond donor properties but may occasionally work as a hydrogen bond acceptor. Obviously, deoxygenation prevents both hydrogen bond donation and acceptance.

![Figure 5.4.1](image_url)

When incorporating hydroxyl groups isosteres in carbohydrates one needs to develop a protecting group strategy leaving only the desired hydroxyl group accessible for transformation. In addition, a saccharide with appropriate stereochemistry at the chosen position needs to be selected as starting material. The protecting group strategy used for preparation of 4-deoxygenated galactose, could also be employed for synthesis of the 4-flouro and 4-methoxy analogues. Methylation of 2,3,6-tri-O-benzyl protected methyl galactoside using sodium hydride and methyl iodide provided the 4-methoxy analogue (Figure 5.4.2). The flourine atom was introduced by conversion of the free 4-hydroxyl...
group in methyl 2,3,6-tri-O-benzyl \([-\text{D-glucopyranoside} \text{ into a trifluoromethane sulfonate. This was further substituted for fluorine, with inversion of configuration, using tris-(dimethylamino)sulfonium difluoro-trimethylsилicate(TASF). After changing to O-acetyl protecting groups and conversion to glycosyl bromides the acquired glycosyl donors could be used in silver silicate promoted glycosylation of a protected derivative of hydroxylysine (section 5.3).}

\[
\begin{align*}
\text{a) } & \text{NaH, MeI, DMF} \\
\text{b) } & \text{H}_2 (5 \text{atm}), \text{Pd-C, HOAc} \\
\text{c) } & \text{H}_2 \text{SO}_4, \text{HOAc/CH}_2\text{Cl}_2 \\
\text{d) } & \text{HOAc/MeI, pyridine, DMAP} \\
\end{align*}
\]

Figure 5.4.2
6. Biological tools

Assays can be set up in numerous ways when setting out to study interactions in biological systems. For example, one has to consider the availability of cells, proteins and other reagents, the type of interaction to base the assay on, speed and throughput as well as detection techniques. This chapter provides a brief description of the assays used to probe the main biological goals of this thesis; the peptide A\(^3\) interaction and the T cell response to that complex.

6.1. Peptide binding assay(s)

Three closely related versions of a competition assay between a tracer peptide and the (glyco)peptide under evaluation was performed to test peptide binding to the A\(^3\) molecule. The competition assay approach was chosen since A\(^3\) molecules are very unstable in the empty state. An additional advantage is that the use of a tracer peptide removes the need to attach a label for detection to all the different test peptides, which would be synthetically time-consuming and also could interfere with the binding to A\(^3\). The three variants of the assay utilised isolated, recombinant and cell surface located A\(^3\) molecules, respectively.

In Paper I and II peptide binding to isolated, affinity purified A\(^4\) molecules was studied. The A\(^3\) molecules were purified from mouse spleen APCs using affinity-chromatography, which required the transmembrane region of A\(^4\) to be protected by a detergent.\(^{36}\) This purification method depends on that the antibodies used for affinity-chromatography are of high quality. Attempts to purify A\(^3\) using preparative native gel failed due to decomposition of the \(\beta\)-chain. Peptide binding was then determined by incubation of a fixed concentration of a biotinylated CLIP peptide (tracer peptide) and different concentrations of the test peptides together with the A\(^4\) molecules (Figure 6.1.1A). The resulting A\(^3\)-peptide complexes were captured by antibodies coated in 96-well microtiter plates. Europium (Eu\(^{3+}\)) labelled streptavidin was added,
and the extremely high affinity of streptavidin for biotin allowed a strong complex to form with the $A^9$ bound, biotinylated CLIP peptides. The bound $Eu^{3+}$ was further detected using the Dissociation-Enhanced Lanthanide FluoroImmunoAssay, commercially available as the DELFIA kit. The resulting micelle fluorescence signal is relatively long-lived, which allows decline of short-lived unspecific fluorescence due to background emission in biological samples before detection. This time-resolved fluorescence provides a very selective and sensitive detection method, as compared to use of normal fluorescence. Quantification of the decrease in fluorescence upon incubation with increasing concentrations of test peptide allows the IC50-values for binding of the peptides to $A^9$ to be determined.

**Figure 6.1.1**

During the progress of the project described in this thesis, a recombinant $A^9$ molecule ($rA^9$) was generated by Dzambazov et al., which provided access to soluble $A^9$ molecules. The molecules could be cultured in insects cells and took away the need for using laboratory mice as source for $A^9$. One disadvantage was
that the rAq turned out to be even more sensitive to aggregation and denaturation than native Aq molecules. However, the same type of competition assay could be performed, as described above, but now using rAq molecules instead of native Aq (Figure 6.1.1B). The peptide-Aq interactions in Paper V, and the validation set in Paper IV, were studied using rAq.

In Paper I and IV, peptide-Aq binding was studied using whole cells rich in Aq molecules (Figure 6.1.2). The cells offered an easy access to Aq with no need for time-consuming purification, or use of laboratory animals. In analogy with the other assays described above competition between binding of a biotinylated CLIP tracer peptide and the test-peptide to cell bound Aq molecule was determined. Phycoerytrin (PE) labelled streptavidin was used to detect the biotinylated CLIP peptide. The PE fluorescent dye absorbs green light while emitting red light as detected in a flow cytometry analysis. While avoiding purification of Aq molecules, this cell-based assay is not as sensitive as the time-resolved fluorescence assay and generally results in higher background fluorescence.

### 6.2. T cell stimulation assay

The specificity of antigen recognition by Aq restricted T cells can be evaluated using T cell hybridomas. A set of collagen specific T cell hybridomas has been established earlier by fusion of tumor cells and collagen sensitive T cell clones. The long lived T cell hybridoma retains the
THE MHC-GLYCOPEPTIDE-T CELL INTERACTION IN CIA

antigen specificity of the original T cell clone and responds by secretion of interleukin-2 (IL2) into the surrounding media upon TCR mediated recognition of antigen. In order to study the specificity of the TCR-peptide-A\(^3\) interaction T cell hybridomas were co-cultered with mouse spleen APC’s that carry A\(^3\) on their surface and different concentrations of the (glyco)peptide under evaluation (Figure 6.2.1A). The IL2 secreted by the hybridomas was quantified using antibody detection (Paper III) or the CTLL T-cell clone (Papers I-III, V),\(^{76}\) which depend on IL2 for growth (Figure 6.2.1B). CTLL growth was then detected by incorporation of \(^3\)H thymidin in their DNA, as quantified by radioactivity measurement.

Figure 6.2.1
7. Computational tools

Besides the synthetic and biological tools used to study the MHC-glycopeptide-T cell interaction, could statistical/computational tools be used to pick informative molecules to synthesise and test (diverse and adequate sets of peptides) and also for aid in evaluating data. This chapter provides a brief description of the computational tools used to probe the relationship between a structure of a peptide and its possibility to bind to the A^{4} molecule.

7.1. QSAR

In order to correlate molecular structures with a biological response of interest a structure-activity relationship (SAR) can be established. The assumption underlying this approach is that changes in the biological response are related to changes in the chemical structure of the molecules involved. Hence studying a targeted biological response to a set of molecules with systematically varied chemical structures could reveal the SAR of substances that induce the response. If the chemical structures can be described numerically, using quantitative structural descriptors, a quantitative structure-activity relationship (QSAR) may be formed and described as a mathematical equation. Such mathematical models may provide indications of the characteristics of unknown compounds that may be involved in the studied interaction(s) and facilitate attempts to design new compounds with desirable characteristics for stimulating or inhibiting the biological response.

Hansch is generally considered to be the pioneer in QSAR, i.e. in correlating biological activity to structurally related molecules. Since then QSAR’s have been successfully used in a variety of applications, including drug-discovery, toxicology studies, and the early peptide structure activity analyses started around 1960.
When designing a QSAR study one needs to select a well-balanced set of molecules, ideally by statistical methods. The data must be homologous and “act on local basis” i.e. all of the compounds should interact with the same target by the same mechanism. In addition, both the described chemical structure characteristics, and the measured responses must be relevant to the studied biological phenomenon. The selection of peptides to be synthesised and tested can be based on either the building-block (amino-acid) space or the final products (peptides). There are no difference in efficiency between the two approaches, provided that the building block space is carefully investigated and the diversity of the selected compounds is both adequate and appropriate. However, the building block approach generates QSAR data that are easier to interpret and use, since molecular characteristics of the amino acids at specific positions in the peptide can be correlated to the biological response. The development of a QSAR could be divided into six steps: 1. Numerical description of the chemical structures (sections 7.2 and 7.3). 2. Selection of the peptides to be synthesised and tested (section 7.4) 3. Synthesis and biological testing (chapter 5-6) 4. Construction of a mathematical model correlating chemical structure with the investigated biological response (section 7.5) 5. Verification of the model (section 7.6) and 6. Use of the model to postulate new compounds with desirable characteristics.

In Paper IV a QSAR model was established which related peptide structure characteristics to the binding affinity for the \( A^b \) molecule. The design was made in building-block to allow the influence of amino acids with specific characteristics at specific positions to be interpreted.

### 7.2. Molecular descriptors in QSAR

Structural descriptors are used to describe different molecular properties in terms of mathematically interpretative numbers. When used to describe molecules for subsequent QSAR analysis, the structural descriptors should collectively be able to describe properties that are relevant to the studied biological activity. If protein-ligand interactions are investigated important properties generally include hydrogen-bonding, and hydrophobic, electrostatic and steric interactions, i.e. characteristics that influence binding.
In the first published peptide QSAR study that demonstrated a significant predictive capability, the descriptors used to describe amino acids were based on 20 experimentally determined physical-chemical parameters, including MW, pKa, NMR shifts, and hydrophobicity-hydrophilicity scales. The dimensionality of the data was compressed using principal component analysis (section 7.3), whereby the 20 descriptor values for each amino acid were compressed to a reduced number of values (score-values) that retained most of the relevant information. These new values effectively described the latent principal properties of the molecules’ size and hydrophobicity/hydrophilicity and were referred to as “z-scales”. In several subsequent studies these z scales were improved by adding more experimentally determined descriptors (such as HPLC and TLC retention times) and calculated descriptors (e.g. surface area, HOMO-LUMO, atomic charge, dipole and surface area) and expanded by including non-coded amino acids.

A recent comparison between a set of physico-chemical and calculated descriptors revealed that they contained similar latent information. However, the calculated descriptors have several advantages, notably reducing the need for experimental testing, which saves time and money. Using calculated descriptors it is also possible to postulate characteristics of compounds that may not even have been synthesised to date. Today various computer programs that are capable of generating calculated descriptors are available. Most importantly, the majority of the chosen descriptors must be interpretable with regard to the purpose of the investigation. Once a QSAR has been established it should be possible to translate “backwards” from the numerical values to chemical structures, and thus allow the prediction of new molecular structures with desirable biological effects.

In Paper V twenty-eight calculated molecular descriptors were used to describe the twenty coded amino acids (Table 7.2.1). The descriptors describe properties such as size, polarity, lipophilicity, shape, flexibility and hydrogen-bonding properties, all of which are important in protein-peptide interactions. The data were further compressed (using principal component analysis) to underlying principal properties.
7.3. Principal component analysis

PRINCIPAL COMPONENT ANALYSIS (PCA) is a multivariate projection method, that has been thoroughly described in the literature. From a set of original data a new set of latent variables can be generated by PCA that describe the underlying variation of the original data. PCA has proved to be very useful in diverse disciplines, ranging from forestry to pharmaceutical sciences, and in a huge range of investigations involving multivariate characterisation, process modelling and/or classification due to its power to detect latent trends and grouping in data. Furthermore, the detected latent variables can be used as design variables in multivariate design for diversity-based compound selection.

Briefly, a multivariate data matrix \( \mathbf{X} \) (Figure 7.3.1) representing \( N \) observations of \( K \) variables can be viewed graphically as \( N \) points in \( K \)-dimensional space. The matrix may be either “long and lean” \( (N>K) \) or “short and fat” \( (K>N) \), and if there are many variables they are often correlated to each other. PCA can then
be used to compress the data and extract variation in $X$ by generating a set of new latent variables. The $N$ observations in $K$-space are thereby projected onto a straight line/vector ($A=1$) or $A$ dimensional plane so that the squared residuals are minimised i.e. the maximum variation of $X$ is explained. The first vector, referred to as principal component 1 (PC1) describes the largest variation in $X$. The second principal component (PC2) is placed orthogonal to the first and explains the second greatest degree of variation and so on.

In many data sets only a small number of PCs are required to explain most of the variation in $X$. The observations projected onto the PC line/plane are assigned new values called scores ($t$) and loadings ($p$). While the score value denotes its position on the PC’s hyperplane, loadings ($p$) describe the relative contribution of each variable to the different principal components and thus the cause of separation in scores. Individual plots of scores and loading planes can be superimposed, facilitating interpretation of the cause of explained variation in $X$. Groupings and trends in the observations (and differences and similarities in the structures) can then be studied. A PCA can be calculated using the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm. NIPALS approaches are considered to be more rapid than other methods, where all components are calculated simultaneously, and they can handle some degree of missing data.

Mathematically the new PC-based axes can be expressed as follows:

$$X = 1x + t_1 p'_1 + t_2 p'_2 + t_3 p'_3 + \ldots + t_A p'_A + E$$

The data that cannot be explained by the model are referred to as residuals ($E$), and collectively form the residual matrix ($E$). Extreme outliers of single observations may “tilt” the model in the selection of PC’s, as seen in score plots, while moderate outliers can be detected by studying their distance to the model.
in X (dModX). The reasons why observations appear to be outliers should be carefully examined, and then possibly excluded from the PCA. The number of significant principal components was decided using their eigenvalues, a Scree plot and chemical interpretation of the loadings for the corresponding components.

In Paper IV PCA was used to compress 28 molecular descriptors (variables) of the 20 coded amino acids (objects) to just 3-5 orthogonal principal components. Prior to PCA, the molecular descriptors were scaled to unit variance to give them all equal importance in the PCA and avoid bias due to differences in the numerical ranges covered by the molecular descriptors. Amino acids with similar characteristics group together in a t1/t2 score plot, based on the two PC’s describing the most variation (Figure 7.3.2). A reduced and representing number of amino acids was chosen with retained chemical diversity in the building block (amino-acid) space and thereby with diverse characteristics. The diversity in chemical diversity, which potentially account for biological diversity, are basis in creating a well balanced peptide test set critical for a successful QSAR analysis.

![Figure 7.3.2](image)

7.4. Selection by D-optimal design

A well balanced set of molecules (here peptides or amino acid building blocks) is essential for establishing a successful QSAR
model with high predictive power. This “training” set should include peptides with diverse chemical structures and ideally span a wide range of affinities for the biological target. The traditional COST approach, which considers one separate factor at a time, does not necessarily give information about the optimum biological response and usually requires tests with large numbers of molecules. Apart from the cherry-picking approach, which could be used with relatively small data sets, a well-balanced set of molecules is best selected using statistical methods (i.e. using statistical molecular design/SMD). Such methods can reveal interaction effects by changing several factors at a time and thus reduce the number of required experiments.

D-optimal design is a commonly used SMD method to select from all possible molecules a subset with the maximum chemical diversity covering as many different characteristics as possible. Mathematically this is performed by an algorithm that maximises the determinant of the variance-covariance matrix $X^\top X$ of the model matrix $X$. This means that a subset of molecules is chosen that spans a maximal volume of the chemical space (Figure 7.4.1). Since molecules representing the extremities of the space are picked, which by definition are extreme points, it is sometimes necessary to restrict the studied space. It is also important to include molecule(s) corresponding to (a) centre point(s) in order to detect curvatures within the data.

In the study described in Paper IV a virtually combined peptide data set consisting of 4500 peptides was reduced by D-optimal design to a well balanced subset of 22 peptides, that could be efficiently used in QSAR modelling.

![Figure 7.4.1](image-url)
7.5. Regression methods and PLS

Regression methods can be used to correlate sets of variables, also called factors \((X)\) with response data \((Y)\), to identify the relationships between \(x\) and \(y\) in terms of a mathematical model. When used to establish a QSAR model, the relationship between molecular structure variables \((X)\) and the biological response of interest \((Y)\) is studied. Two such regression methods that are commonly used include Multiple Linear Regression (MLR) and Projections to Latent Structures by means of Partial Least Squares (PLS). The latter has several advantages.\(^{80, 87}\) In MLR the variables must be independent of each other (uncorrelated), the number of factors must be fewer than the number of observations \((K<N)\) and it is assumed that there are no errors in the \(X\) data. PLS on the other hand can handle correlated data, matrices where \(K>N\) and it accounts for noise in both the \(X\) and \(Y\) data. PLS can also cope with multi-\(y\) data sets, i.e. data sets with more than one response variable.\(^{80}\)

![Diagram of PLS method](image)

Figure 7.5.1

The PLS method, introduced in the 1970’s,\(^{92, 93}\) is in many respects similar to the PCA technique described in section 7.3. The factors and responses may be viewed graphically in \(K\) and \(M\)-space, respectively (figure 7.5.1), and projected
THE MHC-GLYCOPEPTIDE-T CELL INTERACTION IN CIA

onto latent lines in X and Y. PLS then maximises the covariance between the latent variables of the matrix (X) and the latent variables in the response matrix (Y), and the responses are then modelled as linear combinations of these latent X variables. These principal components (PC’s) are extracted in an iterative process to find the direction in X and Y space where the PC’s (e.g. u1 and t1, figure 7.5.1) covariance is maximised by minimising the partial least squares. As in PCA the PC’s are all orthogonal to each other. The PLS equation can be explained as linear combinations as follows:

\[
y = a_1t_1 + a_2t_2 + a_3t_3 + \ldots + a_nt_n
\]
\[
t_1 = b_{11}x_1 + b_{12}x_2 + \ldots + b_{1p}x_p
\]
\[
t_2 = b_{21}x_1 + b_{22}x_2 + \ldots + b_{2p}x_p
\]
\[
t_i = b_{i1}x_1 + b_{i2}x_2 + \ldots + b_{ip}x_p
\]

If the Y variables are correlated to each other multi-Y PLS can be used, which yields a more stable model than single-Y PLS, handles missing or deviating data better, allows higher degrees of freedom and (thus) provides more accurate models.

In Paper IV the relationship between peptide structural variables (X) and the biological response (Y), determined as binding to the A\(^q\) molecule, was studied. The X-matrix consisted of a well balanced set of peptides selected by D-optimal design represented by a combination of principal components of the included amino-acids in corresponding order (variables). The Y-matrix described the response at different concentrations of the peptides (multi-Y). Both X and Y were scaled to unit variance before regression.

7.6. Model validation

A n ACQUIRED QSAR MODEL (like other PLS or PCA models) is a mathematical model describing the studied interactions/properties and should be considered as such rather than an absolutely true description of the relationships concerned. The performance of models needs to be tested and they should be validated either internally or externally. While internal validation tests the model using objects that were included in its construction, an external
validation tests the model using a new data set. External validation is always by far the best way of verifying a model, but it cannot always be applied, due to time and/or cost constraints or the limited availability of compounds. The observations/objects chosen for an external test set should be diverse, according to criteria mentioned above, and challenge the validity of the model throughout all of the space it is supposed to cover.

The quality of any model can be described by goodness of fit ($R^2_Y$), goodness of prediction ($Q^2$) or permutation tests, and if it is not possible to use an external test set one must rely entirely on these diagnostic tools. $R^2_Y$ describes the proportion of the variation in $Y$ that is explained by the model, i.e. the fit of the model, and is calculated as:

$$R^2=1-(RSS/TSS)$$

RSS is the residual sum of squares (representing the variation that is not explained by the model) and TSS is the total sum of squares of $Y$ corrected for the mean. The predictive ability ($Q^2$) of the model can be calculated using cross-validation, the most common method used for internal validation, which also allows the number of components to be estimated. The data are first divided into a chosen number of classes and each class should be left out once throughout the modelling procedure. If the classes contain only one object, the leave one out (LOO) approach, can generally be used to determine the highest $Q^2$, but the robustness of the model may be poorly accounted for. Furthermore, with designed data the LOO approach may also give false indications of poor predictability since leaving one out may result in an entire “corner” in the chemical space being missed. The cross-validated $Q2$ are calculated according to:

$$Q^2=1-(PRESS/TSS)$$

PRESS is the predicted residual sum of squares when all objects have been left out of the modelling once.

The peptide QSAR model described in Paper IV was validated both internally, using cross-validation based on seven classes, and by an external test set consisting of six new objects (peptides). The six peptides of the external test set were selected to test the predictability through-out the entire model space.
THE MHC-GLYCOPEPTIDE-T CELL INTERACTION IN CIA

(Figure 7.6.1). Pairs of peptides were chosen that, according to the QSAR model, was predicted to elicit strong, moderate and very weak responses, i.e. have high, moderate and very weak binding affinity, respectively.

Figure 7.6.1
8. Results and discussion

The MHC-glycopeptide-T cell three-component system, associated with CIA in mice, was studied in detail by elucidating the binding properties of the collagen glycopeptide to A\(^q\) class II MHC molecules, as well as its T-cell receptor (TCR) interactions points. The separate techniques used, such as solid-phase peptide synthesis, amide bond and hydroxyl group isosters, statistical molecular design and biological tools have been presented earlier in their respective chapters, and the main results of Paper I-V are discussed in this chapter.

8.1. The minimal peptide core based on the CII glycopeptide, capable of binding to the A\(^q\) molecule as well as stimulating a full T-cell response

The immunodominant part of collagen from different species, which is capable of inducing a proliferative T cell response in H-2q mice has previously been located to residues 256-270 of CII. This finding was based on evaluation of type II collagen that had undergone chemical cleavage.
and also of a series of overlapping synthetic peptides. Other class II MHC molecules have been found to be able to present peptides as short as hexa- or heptamers, which suggests that the fifteenmer CII256-270 could be considerable truncated with retained Aq binding. A study with overlapping peptides and alanine scans, further indicated the T cell determinant region to reside within residues 260-267 for the naked, non-glycosylated peptides i.e. a quite limited region. The highly interesting posttranslationally modified CII256-270 peptide, with a galactosyl moiety on the central hydroxylysine 264, is recognised by specific, sugar dependent T cells. The influences of (glyco)peptide length for stimulation of these T cells has not been studied earlier.

In Paper I we studied the length aspects of the CII (glyco)peptide required for binding to the Aq molecule and for T cell stimulation. Two series of truncated non-glycosylated peptides, N- respectively C-terminally truncated based on the longer CII257-275 peptide, were synthesised on solid phase (Figure 8.1.1). They were prepared as N-terminal acetates and C-terminal amides, and tested for binding to the Aq molecule. This was done in two separate competition assays using whole cells expressing Aq, or affinity purified mouse-spleen derived Aq molecules. The set of N-terminally truncated peptides lost binding, in an all or nothing manner, after removing isoleucine 260, confirming the role of this residue as an anchor position. In the set of C-terminally truncated peptides there was a gradual loss in binding as the peptides were shortened, and the heptamer CII260-266 constituted the minimal epitope that is able to display at least weak binding to Aq. The ability of the truncated peptides to stimulate T cells was evaluated using T-cell hybridomas, established earlier from T-cells acquired after immunisation of mice with type II collagen. All the N-terminally truncated peptides that were capable of binding to Aq, i.e. N-terminal peptides starting at amino acid 260 or earlier, were recognised by the tested T cells hybridomas. The
C-terminally truncated peptides further revealed CII260-267 to be the minimal peptide epitope capable of binding to Aq with sufficient affinity, and also presenting appropriate T cell contact residues so that a full response (comparable to the longer counterpart) is produced from the T cell hybridomas. Once another amino acid is truncated, resulting in peptide CII260-266, the response was still clearly detectable but dropped at least two orders of magnitude. Synthesis and evaluation of the minimal T cell epitope CII260-267 in its glycosylated form, i.e. carrying a D-galactosylated hydroxylsine moiety at position 264 (Figure 8.1.2), revealed that the glycopeptide contained all structural elements required to stimulate carbohydrate-specific T cell hybridomas obtained in arthritic mice. Thus based on the two sets of truncated peptides from rat collagen, an Aq binder containing all critical T cell contacts points was found within the octamere peptide CII260-267. A shorter peptide lead constitutes an advantage in future drug-discovery projects, for instance by allowing more facile structural manipulation.

**Figure 8.1.2**

### 8.2. Side-chain T cell contact points of the collagen glycopeptide that are required for T-cell stimulation

**Paper II** addressed the issue of identifying T cell contact points of the minimal glycopeptide other than the earlier studied posttranslationally modified galactosyl moiety on the central hydroxylsine 264. Alanine substituted glycopeptides were therefore used to study the role of side chains in stimulation of sugar restricted T-cell hybridomas, selected from five groups previously found to have different carbohydrate specificity. A heavily alanine and glycine substituted peptide, lacking side-chains with hydrogen bond donor or acceptor interaction points to the TCR within the established minimal peptide core, was synthesised and tested for T cell recognition. Three out of five tested T-cell hybridomas, from different carbohydrate specificity groups, responded to the poly-substituted peptide just as
well as to the corresponding glycosylated collagen peptide. This means that hybridomas from these three groups do not need any side chain residue for stimulation, besides the galactosylated hydroxylysine 264. The T-cell responses towards a peptide, having a single alanine substitution at position 266, and to the corresponding mouse peptide, revealed that hybridomas from the two remaining groups are dependent on glutamic acid 266 as a critical T cell contact point. This position possesses the only amino acid that distinguish rat from mice collagen within the immunodominant region, where the glutamic acid in rat is replaced by aspartic acid in mice. The residue at position 266 is a highly interesting residue in autoimmune responses and is e.g. responsible for different results in earlier vaccination experiments using rat or mice collagen.

The results obtained with the poly-alanine substituted glycopeptide further implied that there could be T cell hybridomas that respond to highly modified glycopeptides as long as the anchor residues for the P1 and P4 pockets of A9, and the galactosylated hydroxylysine moiety, are retained in the correct positions. To test this hypothesis we designed a completely different glycopeptide based on the peptide backbone of the self peptide CLIP. The CLIP peptide was modified by introducing a galactosylated hydroxylysine at a central position, corresponding to position 264 of CII, as indicated in a comparative modelling study (Figure 8.2.1). The lack of T-cell recognition of this glycopeptide was thought to depend on disruption of TCR binding due to the large arginine side chain (corresponding to position 261 in CII) pointing up
towards the TCR adjacent to the galactose residue. Even though the peptide side chains are not involved in T-cell signalling for the majority of the sugar restricted hybridomas, an APL must not contain side chains preventing binding to A8 or interrupting critical TCR interactions.

8.3. Interactions of the backbone amide bonds

BESIDES INTERACTIONS INVOLVING the peptide side chains, backbone amide bonds could also contribute both in MHC binding and interactions with the TCR. In fact, the conserved class II MHC molecule seems to form strong interactions with the main chain atoms of the bound peptide. In a study of peptides that bind to the MHCII molecules DR1 and DR4 (both linked to RA in man) the amide bonds were found to be critical for most studied positions, indicating the great need for hydrogen bonding between the backbone amides and DR1 or DR4. In Paper II, a methylene ether isostere which replaced one of the amide bonds was used to highlight the role of the peptide backbone in interactions with A8, and at the same time stabilise the glycopeptide towards enzymatic degradation. The dipeptide isostere was incorporated instead of isoleucine 260 and alanine 261, in the minimal glycopeptide epitope CII260-267,Gal264 (Figure 8.3.1). The choice of isosteric position was based on its location between the two essential anchor residues, a wish to offer protection against N-terminal peptidases and synthetic feasibility. Two versions of the isosteric peptide were tested in the competition assays using affinity purified A8 molecules, one N-terminally acetylated and the other with a free amine at the N-terminus. The lack of binding of the peptide having a free –amino group was most likely due to that the positive charge at the N-terminus prevented binding in the binding cleaf. The N-terminally acetylated isosteric glycopeptide bound but required a twenty times higher concentration to reach the same level of binding as the reference native rat peptide CII259-273. This most likely reflects that the Ile260-Ala261 amide bond is involved in hydrogen bonding to the A8 molecule. In T-cell hybridoma assays, this peptide was recognised by two out of four groups of hybridomas having different specificity for the galactose moiety, but 100-1000 times higher concentrations were required to reach a maximum T
cell response. This probably reflects the weaker binding to A^q, and could also indicate an interaction between the TCR and the Ile260-Ala261 amide bond.

![Dipeptide isostere Building block](image1)

![Isostere peptide](image2)

Figure 8.3.1

An in detail comparison of the specificity of the different T-cell hybridomas for the side chain of glutamic acid 266, the Ile260-Ala261 amide bond and the fine-specificity for the galactose moiety, allowed us to propose a refined model for TCR interactions with the glycopeptide-MHC complex. We hypothesised that the TCR’s of the different hybridomas recognise the galactose residue on hydroxylysine from different sides (Figure 8.3.2). It is also possible that the galactose moiety displays different epitopes to the hybridomas in order to allow formation of critical interactions. This should be facilitated by the flexibility of the glycosidic bond to hydroxylysine.

![Figure 8.3.2](image3)
8.4. The importance of hydroxyl group 4 of the galactose moiety for T-cell recognition

While the T cell hybridomas from the four groups of sugar dependent T cells hybridomas showed different specificity for the 2-, 3-, or 6-hydroxyl groups of the galactose moiety, all hybridomas were sensitive to deoxygenation at C-4, which resulted in reduction or complete loss of the response. In Paper III we continued to probe the T cell specificity for the galactose moiety by focusing on the critical 4-hydroxyl group. The specific hydrogen bond donor-acceptor characteristics of HO-4 were studied using glycopeptide analogues modified at C-4 by O-methylation or exchange of HO-4 for a fluorine atom. The O-methylated analogue has lost its hydrogen bond donor property and is more sterically hindered as compared to the hydroxyl group, while the fluorine analogue offers a mimic that is similar in size and polarization and possibly acts as a potential weak hydrogen bond acceptor. In Paper III we had chosen not to consider fluorine as a hydrogen bond acceptor, due to the ongoing critical debate on fluorine as acceptor. However, polar components of ligand-receptor interactions most likely reflect an at least weak hydrogen bonding interaction. In fact, interpretation of the data in Paper III is facilitated if fluorine is viewed as a weak hydrogen bond acceptor. None of the modified analogues (HO-4 deoxygenated, O-methylated or replaced by fluorine) induced a full T cell response, i.e. identical to the galactose moiety, in T cell assays using representative hybridomas from each subgroup. A complete lack of response to the analogues from hybridomas in two of the groups, implies that HO-4 both donates and accepts hydrogen bonds from the T-cell receptors in these hybridomas (Figure 8.3.2). The third group of T cell hybridomas is possibly involved in polar interactions (most likely weak hydrogen bonding) with the T-cell receptor. The hybridoma in the fourth group shows weaker but significant responses to the modified analogues, indicating no direct hydrogen bonding for this group of hybridomas. These results verify the role of HO-4 as an important contact point for the TCR, where very minor changes drastically influence the TCR response. For other antigens, such minor changes have previously been shown to be able to drastically modify the response of a T-cell, including sending it into an anergic state.
8.5. A QSAR model for peptide binding to the A\textsuperscript{d} molecule

Despite the many substitutions in the polyalanine substituted CII glycopeptide (Paper III) it retained almost all binding to the A\textsuperscript{d} MHCII molecule; in a binding assay using purified A\textsuperscript{d} as little as a twofold higher concentration was needed to reach the same binding as for the native rat glycopeptide. This correlates well with the finding that polyalanine substituted peptides, with anchor residues intact, retained binding to the closely related DR1 class II MHC molecule.\textsuperscript{98} The majority of the peptide side chains are thus not required for high affinity binding to DR1, suggesting that even bigger changes in amino acid sequence could be tolerated provided that the anchor residues are retained.\textsuperscript{99}

The peptide requirements, “position for position”, for binding to the A\textsuperscript{d} molecule were studied in detail in Paper IV, using statistical molecular design (chapter 7) and a QSAR model was established. Based on the minimal CII binding epitope, CII260-267, five amino acid positions, i.e. 261, 262, 265, 266 and 267 were selected for investigation. The anchor residues, Ile260 and Phe263, and the galactose bearing hydroxylysine 264 were left intact. If peptides consisting of all possible combinations of the coded amino acids at the five selected positions should have been synthesised and tested, this would have resulted in an enormous study based on 3,200,000 peptides. Instead, by using the different molecular characteristics of the amino acids, as described by 28 molecular descriptors, 5-6 representative amino acids could be chosen for each position without substantial loss of information. The multivariate characterisation was performed using PCA and the amino acids were chosen to maximize the selected chemical space. All together 4500 virtually combined peptides were selected, based on all reduced possible combinations of amino acids, and this library was further compressed using D-optimal design to a (substantially reduced) peptide library consisting of only 22+3 peptides. The library with diverse and representative molecular characteristics was synthesized on solid phase and evaluated for binding to A\textsuperscript{d} molecules in a FACS-analysis competition assay. The quantitative PCA score values were further used to mathematically describe each amino acid, which in combination, corresponding to respective amino acid sequence, described the peptides of the reduced library. These characteristics were then correlated to the biological response, given as %inhibition. A linear regression correlating X (peptide structure) with Y (binding to A\textsuperscript{d}-molecule) was
performed using the PLS method (NIPALS algorithm). This allowed a high quality QSAR model to be established, explaining 66% of the biological response. Interpretation of the model coefficients and regression weights allowed backtracking of the desirable characteristics to suggested amino acids (in the PCA).

<table>
<thead>
<tr>
<th>Peptide pos.</th>
<th>Pos1 P2</th>
<th>Pos2 P3</th>
<th>Pos3 P6</th>
<th>Pos4 P7</th>
<th>Pos5 P8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Character</td>
<td>small rigid</td>
<td>small rigid</td>
<td>hydrophobic flexible</td>
<td>large flexible (H-donor, acceptors) (aromatic)</td>
<td>large flexible (H-donor, acceptors) (aromatic)</td>
</tr>
<tr>
<td>Character</td>
<td>large flexible</td>
<td>large flexible</td>
<td>polar H-donor, acceptors</td>
<td>small rigid (aliphatic)</td>
<td>small rigid (polar)</td>
</tr>
</tbody>
</table>

**Dominating descriptors Spartan/WHIM**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Volume 1K, 2K, PHI</th>
<th>Volume 1K, 2K, PHI</th>
<th>Volume 1K, 2K, PHI</th>
<th>Volume 1K, 2K, PHI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation:</td>
<td>negative</td>
<td>negative</td>
<td>positive</td>
<td>positive</td>
</tr>
</tbody>
</table>

**Suggested binding site**

| Aα molecule | smaller polar | smaller polar | hydrophobic | big pocket positive | big pocket positive |

Amino acids are represented by their conventional one letter code. *Varied position in peptide. †Suggested corresponding position in the Aα molecule. Descriptor abbreviations: K1, path1 Keir shape index; K2, path2 Keir shape index; PHI, Keir flexibility index; LOGPC, LogP; AM1SMM2, Aqua solvatation grade.

Figure 8.5.1

Amino acid positions 266 and 267 in the CII scaffold had the largest influence on peptide binding to the Aα molecule according to the regression model. The variables describing size and flexibility are the main contributors for both positions, and in order to promote strong binding these amino acids should preferably be large and flexible (Figure 8.5.1). While slightly polar residues are preferred in position 266, hydrophobic amino acids are favoured in position 267. Hydrophobic and flexible amino acids of moderate size are favoured in position 265. Positions 261 and 262 have a small influence on binding in the acquired
model, because all changes gave similar biological responses. While amino acids of moderate size are tolerated, there is however, a slight preference for smaller and more rigid amino acids at these two positions. The influential descriptors and suggested amino acids at the five positions are summarised in figure 8.5.1.

The QSAR model was verified using an external validation set of six peptides (chapter 7.6). These were designed to represent two good, two average and two poor binders as predicted by the QSAR model. Following synthesis, the peptides were tested in the binding assay to rA\(q\) molecules. After comparing the inhibition curves, the peptides were ranked relatively each other, and the results confirmed the predictability of the QSAR model.

Analysis of the preferred molecular properties of the peptide also gave indications of the properties of the A\(q\) binding site, a highly valuable information since the A\(q\) molecule has not yet been crystallised. Preferences for small and rigid amino acids in positions 261 and 262 of CII, corresponding to binding pocket P2-P3 in A\(q\), suggest binding pockets of limited size in this area between the P1 and P4 anchor pockets. The P6 pocket, occupied by residue 265 in CII, is implied to be hydrophobic and of medium size. Both the P7 and P8 pockets appear to be large binding pockets. The P8 pocket is likely strictly hydrophobic, while P7 harbours some polar surface area. While waiting for the A\(q\) MHC II molecule to be crystallised, we can conclude that the characteristics for the binding pockets, implied by the QSAR model, agree very well with a structural model of A\(q\) obtained independently by comparative modelling.\(^{36}\)

### 8.6. The stereochemical requirements for glycopeptide binding to the A\(q\) molecule and for T-cell recognition

The acquired QSAR model for peptide binding to the A\(q\) molecule was built exclusively on the naturally occurring coded L-amino acids. This gave us a deeper knowledge of which characteristics and amino acids that are tolerated and preferred in the binding site of the A\(q\)-molecule. The model does, however, not take into consideration the effects of changing the C\(a\) configuration of the amino acid building blocks. Moreover, our studies of the critical T cell contact points were also restricted to L-amino acids, except for a
study with galactose on D-hydroxylysine, which resulted in loss of binding to the A\(^8\) molecule.\(^{36}\) In Paper V stereochemical requirements for glycopeptide binding to the A\(^8\) molecule, as well as for T-cell recognition, were studied.

Three positions within the minimal binding epitope CII260-267 (positions 261, 266 and 267, respectively) were investigated for their stereochemical requirements by substitution of L- for D-amino acids (Figure 8.6.1B). The positions lacking stereocentra (Gly262 and 264), the critical anchor residues (Ile260 and Phe263) and the T cell contact (Gal-Hyl264) were not altered in this study. The amino acids at the three positions were separately replaced by a D-alanine, i.e. the smallest amino acid harbouring a stereocentra, to introduce the smallest possible interference e.g. in binding to A\(^8\). An additional peptide where L-glutamic acid 266 was replaced by a D-glutamic acid was also designed, to be able to investigate the stereochemical requirements of the T-cell hybridomas dependent on glutamic acid 266. The glycopeptides were synthesised on solid phase and tested for A\(^8\) binding using recombinant A\(^8\), and also for T-cell stimulation. All the D-alanine substituted glycopeptides retained binding to the A\(^8\)-molecule, indicating no limitations to small stereochemical modifications in the studied positions. On the other hand, the glycopeptide with the much larger D-glutamic acid in position 266 displayed a drastic drop in A\(^8\) binding. Nine T cell hybridomas, representing the five subgroups divided according to the fine-specificity for the galactose moiety, were used to study the T cell stimulation capacity of the D-amino acid substituted peptides. The results revealed that use of D-alanine in position 261 prevents a T-cell response from all hybridomas, possibly through steric hindrance of the TCR (Figure 8.6.1A). In position 267, the same effect upon introducing D-alanine was seen in two out of five groups of hybridomas, while the other hybridomas elicited a weak or medium response. The glycopeptide substituted with D-glutamic acid in position 266 lacked A\(^8\) binding and as a consequence could not stimulate a T-cell response. When this position was substituted with D-alanine, the resulting glycopeptide was found to stimulate three out of five groups of hybridomas. The responding hybridomas are the same ones that were earlier found to be independent of glutamic acid 266, while the glutamic acid dependent hybridomas did not respond. The combined results from the investigation agreed well with the structural features of the complex between the wild type glycopeptide and the A\(^8\) molecule, as revealed by comparative modelling of the complex.\(^{36}\) (Figure 8.6).
### A

<table>
<thead>
<tr>
<th>Position</th>
<th>Cross-section</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>261</td>
<td><img src="image1" alt="Image" /></td>
<td>L-Ala-&gt;D-Ala: TCR steric hindrance in all groups of hybridomas</td>
</tr>
<tr>
<td>266</td>
<td><img src="image2" alt="Image" /></td>
<td>L-Glu-&gt;D-Glu: Intercepts Aβ binding. L-Glu-&gt;D-Ala: T cell recognition by Glu-independent hybridomas.</td>
</tr>
<tr>
<td>267</td>
<td><img src="image3" alt="Image" /></td>
<td>D-Gln-&gt;D-Ala: TCR steric hindrance for some hybridomas</td>
</tr>
</tbody>
</table>

### B

![Image](image4)

**Figure 8.6.1**
9. Concluding remarks and future prospects

The main conclusions that can be drawn from the studies underlying this thesis on the interactions involved in the MHC-glycopeptide-T cell contact can be summarised as follows:

The MHC-peptide interaction

- The heptamer CII260-266 constitutes the minimal epitope that can bind, at least weakly, to H-2A^d.
- The amide bond between Ile260 and Ala261 makes significant contributions to binding to the A^d molecule.
- Amino acids 266 and 267 in the CII peptide make substantial contributions to the peptide’s binding to the A^d molecule, besides the earlier established anchor residues 260 and 263. Position 265 has intermediate influence on binding and positions 261 and 262 have minor influence.
- For a peptide that binds with high affinity to A^d the amino acid at positions CII261 and 262 should be small and rigid, the residue at position CII265 should be hydrophobic and flexible, the residue at position CII266 should be large, flexible and slightly polar, and the residue at position CII267 should be large, flexible and hydrophobic.
- Small D-amino acids at positions 261, 266 and 267 of CII are compatible with binding to A^d, indicating that there a degree of stereochemical freedom at these positions. However, the presence of the large amino acid D-Glu at position 266 effectively disrupts the binding.

The glycopeptide T cell contact points

- CII260-267 is concluded to be the minimal peptide epitope capable of presenting appropriate T cell contact residues to elicit a full
response (comparable to that of the longer counterpart CII256-270) from the T cell hybridomas. The corresponding glycopeptide, CII260-267Gal264, was also found to contain all of the structural elements required to stimulate carbohydrate-specific hybridomas.

- T cell hybridomas from two out of five groups that recognise galactosylated hydroxylysine 264 also need glutamic acid 266 for stimulation, while hybridomas from the other three groups do not require contact with any other specific amino acid side chain.
- Some of the T cell hybridomas from the groups that do not need glutamic acid 266 for stimulation require the amide bond between Ile260 and Ala261.
- The observation that either the Glu266 or the Ile260-Ala261 amide bond may be required for T cell recognition, in addition to Gal-Hyl264, of CII glycopeptides, led to the hypothesis that the T cell receptor recognises the glycopeptide from different sides of the galactose moiety.
- Hydroxyl group 4 of the galactose moiety is an important contact point for the TCR and can interact both as a hydrogen bond donor and acceptor.
- The presence of D-alanine at position 261 prevents a T-cell response for all groups of hybridomas. The Glu266-independent hybridomas are not affected by the D-Ala266 substitution. Some T cell responses are prevented by D-amino acids at position 267, possibly due to steric hindrance.

The studies this thesis is based upon (Papers I-V) elucidated a number of features of A\(^4\)-glycopeptide-TCR interactions that are relevant to autoimmune responses towards collagen and, more generally, to T cell recognition of glycopeptides. There seem to be few key TCR-glycopeptide points of interaction. Besides hydrogen bonding to the galactose hydroxyl groups, the only peptide side-chain confirmed to be important in this respect was that of the glutamic acid at position 266. Evidence that some TCRs interact with the amide bond between amino acid 260-261 in collagen was obtained. However, the analysis of amide contributions to the extensive hydrogen bonding involved in the TCR-peptide interactions is not yet complete and further interesting details may emerge. Several new potential anchor positions, and binding characteristics,
were revealed in the QSAR study and further investigation of the original anchor positions Ile260 and Phe263 may be warranted. It may also be possible to design, synthesise and ultimately test in vaccination experiments a “full” mimetic peptide (non-peptide scaffold) in which all the immunodominant interaction points are retained but specific physico-chemical characteristics are modified to give greater enzymatic stability than the immunodominant peptide. Furthermore, undesirable loss of binding due to i.e. insertion of peptide mimetics could be compensated for by including additional and/or stronger anchor residues based on the established A\textsuperscript{4} binding motif. The reduction of the immunodominant region, to a fragment only seven amino acids long, has reduced the area that needs to be modulated and should thus facilitate such studies.

Even small changes within an immunodominant epitope have been shown to be capable of inducing altered T cell responses, resulting in either anergy or shift in the dominating Th responses. The altered glycopeptides are therefore ideal for use in analyses of immunosuppressing properties and vaccination experiments. The APL’s, including the galactose-isostere peptides, could be tested for their ability to stimulate T cell clones, and the resulting cytokine profile could be assayed to determine their probable effector functions as T\textsubscript{H1}, T\textsubscript{H2} or Tregs. Peptides that induce a partial activating signal response in the T cell hybridoma assays should be particularly interesting to see if they can induce altered T\textsubscript{H2} or Tregs responses. New APL’s could be synthesised in which several combinations of the TCR contact points have been removed to search for a candidate peptide capable of switching the T cell effector functions and associated immunological responses. It would further be interesting to correlate TCR affinity on/off-rates for the peptide-A\textsuperscript{4} complexes (using protective respectively non-protective peptides towards CIA) to the T cells' effector functions, and a T cell affinity window could possibly be established in the same way as for thymus selection for the different outcomes. The QSAR-established binding motif for A\textsuperscript{4}, allows APL’s to be designed with weaker or moderate binding affinity to A\textsuperscript{4} in attempts to alter the T cell responses. Finally, the QSAR-derived A\textsuperscript{4} binding motif also provides opportunities for testing a less subtle method to combat RA by blockading A\textsuperscript{4}, although this would not be the first method of choice. The results presented in this thesis and the underlying studies will hopefully help to accelerate the development of effective new drugs against RA.
10. Acknowledgements

Det är mycket efterlängtat och med en viss overklighetskänsla som jag nu skriver på slutet av avhandlingen....tillika avsnittet som verkar ha högst “impact factor”. Jag har många att TACKA som varit till övärderlig hjälp professionellt och privat:

Min huvudhandledare Jan Kihlberg som med sitt otroligt organiserade och lugna sätt styrt projekten tillbaka på säker mark “Nej, inga liposomer!” och trots boende i Göteborg ständigt tar sig tid för diskussioner, ständigt inflygande för nya inplanerade möten eller via telefonsamtal. Tack också för att du under din strikta image har en stor empatisk sida.

Rikard Holmdahl för sin inspirerande energiska sida som brinner för forskning och för att jag fått komma och jobba på ditt lab.

Mikael Elofsson och Anna Linusson som tidvis fått hoppa in som reservhandledare. Tack också Elof för att du med glimten in the eye tagit på dig jobbet som språk-polis. Tack Anna för att du är så positiv att jobba med och (oftast) tålmodig. (Så har du en fin hund också!)

Dan Johnels som gett mig extra undervisning när jag önskat och visat tilltro till mig.

Carina Sandberg som nitiskt lyckats hålla reda på alla papper, datum och mig.

Alla gamla och nya på organisk kemi som gjort avdelningen till ett trevligt ställe att jobba på! (Så trevligt att jag under avhandlingsskrivandet hemma mest kommit in till jobbet för att luncha eller fika):

Tomas Gustafsson som helt enkelt är superbra både på och utanför jobbet. Tack för alla middagar och filmkvällar (trots att jag brukar somna ifrån slutet) och allt annat du gjort för mig.

Nya (Ida, Sara ”puss-puss”, Sussi) och gamla (Björn, Zhongqing, Baquer, Stina) och ytterst tillfälliga (Suss, Knatti, Mattias) rumskompisar som fått stått ut med en hel del pratande och datautbrott, och nästan alltid varit glada och hjälpsamma. Speciellt min gamla vapendragare Björn Holm (nej vi är fortfarande inte släkt eller gifta) som satt så oavbrutet så t.o.m. en pratkväll som jag lärde mig, förutom diverse (oviktiga?) saker, att uppskatta tystnad :-).och Ida som både är hjälpsam och låtsas intresserad när jag vill prata T celler.

Gamla sockerpojkarna: Johan, Petter, Moge, Wallner Alltid glatt visslande: Fripp Hassarna (H1 och H2) för ni är bara så roliga helt enkelt! Vadduärnsnyggidag-Blompa Party Ing-Marie och alltid lika omtänksamma Anna Kauppi självklart, är några till som ej får glömmas som forgyllt/förgyller avdelningen. Och så alla ni andra förstås!
THE MHC-GLYCOPEPTIDE-T CELL INTERACTION IN CIA


För mina vistelser på medicinsk inflammationsforskning i Lund vill jag speciellt tacka Mikael Vestberg, Alexandra o Johan Bäcklund, Balik Dzhambazov, Robert Bockermann, Sharam, och Hisakata Yamada för den hjälp jag fått på labbet och/eller då ni bjudit med mig på annat.

När man nu bosatt sig så många mil hemifrån så har man ju varit tvungen att skaffa en substitut-familj : Min “klon” Ingrid Söderberg alias Inga-from–Sweden för allt kul och dumt vi gjort (kanske borde vi hinna lite mer nu). Bastuklubbens ordförande Anna Segerman från Mockfjärd för all “träning” på IKSU och djupa/”djupa” diskussioner. Fixartjejen nr 1 Anneli Ågren för alla bara-titta-in-middagar/fika, och respektive familjer (Walle, Bosse, Ola och barn) TACK för att ni är så bra att vädra åsikter med utanför jobbet.


Slutligen vill jag tacka min familj: Pappa Bo, Mamma Margaretha, syster Anki och bror Christoffer för att de alltid “hejar på” mig. Min fyrbenta sambo Freja tackas för ”långa skogspromenader och mysiga hemmakvällar.”

Till alla TACK! Ni som inte orkar läsa avhandlingen också skall veta att jag gömt nästa veckas lottorad i texten...

Lotta 9 oktober 2006

Finally I like to thank for Financial support from “Glycoconjugates in Biological System” sponsored by the Swedish foundation for strategic research. Tack också till GLIBS-Barbro Lowisin o Ulf Lindahl för att ni på ett personlig plan stöttat färdigställandet av avhandlingen.
THE MHC-GLYCOPEPTIDE-T CELL INTERACTION IN CIA

11. References

4. Holmdahl, R., Section of Medicinal Inflammation Research, Lund University, Lund, Sweden.
THE MHC-GLYCOPEPTIDE-T CELL INTERACTION IN CIA


THE MHC-GLYCOPEPTIDE-T CELL INTERACTION IN CIA


THE MHC-GLYCOPEPTIDE-T CELL INTERACTION IN CIA


THE MHC-GLYCOPEPTIDE-T CELL INTERACTION IN CIA

98. Jardetzky, T.; Gorga, J.; Busch, R.; Rothbard, J.; Strominger, J.; Wiley, D., peptide binding to HLA-DR1: a peptide with most residues substituted to alanine retains MHC binding. The EMBO journal 1990, 9, (6), 1797-1803.