The Role of Endoplasmic Reticulum Aminopeptidase 1 (ERAP1) in the Pathogenesis of Ankylosing Spondylitis

by

Nigil Haroon

A thesis submitted in conformity with the requirements for the degree of Doctorate of Philosophy in Medical Sciences

Institute of Medical Sciences
University of Toronto

© Copyright by Nigil Haroon 2012
The Role of Endoplasmic Reticulum Aminopeptidase 1 (ERAP1) in the Pathogenesis of Ankylosing Spondylitis

Nigil Haroon

Doctorate of Philosophy in Medical Sciences
Institute of Medical Sciences
University of Toronto
2012

Abstract

Ankylosing spondylitis (AS) is associated with HLA-B*2704 and B*2705 but not with HLA-B*2706 and B*2709. Genome wide studies recently identified ERAP1 as an important genetic association in AS and could be the missing link in the pathogenesis of AS.

I studied the implications of the two known actions of ERAP1 on AS pathogenesis. For assessing the peptide trimming function, surface HLA-B27 and MHC-I free heavy chain (FHC) expression on peripheral blood mononuclear cells of AS patients were studied. Subsequently, in an in vitro system of C1R cells expressing different AS-associated and AS-neutral HLA-B27 subtypes, I studied the effect of ERAP1 suppression on HLA-B27 and FHC expression. To assess the cytokine receptor shedding function, I studied serum cytokine receptor level variation with ERAP1 polymorphisms and its relationship to disease activity in AS patients. Finally, I studied the effect of variants of ERAP1 and other members of the antigen presentation machinery on radiographic severity in AS patients.
AS patients with the major allele of the *ERAP1 rs27044* polymorphism had higher FHC expression on monocytes. In C1R cells *ERAP1* suppression led to an increase in intracellular FHC (IC-FHC) and B27-peptide complexes identified by a special MARB4 antibody, but only in C1R cells expressing the AS-associated subtypes HLA-B*2704 and B*2705. ERAP1 variants had no effect on serum cytokine receptor levels. Baseline radiographic severity was associated with ERAP1 polymorphism in univariate analysis only. LMP2 variants were associated with baseline radiographic severity in multivariate analysis.

ERAP1 affects peptide presentation and FHC formation by HLA-B27 and could be the missing link in the pathogenesis of AS. ERAP1 through its differential HLA-B27 subtype interaction could explain why certain subtypes of HLA-B27 are associated with AS while others are not. Larger studies are required to look closely at the effect of ERAP1 on radiographic severity and progression in AS.
Acknowledgments

An exceptional journey it has been,
Meandering through the perplexing landscape of ankylosing spondylitis,
An occasional stroll, an occasional skip,
  Largely a sprint
The sun peeking through the clouds
  Blinding at times
  Enlightening otherwise

Rob took care, always with a prayer
  To ease my affair
A home away from home did he bestow
  His love did overflow

Trials made me sigh
  But help was nigh
Florence was my ally
  And helped me fly

David and Tania are sharp
And helped build a castle from scrap

What a journey
  A pleasure for me
For Nisha was ever beside me.
To

Dad and Mom

Who selflessly gave up everything for me
# Table of Contents

Acknowledgments ........................................................................................................ iv

Table of Contents ........................................................................................................ vi

List of Figures .............................................................................................................. xi

List of Tables ............................................................................................................... xiii

List of Abbreviations .................................................................................................. xiv

Chapter 1 Introduction ................................................................................................. 1

1 Ankylosing Spondylitis .............................................................................................. 1

1.1 History .................................................................................................................. 1

1.2 Epidemiology ......................................................................................................... 2

1.3 Clinical Features ................................................................................................... 3

1.3.1 Inflammatory Back Pain (IBP) ........................................................................ 3

1.3.2 Peripheral Arthritis .......................................................................................... 5

1.3.3 Extra-articular Features ................................................................................ 5

1.4 Radiographic Features ......................................................................................... 6

1.5 Classification Criteria ........................................................................................... 8

1.5.1 Modified New York Criteria ........................................................................... 8

1.5.2 ASAS Criteria for Axial Spondyloarthritis ...................................................... 9

1.6 Treatment of Ankylosing Spondylitis .................................................................. 12

1.6.1 Non-pharmacological treatment ..................................................................... 13

1.6.2 Non-steroidal anti-inflammatory drugs .......................................................... 13

1.6.3 Disease modifying anti-rheumatic drugs (DMARDs) .................................... 14

1.6.4 Biological Therapy ......................................................................................... 14

1.7 Pathogenesis ......................................................................................................... 16

1.7.1 Infections and Pathogenesis of AS ............................................................... 16
1.7.2 Cytokines in AS .................................................................................................................. 18
1.7.3 Genetics of ankylosing spondylitis ......................................................................................... 21
1.7.4 Inflammation and New Bone Formation in AS ......................................................................... 29
1.7.5 DKK-1 in the Pathogenesis of Ankylosing Spondylitis ............................................................... 31

1.8 Endoplasmic Reticulum Aminopeptidase 1 ............................................................................. 35
1.8.1 Structure of ERAP1 .............................................................................................................. 37
1.8.2 Genetic association of ERAP1 with AS .................................................................................... 47
1.8.3 Normal functions of ERAP1 .................................................................................................. 49
1.8.4 ERAP1 polymorphisms and AS phenotype ............................................................................. 59
1.8.5 ERAP1 and HLA-B27 pathogenesis ....................................................................................... 60
1.8.6 ERAP1 in other diseases ...................................................................................................... 60

1.9 Aim ........................................................................................................................................ 63
1.10 Hypothesis ............................................................................................................................. 64

Chapter 2 .................................................................................................................................. 65

2 Functional interaction of ERAP1 and HLA-B27 is subtype specific and could be pathogenic in ankylosing spondylitis ......................................................................................... 65

2.1 Introduction ............................................................................................................................. 65

2.2 Methods ................................................................................................................................ 68
2.2.1 Patients ................................................................................................................................ 68
2.2.2 Genotyping .......................................................................................................................... 68
2.2.3 Cell Lines ............................................................................................................................ 70
2.2.4 Optimization of Nucleofection ............................................................................................ 70
2.2.5 ERAP1 Suppression by RNA Interference ............................................................................ 73
2.2.6 Western Blot Assay ............................................................................................................. 73
2.2.7 Fluorescence Activated Cell Sorting (FACS) ...................................................................... 75
2.2.8 Statistics ............................................................................................................................ 78
2.3 Results

2.3.1 Monocytes have the highest level of surface expression of HLA B27 and free heavy chains.

2.3.2 ERAP1 rs27044 polymorphisms affect the expression of free heavy chains on the surface of monocytes.

2.3.3 Surface expression of HLA B27 and FHC was not related to disease activity.

2.3.5 ERAP1 suppression of HLA-B*2705 C1R cells causes changes in intracellular free heavy chains and surface MHC-peptide complexes.

2.3.6 C1R cells expressing different HLA-B*27 subtypes have variable level of total MHC-I expression.

2.3.7 C1R cells expressing AS-susceptible HLA-B*27 subtypes have high MARB4 staining.

2.3.8 ERAP1 suppression leads to an increase in intracellular free heavy chains only in C1R cells expressing AS associated HLA-B*27 subtypes.

2.3.9 ERAP1 suppression leads to increase in MARB4 staining only in C1R cells expressing AS associated HLA-B*27 subtypes.

2.4 Discussion

2.5 Conclusions:

Chapter 3

3 Serum Cytokine Receptors in Ankylosing Spondylitis: Relationship to inflammatory markers and Endoplasmic Reticulum Aminopeptidase Polymorphisms

3.1 Background

3.2 PATIENTS AND METHODS

3.2.1 Patients

3.2.2 Genotyping

3.2.3 Serum cytokine and soluble cytokine receptor levels

3.2.4 Statistical analysis

3.3 RESULTS
3.3.1 Patients
3.3.2 Genotyping
3.3.3 Soluble cytokine receptor levels
3.3.4 No significant effect of ERAP1 polymorphism on serum cytokine receptor level
3.3.5 Serum cytokine levels
3.4 Discussion
3.5 Conclusion

Chapter 4

4 Radiographic severity in ankylosing spondylitis: The effect of genetic polymorphisms in ERAP1 and other elements of the antigen presenting machinery

4.1 Background
4.2 Methods
  4.2.1 Patients
  4.2.2 Radiographic Scoring
  4.2.3 Genotyping
  4.2.4 Statistical Analysis
4.3 Results
  4.3.1 Patients
  4.3.2 Baseline radiographic severity can be influenced by polymorphisms in LMP2
  4.3.3 Baseline mSASSS is the strongest predictor of progression
4.4 Discussion
4.5 Conclusions

Chapter 5

5 Conclusions and Future Directions

5.1 Conclusions
5.2 Future Directions

5.2.1 Studies in Ankylosing Spondylitis Patients

5.2.2 In vitro studies

References
List of Figures

Figure 1.1 The Canonical Wnt-β Catenin Pathway ........................................33
Figure 1.2 The Chromosome 5q15 locus with ERAP1 and other surrounding genes ....38
Figure 1.3 Nucleotide Change in the ERAP1 Polymorphisms Studied ......................40
Figure 1.4 Residue Change in the ERAP1 Polymorphisms Studied ........................41
Figure 1.5 ERAP1 Crystal Structure: Open Configuration ..................................43
Figure 1.6 ERAP1 Crystal Structure: Closed Configuration .................................45
Figure 1.7 Schematic Diagram to show the role of endoplasmic reticulum aminopeptidases in antigen processing and presentation .......................................50
Figure 1.8 ERAP1 Peptide specificity .................................................................53
Figure 2.1 Western Blot and analysis showing ERAP1 suppression with ERAP1 siRNA compared to the respective negative control siRNA ...............................74
Figure 2.2 Flow Assisted Cell Sorting of Peripheral Blood Mononuclear Cells ..........76
Figure 2.3 Total HLA-B*27 and Free Heavy Chain Expression was highest in the monocyte population among the peripheral blood mononuclear cells ...............84
Figure 2.4 Flow Cytometry Analysis of MHC-I Free Heavy Chain Expression on Peripheral Blood Mononuclear Cells from Patients with Ankylosing Spondylitis .........86
Figure 2.5 Comparing the effects of ERAP1 rs27044 SNP on free heavy chain and Intact HLA-B27 expression on different cell populations ..................................89
Figure 2.6 Comparing the effects of ERAP1 rs30187 SNP on free heavy chain and Intact HLA-B27 expression on different cell populations .................................91
Figure 2.7 Comparison of inflammatory markers across the different genotypes ..........93

Figure 2.8 Effect of siRNA induced ERAP1 suppression on HLA-B*27 expression in C1R cells expressing HLA-B*2705 ........................................................................................................95

Figure 2.9 Western Blot and FACS analysis of C1R cells at baseline. ......................98

Figure 2.10 The effect of ERAP1 Suppression on Intracellular Free Heavy Chain Expression in C1R cells expressing different HLA-B*27 subtypes.................................101

Figure 2.11 Percent of cells with high expression of intracellular free heavy chains....103

Figure 2.12 The effect of ERAP1 suppression on HLA-B*27 complexes recognized by MARB4 antibody on the surface of C1R cells expressing different HLA-B*27 subtypes ..............................................................105

Figure 3.1 Correlation of TNFRI with CRP and ESR.................................................122

Figure 3.2 No correlation of CRP, ESR and BASDAI with serum IL-6Rα and IL-1RII..125

Figure 3.3 Serum Cytokine Receptors in patients with different ERAP1 polymorphisms ......................................................................................................................127

Figure 4.1 Modified Stoke’s Ankylosing Spondylitis Spine Score.............................136

Figure 4.2 Cumulative Probability Plot of patients with progression of radiographic scores .........................................................................................................................147

Figure 4.3 ROC Curve for regression model with baseline mSASSS predicting radiographic progression ........................................................................................................150

Figure 4.4 Hypothetical mechanism for LMP2 and ERAP1 variants affecting radiographic progression in AS .................................................................................................155
List of Tables

Table 1.1 Modified New York criteria for ankylosing spondylitis............................................. 10

Table 1.2 ASAS classification criteria for axial Spondyloarthritis ............................................ 11

Table 1.3 Hypothesis of the role of HLA-B27 in the pathogenesis of Ankylosing Spondylitis .......................................................................................................................... 23

Table 1.4 Single Nucleotide Polymorphisms of ERAP1 with Nucleotide and Residue Change.......................................................................................................................... 41

Table 2.1 The Amaxa cell line optimization nucleofector kit .......................................................... 72

Table 2.2 Clinical and laboratory features of patients with different ERAP1 rs27044 genotypes .......................................................................................................................... 81

Table 2.3 Clinical and laboratory features of patients with different ERAP1 rs30187 genotypes .......................................................................................................................... 82

Table 4.1 Genotype frequency of the tested SNPs in the study patients................................. 138

Table 4.2 Demographic features of the two AS cohorts .............................................................. 143

Table 4.3 Univariate analysis of predictors of baseline radiographic severity ....................... 145

Table 4.4 Predicting patients who are likely to progress on follow up................................. 148
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-LAP</td>
<td>Adipocyte-derived Leucine Aminopeptidase</td>
</tr>
<tr>
<td>AP-1</td>
<td>Activated Protein 1</td>
</tr>
<tr>
<td>AR</td>
<td>Attributable Risk</td>
</tr>
<tr>
<td>ARTS-1</td>
<td>Aminopeptidase Regulating Tumor necrosis factor receptor I Shedding 1</td>
</tr>
<tr>
<td>AS</td>
<td>Ankylosing Spondylitis</td>
</tr>
<tr>
<td>ASAS</td>
<td>Assessment of Spondyloarthritis International Society</td>
</tr>
<tr>
<td>AUC</td>
<td>Area Under Curve</td>
</tr>
<tr>
<td>BASDAI</td>
<td>Bath Ankylosing Spondylitis Disease Activity Index</td>
</tr>
<tr>
<td>BASFI</td>
<td>Bath Ankylosing Spondylitis Functional Index</td>
</tr>
<tr>
<td>BME</td>
<td>Bone Marrow Edema</td>
</tr>
<tr>
<td>β2m</td>
<td>β2-Microglobulin</td>
</tr>
<tr>
<td>CAST</td>
<td>Calpastatin</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence Interval</td>
</tr>
<tr>
<td>CIL</td>
<td>Corner Inflammatory Lesions</td>
</tr>
<tr>
<td>COXIBs</td>
<td>Cyclooxygenase II Inhibitors</td>
</tr>
<tr>
<td>CRP</td>
<td>C Reactive Protein</td>
</tr>
<tr>
<td>CTX-II</td>
<td>Cross-linking Telopeptide of type II</td>
</tr>
<tr>
<td>DDI</td>
<td>Damage Duration Index</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>DISH</td>
<td>Diffuse Idiopathic Skeletal Hyperostosis</td>
</tr>
<tr>
<td>DKK1</td>
<td>Dickopf 1</td>
</tr>
<tr>
<td>DMARDs</td>
<td>Disease modifying anti-rheumatic drugs</td>
</tr>
<tr>
<td>EMS</td>
<td>Early Morning Stiffness</td>
</tr>
<tr>
<td>ERAAP</td>
<td>Endoplasmic Reticulum Aminopeptidase Associated with Antigen Processing</td>
</tr>
<tr>
<td>ERAP1</td>
<td>Endoplasmic Reticulum Aminopeptidase 1</td>
</tr>
<tr>
<td>ERAP2</td>
<td>Endoplasmic Reticulum Aminopeptidase 2</td>
</tr>
<tr>
<td>ESR</td>
<td>Erythrocyte Sedimentation Rate</td>
</tr>
<tr>
<td>FBS</td>
<td>Fetal Bovine Serum</td>
</tr>
<tr>
<td>FHC</td>
<td>Free Heavy Chain</td>
</tr>
<tr>
<td>GFP</td>
<td>Green Fluorescence Protein</td>
</tr>
<tr>
<td>GPR</td>
<td>Global Posture Re-education</td>
</tr>
<tr>
<td>GSK-3β</td>
<td>Glycogen Synthase Kinase-3β</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome Wide Association Study</td>
</tr>
<tr>
<td>HWE</td>
<td>Hardy Weinberg Equilibrium</td>
</tr>
<tr>
<td>HLA</td>
<td>Human Leukocyte Antigen</td>
</tr>
<tr>
<td>HSP60</td>
<td>Heat Shock Protein-60kD</td>
</tr>
<tr>
<td>HUGO</td>
<td>Human Genome Organization</td>
</tr>
<tr>
<td>IBD</td>
<td>Inflammatory Bowel Disease</td>
</tr>
<tr>
<td>IBP</td>
<td>Inflammatory Back Pain</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>ICC</td>
<td>Inter-class Correlation Coefficients</td>
</tr>
<tr>
<td>IC-FHC</td>
<td>Intracellular Free Heavy Chain</td>
</tr>
<tr>
<td>IFNγ</td>
<td>Interferon Gamma</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>Int1</td>
<td>Integration 1</td>
</tr>
<tr>
<td>IQR</td>
<td>Inter Quartile Range</td>
</tr>
<tr>
<td>JAK</td>
<td>Janus Kinase</td>
</tr>
<tr>
<td>KIR</td>
<td>Killer cell Immunoglobulin-like Receptors</td>
</tr>
<tr>
<td>L-Amc</td>
<td>Leucine-7-amino-4-methylcoumarin</td>
</tr>
<tr>
<td>LIR</td>
<td>Leukocyte Immunoglobulin-like Receptors</td>
</tr>
<tr>
<td>LD</td>
<td>Linkage Disequilibrium</td>
</tr>
<tr>
<td>LMP2</td>
<td>Large Multifunctional Peptidase 2</td>
</tr>
<tr>
<td>L-RAP</td>
<td>Leukocyte-derived Arginine Aminopeptidase</td>
</tr>
<tr>
<td>LRP</td>
<td>LDL-receptor like protein</td>
</tr>
<tr>
<td>MAF</td>
<td>Minor Allele Frequency</td>
</tr>
<tr>
<td>MFI</td>
<td>Mean Fluorescence Intensities</td>
</tr>
<tr>
<td>MHC</td>
<td>Major Histocompatibility Complex</td>
</tr>
<tr>
<td>MMP3</td>
<td>Matrix Metalloproteinase 3</td>
</tr>
<tr>
<td>MMTV</td>
<td>Mouse Mammary Tumor Virus</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>MS</td>
<td>Mass Spectroscopy</td>
</tr>
<tr>
<td>mSASSS</td>
<td>Modified Stoke’s Ankylosing Spondylitis Spine Score</td>
</tr>
<tr>
<td>ΔmSASSS</td>
<td>Change in Modified Stoke’s Ankylosing Spondylitis Spine Score</td>
</tr>
<tr>
<td>mTNF</td>
<td>Membrane-bound Tumor Necrosis Factor-α</td>
</tr>
<tr>
<td>MTX</td>
<td>Methotrexate</td>
</tr>
<tr>
<td>NASC</td>
<td>North American Spondylytis Consortium</td>
</tr>
<tr>
<td>NC</td>
<td>Negative Control</td>
</tr>
<tr>
<td>NF-κB</td>
<td>Nuclear Factor-κB</td>
</tr>
<tr>
<td>OR</td>
<td>Odds Ratio</td>
</tr>
<tr>
<td>PASW</td>
<td>Predictive Analytics Software</td>
</tr>
<tr>
<td>PBMC</td>
<td>Peripheral Blood Mononuclear Cells</td>
</tr>
<tr>
<td>pGR</td>
<td>Peptide- Glucagon Receptor</td>
</tr>
<tr>
<td>PI</td>
<td>Propidium Iodide</td>
</tr>
<tr>
<td>PILS-AP</td>
<td>Puromycin-Insensitive Leucyl Aminopeptidase</td>
</tr>
<tr>
<td>pLMP2</td>
<td>Peptide- Latent Membrane Protein 2</td>
</tr>
<tr>
<td>pVIPR</td>
<td>Peptide- Vasoactive Intestinal Polypeptide Receptor</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomized Control Trial</td>
</tr>
<tr>
<td>R-Amc</td>
<td>Arginine-7-amino-4-methylcoumarin</td>
</tr>
<tr>
<td>RBMX</td>
<td>RNA-Binding Motif gene, X chromosome</td>
</tr>
<tr>
<td>ROC</td>
<td>Receiver Operating Characteristic Curve</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>SDC</td>
<td>Smallest Detectable Change</td>
</tr>
<tr>
<td>SIJ</td>
<td>Sacroiliac Joints</td>
</tr>
<tr>
<td>siRNA</td>
<td>Short Interfering RNA</td>
</tr>
<tr>
<td>SNP</td>
<td>Single Nucleotide Polymorphisms</td>
</tr>
<tr>
<td>STIR</td>
<td>Short Tau Inversion Recovery</td>
</tr>
<tr>
<td>sTNF</td>
<td>Soluble Tumor Necrosis Factor-α</td>
</tr>
<tr>
<td>TAP</td>
<td>Transporter associated with Antigen Processing</td>
</tr>
<tr>
<td>TASC</td>
<td>The Australo-Anglo-American Spondylitis Consortium</td>
</tr>
<tr>
<td>TCF</td>
<td>T Cell Factor</td>
</tr>
<tr>
<td>TIFF3</td>
<td>Tricorn-Interacting Factor F3</td>
</tr>
<tr>
<td>TNFα</td>
<td>Tumor Necrosis Factor-α</td>
</tr>
<tr>
<td>UPR</td>
<td>Unfolded Protein Responses</td>
</tr>
<tr>
<td>WB</td>
<td>Western Blot</td>
</tr>
<tr>
<td>Wg</td>
<td>Wingless</td>
</tr>
<tr>
<td>WTCCC</td>
<td>Welcome Trust Case Control Consortium</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

A part of this chapter describing the biology of Endoplasmic Reticulum Aminopeptidase 1 was published as a review in Nature Review’s Rheumatology. The citation for this article is: Haroon N and Inman RD. Endoplasmic reticulum aminopeptidases: biology and pathogenic potential. Nat Rev Rheumatol. 2010 Aug;6(8):461-7. RDI revised and approved the draft.

1 Ankylosing Spondylitis

Ankylosing spondylitis (AS) is a chronic inflammatory rheumatic disease characterized by arthritis affecting predominantly the joints of the spine. AS classically begins with inflammation in the sacroiliac joints but over time affects the rest of the spine. Uncontrolled inflammation can eventuate in osteoproliferation, with pathologic new bone formation and eventually bony ankylosis of the spine. The cause of AS is unknown and the basic mechanisms underlying chronic inflammation remain undefined.

Patient with AS can have significant impairment of their quality of life[1-4]. Pain and stiffness can result in functional and work disability[5-10]. Long term sequelae including increased risk of cardiovascular morbidity and mortality are now well recognized in AS[11-14].

1.1 History

The existence of AS has been traced to biblical times with at least three of the Pharaohs of Egypt having suffered from the disease[15]. In 1963 Bernard Connor, an
Irish Physician, described an unusual skeleton from a French cemetery in which the sacrum, ilium, lumbar and thoracic spine as well as some ribs were fused like a single bone[16]. Remarkably by dissection he identified that ankylosis occurred along the annulus fibrosus while the nucleus pulposus was free of such change. This is consistent with the current concept of ankylosis in AS. AS is also known as Marie-Strümpell disease, after the clinicians who published the first clinical reports of AS[17,18]. Forestier demonstrated sacroiliitis and syndesmophytes formation, the two cardinal features of AS[19,20]. Syndesmophytes refer to new bone formation seen at the vertebral corners. These can be seen as bony spurs in x-rays developing from the upper and lower vertebral corners, both anteriorly and posteriorly.

1.2 Epidemiology

The prevalence of AS varies from 0.2-0.8%, depending on the population studied and carries with it health care costs which are comparable to rheumatoid arthritis[21-23]. The peak onset of AS is in the third decade of life and imposes on patients significant pain and disability at a key time of life. Males are affected two to three times more frequently than females[24]. This ratio might however be an overestimate as the degree of suspicion of AS in women is lower and they usually have a longer delay in diagnosis compared to men[25]. With improved diagnostic criteria and the advent of MRI, more females with AS may be identified. Females may have more pain and radiographic changes in the neck while men have more lumbar involvement[26,27]. This could also reduce the degree of suspicion of AS in females.
There is a strong association of AS with HLA-B27 with 90% of AS patients having the gene\[28\]. The prevalence of AS tends to parallel the frequency of HLA-B27 in the population\[29\]. HLA-B27 is almost absent in the Australian aboriginal population where AS is virtually non-existant\[30\]. In contrast to a low prevalence of HLA-B27 and AS in Japan\[31\] and Africa\[32-35\], Alaskan Eskimos, Inuits and Siberian Chukotkas have a high prevalence of both HLA-B27 and spondyloarthropathies, reaching 10% in Haida Indians\[36-41\].

1.3 Clinical Features

1.3.1 Inflammatory Back Pain (IBP)

Inflammatory Back Pain (IBP) is the characteristic feature distinguishing inflammatory arthritis of the spine from mechanical causes of back ache. Calin et al. established criteria to identify IBP with 95% sensitivity and 85% specificity in 1977 \[42\]. The criteria would be satisfied if 4 out of 5 of the following features were present: (i) age of onset less than 40 years (ii) insidious onset (iii) more than 3 months duration (iv) improvement with exercise and (v) early morning stiffness. These criteria were further modified by Rudwaleit et al to improve the sensitivity and specificity, with the emphasis being on differentiating patients with AS from those with mechanical back pain \[43\].

Despite these criteria for identifying IBP, the physician’s opinion in this regard is very subjective. To overcome this hurdle, the aim of developing new criteria for IBP was to approximate expert opinion as closely as possible. If this goal is met, early diagnosis of axial spondyloarthritis can become a reality. The Assessment of
Spondyloarthritis International Society (ASAS) has now developed new criteria for IBP with this in mind [44]. Various clinical history items were considered in developing the new criteria, with the gold standard being expert opinion. Thirteen AS experts from 9 countries in Europe and North America evaluated 20 patients and formulated an opinion whether IBP was present or not, with no knowledge of the final diagnosis of the patient. All patients had chronic back pain and were undergoing diagnostic work up at a rheumatology clinic. This kind of analysis removes the heavy bias affecting decision-making in IBP with prior knowledge of the diagnosis.

The new ‘IBP according to experts’ criteria had the following five parameters: (1) age at onset <40 years (2) insidious onset (3) improvement with exercise (4) no improvement with rest and (5) pain at night with improvement on getting up [44]. Despite the sound methodology used, the specificity and sensitivity of these criteria were not significantly different from previous criteria in a validation study on 648 patients with chronic back pain [44]. It should be noted that some cardinal parameters were not retained after the procedure. Early morning stiffness (EMS) was not included in the new criteria. It is known that EMS can be seen both in patients with IBP and with mechanical pain, but if quantified as greater or less than 30 minutes in duration, it might be a useful differentiating parameter. However in this study EMS did not stand out in the regression analysis. Similarly pain in the second half of the night might be a more specific feature compared to just nocturnal pain alone. This timing of night pain was not specified in this study [44]. Unlike traditional practice of combining worsening of pain with rest and improvement with exercise as a single parameter, here they have been included separately as both are independently associated with IBP. Despite the
inherent problems of using ‘IBP according to experts’ as the gold standard, this undertaking is a step in the right direction and continued efforts in this direction could give us better criteria that can contribute to decreased diagnostic delays in AS.

1.3.2 Peripheral Arthritis

Peripheral arthritis is seen in up to 30% of patients with AS. Classically, it is an asymmetrical oligoarthritis with a predilection for the lower extremities. Thus patients usually present with pain and swelling affecting knees and ankles. Other joints can also be affected, but typically less than 6 joints are affected at a time. Arthritis can be associated with tenosynovitis and enthesitis. Common enthesial sites affected are Achilles tendon insertion, plantar fasciitis, tibial tuberosity, peripatellar tendon insertions, spinal enthesis, iliac crest and iliac spines as well as costochondral junctions. Hip joint involvement is a marker of severity of AS. It is more common with younger age of onset but can occur at any point in the course of AS.

1.3.3 Extra-articular Features

The common extra-articular manifestations associated with AS are inflammatory bowel disease (IBD), uveitis and psoriasis. Acute anterior uveitis is seen in up to 40% of AS patients [45]. Classical presentation includes unilateral blurring of vision, photophobia, headache and redness. An acute attack of uveitis can last for a few weeks. HLA-B27 positive patients have a higher risk of uveitis in AS [46]. Uveitis can be recurrent and sometimes chronic with damage to the eyes, if not treated appropriately and promptly.
Inflammatory bowel disease (IBD) and psoriasis occur in approximately 10% of AS cohorts, respectively [47,48]. Asymptomatic microscopic colitis can occur in a higher percentage (40-60%) of patients with spondyloarthritis compared to other inflammatory arthritis [49-53]. Primary AS differs from that associated with psoriasis and IBD with an early onset of illness and more severe spinal restriction [54]. Chronic inflammation in the gut is associated with higher prevalence spinal restriction, bamboo spine and sacroiliitis [55].

Other extra articular manifestations include fever, fatigue, aortic insufficiency, cardiac conduction defects, and pulmonary fibrosis. There is a higher incidence of IgA nephropathy in AS. Amyloidosis is rare in AS. Somewhat paradoxically in the context of the bone forming tendency in AS, osteoporosis and osteopenia is frequent [56,57]. Males and patients with decreased functional capacity have lower bone density[58]. There is a high prevalence of vertebral fractures that are likely underdiagnosed [59]. Age, male gender, poor functional capacity, higher occiput to wall distance and higher radiographic severity are risk factors for fracture in AS [59].

1.4 Radiographic Features

Bilateral sacroiliitis identified by x-ray is ordinarily required to make a diagnosis of AS. Hence there can be a delay in diagnosis till radiographic changes appear. Sacroiliitis manifests as subchondral sclerosis, joint erosions and ankylosis in progressively increasing grades of severity. A basic radiographic assessment of AS should include X-rays of the spine as well. Erosions of the vertebral corners can be seen. When it is present on both anterior corners of the vertebral body, the normal anterior concavity is lost and this is referred to as squaring. Post-inflammatory changes
and new bone formation is thought to mediate the sclerosis of vertebral corners and they appear as ‘shiny corners’.

Exuberant new bone formation manifests as bony spurs developing from the vertebral corners. These bony spurs are called syndesmophytes when they extend from the vertebral corner in the direction of the spinal axis. They can grow and ‘bridge’ across to the adjacent vertebral corner when they are called ‘bridging syndesmophytes’. In addition to the anterior elements of the spine, posterior elements including facet joints can undergo ankylosis and need to be assessed. The end result can be extensive bridging syndesmophytes from cervical to sacral vertebrae and this is called a ‘bamboo spine’.

The introduction of Magnetic Resonance Imaging (MRI) in the diagnosis of spondyloarthritis has greatly improved the sensitivity for diagnosing AS. In the new ASAS criteria for axial spondylarthritits, MRI evidence of inflammation in the sacroiliac joints is sufficient. Such patients, unless they have sacroiliitis by X-ray imaging, cannot be labeled as AS as defined by the modified New York criteria discussed below. The short tau inversion recovery (STIR) sequences are special sequences which suppress fat signals and bring out signals from free water as present in bone marrow edema (BME) [60]. In T2 STIR sequences active inflammation can be seen as hyperintense areas in the spine with the main areas being vertebral corners (corner inflammatory lesions or CILs), SI joints, facet joints, costovertebral and costotransverse joints. In T1w sequences, structural changes like erosions, ankylosis and fat replacement can be seen. Erosions are seen in the vertebral corners as well in the sacroiliac joints. Fatty replacement can be seen in the vertebral corners and in the periarticular areas of
the sacroiliac (SI) joints. When present as hyperintense lesions at the vertebral corners they are referred to by some authors as ‘fatty Romanus lesion’ (as the location is the same as Romanus lesions in X-ray) [61]. It should however be noted that erosions are better seen with CT scan [62].

The ASAS/OMERACT MRI group has published consensus criteria for diagnostic MRI in axial SpA [63]. Neither active vertebral lesions nor chronic sacroiliac and vertebral lesions qualify as a diagnostic MRI finding. BME should be located in the subchondral or periarticular areas of the SI joint. At least two separate BME lesions in one slice or one BME lesion that is visible in two consecutive slices is the minimum intensity of lesion required [63]. MRI can also be used as a tool to judge efficacy of anti-TNF agents as it is sensitive to change in spinal inflammation. In the ASSERT study, 194 patients on infliximab (dose of 5mg/kg) were compared to 72 on placebo. After 6 months of therapy, there was almost complete elimination of MRI inflammation in the infliximab group with no significant change in the placebo group. In the CANDLE study which used low dose infliximab (3mg/kg) [64] there was significantly more reduction in MRI inflammation even at 12 weeks in patients on active therapy compared to placebo (p<0.002). All patients showed a decrease in inflammation in the Infliximab group. This effect is by no means limited to Infliximab and has been shown with etanercept and adalimumab [65-67].

1.5 Classification Criteria

1.5.1 Modified New York Criteria
The modified New-York criteria (Table 1.1) are the currently accepted standard for the diagnosis and classification of AS [68]. A diagnosis of AS is made if sacroiliitis is associated with low back pain or restriction of spinal mobility. The X-ray criterion is required to make a diagnosis of AS. Hence, although specificity is good, the sensitivity decreases, especially in early stages of AS. This raises the criticism that the Modified NY criteria leads to a delay in diagnosis of AS. To meet the demand for more sensitive criteria, the ASAS group has developed new criteria for classifying axial spondyloarthritis.

1.5.2 ASAS Criteria for Axial Spondyloarthritis

The revised classification criteria for axial SpA (Table 1.2) was proposed and validated by ASAS [69,70]. This effort from the ASAS group stems from an unmet need for early diagnosis of SpA with predominant axial symptoms. To make a diagnosis of AS according to the modified New York criteria, there should be X-ray evidence of bilateral grade II or at least unilateral grade III sacroiliitis [68]. The delay in diagnosis of AS can be as high as 8 to 10 years [25,71,72]. It might be many years before any changes appear in X-rays and it is now known that MRI can precede X-ray changes by as much as 8 years [73]. There is tremendous progress in early diagnosis of RA and the benefits of the same are being seen in the form of better control and even healing of erosions. There is no such data in AS and the first step in initiating a trial is to define early SpA. The development of the ASAS criteria for axial SpA is an attempt at resolving this gap.
Table 1.1 Modified New York criteria for ankylosing spondylitis

I. Clinical criteria
a) Low back pain and stiffness for more than 3 months which improves with exercise, but is not relieved by rest.
b) Limitation of motion of the lumbar spine in both the sagittal and frontal planes.
c) Limitation of chest expansion relative to normal values corrected for age and sex.

II. Radiologic criterion
Sacroiliitis grade 2 bilaterally or sacroiliitis grade 3-4 unilaterally.

A diagnosis of AS is made if the radiographic criteria is associated with one clinical criteria.
Table 1.2 ASAS classification criteria for axial Spondyloarthritis

**Entry Criteria:** Patients should be 45 years of age or younger and have back pain for more than 3 months for this criteria to be applicable.

These patients can be classified as axial SpA if they have either

A. Imaging evidence of sacroiliitis and at least 1 other SpA feature

OR

B. HLA B27 and at least 2 other SpA features

**Imaging evidence:** Definite sacroiliitis according to modified New York criteria or MRI showing definite bone marrow edema or osteitis in sacroiliac joints.

**SpA features:**
1. Inflammatory back pain
2. Arthritis
3. Enthesitis
4. Uveitis
5. Dactylitis
6. Psoriasis
7. Inflammatory bowel disease
8. Good NSAID response
9. Family history of SpA
10. HLA B27
11. Elevated CRP
For the first time, MRI has been introduced as an accepted imaging modality for identifying sacroiliitis and a patient can be classified as axial SpA without any X-ray changes (Table 1.2). Another novelty of this classification is a non-imaging arm, wherein a patient with HLA-B27 and no radiographic evidence of sacroiliitis can potentially qualify to have SpA [69,70]. There are several criticisms and the predominant one is that it is too inclusive. A good example would be that of a person with long standing mechanical back pain and a family history of SpA. There is a 7% chance that this person is HLA B27 positive and he could be erroneously classified as axial SpA.

1.6 Treatment of Ankylosing Spondylitis

In AS, the goal of treatment is control of symptoms. Pain, stiffness, fatigue, peripheral arthritis, dactylitis, enthesitis and tenosynovitis usually responds to the current modalities of treatment available for AS patients. Unfortunately we do not have any proven modalities for disease modification or cure. There is some evidence with anti-inflammatory agents that treatment over prolonged periods of time may result in some slowing of radiographic damage. Assessment of treatment response is based on patient reported outcome measures like the Bath AS Disease Activity Index (BASDAI), Bath AS Functional Index (BASFI) and Total Back Pain (TBP) score. BASDAI is calculated from a questionnaire that assesses symptoms like back pain, entheseseal/joint pain, duration and severity of EMS. The score is averaged over a 0-10 scale with a score of 4 indicating inadequate control of symptoms. When assessing treatment response, the minimum clinically significant difference in BASDAI is estimated to be 20% or 2 units in a scale of 0-10. Thus if the BASDAI improves from 6 to 4 during treatment, this is a clinically significant change.
Improvement in functional ability, or the ability to do routine and leisure activities is captured in a questionnaire and the average score gives the BASFI. Improvement in BASFI gives us an idea if the treatment actually changed the ability of the patient to do physical activity.

1.6.1 Non-pharmacological treatment

There has been a surge of trials on the effect of physiotherapy in AS during the last decade. There is strong evidence for clinically significant improvement with regular physiotherapy and multimodal exercise therapy [74,75]. The aim of physiotherapy is to improve and maintain global health and improve the functional status of patients. Group therapies and supervised therapies are more effective than home based physiotherapy [76]. Global posture re-education (GPR) refers to a specific physiotherapy technique using gentle exercises on muscles groups aiming to correct joint alignment, muscle strengthening, and stretching of muscles that have undergone contracture. GPR resulted in better BASDAI and BASFI scores than routine physiotherapy in a recently published randomized controlled trial (RCT) [77]. Occupational therapy and balneotherapy are additional adjunctive modalities of treatment in AS [78,79].

1.6.2 Non-steroidal anti-inflammatory drugs

NSAIDs have remained the mainstay of treatment of AS for many years. NSAIDs effectively reduce the signs and symptoms of AS [80,81]. This is sufficiently reproducible that response to NSAIDs has been postulated as a classification criterion
for inflammatory back pain. Adverse effects like gastrointestinal toxicity and nephrotoxicity remain concerns however. With the introduction of selective cyclooxygenase II inhibitors (COXIBs), the concern of serious GI side effects has decreased [82,83]. Continuous use of NSAIDs help in abiding to a good exercise program and indirectly improves the disease outcome. Recent data on the reduction of radiological progression with continuous use instead of as-needed use of NSAIDs is also very provocative [84].

1.6.3 Disease modifying anti-rheumatic drugs (DMARDs)

Sulfasalazine is effective in patients with AS and peripheral arthritis, however, in pure axial disease, its efficacy is limited [85]. A few studies and a meta-analysis have shown some benefit in reducing the inflammatory pain in peripheral joints and thus helping reduce the NSAID requirement [86,87]. Methotrexate (MTX) has been used with some benefit in patients with AS. There are conflicting reports on the role of MTX in peripheral as well as axial disease [88-90]. The only RCT of MTX in AS showed no benefit of AS either on axial or peripheral arthritis [91,92]. Leflunomide was not useful in spinal symptoms in an open label and a double blind RCT in AS [93,94].

1.6.4 Biological Therapy

Biological therapy in AS primarily involves blocking tumor necrosis factor α (TNFα). Currently there are 4 FDA-approved agents for blocking TNF: infliximab, etanercept, adalimumab and golimumab. These agents have proved to be effective for controlling
inflammation in AS. The overall efficacy of all anti-TNF agents is comparable, with a 50% reduction in BASDAI in 50-80% of patients.

Infliximab (cA2, Remicade, Fa Essex/Centocor) is a chimeric monoclonal IgG1 antibody formed from human constant and murine variable regions. It binds to both soluble (sTNF) and membrane bound forms (mTNF) of TNF, neutralizing its action. There are conflicting reports on induction of apoptosis of T cells and macrophages after infliximab binds to the membrane bound receptor [95-97]. Recently it was also shown to induce regulatory macrophages [98]. It is most commonly used at a dose of 5 mg/kg per dose on 0, 2, 6 and 14 weeks and every 8 weeks thereafter. Significant reduction in disease activity, improvement in productivity and quality of life have been reported with infliximab.

Etanercept (Enbrel®, Fa Lederle/Wyeth) is a recombinant human soluble TNF receptor. It is formed by the fusion of the extracellular portion of the p75 component of the TNF receptor and the Fc portion of IgG1. Unlike infliximab, it affects sTNF predominantly with minimal effect on mTNF. Moreover, it does not cause complement fixation and does not induce lysis of TNF-bearing cells such as monocytes and macrophages. It does not decrease T-cell production of TNFα [99]. The standard dose of etanercept has been 25 mg twice weekly, but 50 mg once weekly dose was found to be at least as effective as the twice weekly dose [100].

Adalimumab (HUMIRA®, Abbott) is a fully humanized monoclonal antibody against TNF [101]. Using phage display technology TNFα-specific human heavy and light chain variable regions were derived and fused to the Fc portion of human IgG1. Owing to the
exclusive human sequences, the chances of immunogenicity and antibody formation are less. Golimumab is an IgG1κ, fully human monoclonal anti-TNF antibody administered as subcutaneous injections once a month. It is administered at a dose of 50 mg once monthly [102].

1.7 Pathogenesis

The exact pathogenesis of ankylosing spondylitis has not been resolved. The role of genetics is clear from the high concordance of AS in twins [103,104]. However, even in monozygotic twins the concordance is only 40-50%[104]. Thus there are environmental influences in the onset of AS.

1.7.1 Infections and Pathogenesis of AS

Infections are known to trigger reactive arthritis. However a link between infections and AS is not very clear. There are several theories supporting a place for bacterial infections in the pathogenesis of AS but the exact role is not understood. Attempts at identifying bacterial DNA from Si joints using a nested PCR technique were unsuccessful [105].

Among the numerous bacterial infections that have been suggested to play a role in the pathogenesis of AS, Klebsiella pneumoniae is the most strongly implicated. There is sequence similarity of the nitrogenase and pullulanase D proteins in Klebsiella with HLA-B27 [106]. Thus molecular mimicry leads to cross reacting antibodies that develop following a Klebsiella infection. Cross-reactivity with HLA-B27 can lead to abnormal immune responses. Both CD4 and CD8 T cell responses have been shown in
AS patients against the 60 kD heat shock protein (HSP60) of Klebsiella [107]. Humoral responses against HSP60 as well as *Klebsiella pneumoniae* nitrogenase have been demonstrated [108,109]. Antigen processing of these Klebsiella proteins followed by presentation on HLA-B27 is likely as chloroquine and Brefeldin A block the cellular response to the antigens [107]. This is further strengthened by the observation that HLA-B27 positive individuals have more humoral response to these antigens compared to HLA-B27 negative individuals [109]. The levels of IgA and IgG antibodies against the Klebsiella proteins have been shown to correlate with C-reactive protein (CRP), a marker of inflammation [108]. There is also a suggestion that abnormal persistence of the bacteria is primarily responsible rather than the immune response. By Elispot assay it was shown that twins who have AS have reduced reactivity against *K pneumoniae*, *S pyogenes* and *C albicans* compared to the healthy monozygotic twins, even though the humoral response is similar [110]. The fact that HLA-B27 transgenic rats are normal when raised in a germfree environment but develop a spondyloarthritis-like illness otherwise, suggests that infection or some other second signal like Toll Like Receptor (TLR) activation may be important in pathogenesis [111].

The above-mentioned evidence is quite compelling, yet a role for Klebsiella in the pathogenesis of AS has not been established due to the equally prevalent negative reports. In an extensive study looking at 4 candidate organisms- *K. pneumoniae*, *Salmonella typhimurium*, *Yersinia enterocolitica* and *Chlamydia trachomatis*, there was no difference in the cellular or humoral response between AS patients and their family members [112]. There is also evidence to show that Klebsiella isolation from the gut does not differ in AS and any other rheumatic disease [113]. The same study did not
find any association of the antibody levels with inflammatory markers[113]. Invoking molecular mimicry would imply an element of autoimmunity in AS but there is no evidence of an autoimmune process in the vast majority of these patients.

1.7.2 Cytokines in AS

Tumor necrosis factor α (TNF) is a pleiotropic cytokine secreted mainly by monocytes and macrophages. It is involved in numerous biological processes including inflammation, apoptosis, septic shock, immunity, arthritis and other autoimmune diseases. It stimulates prostaglandin production by fibroblasts and induces adhesion molecules, chemokines and protease enzymes [114-116]. Acting through two receptors, TNFR1 (p55) and TNFR2 (p75), followed by a complex interplay of intracellular molecules, TNF activates C-Jun N-terminal kinase (JNK) and IκB Kinase leading to activation of activated protein 1 (AP-1) and nuclear factor-κB (NF-κB) respectively which are important transcription factors mediating inflammation [117,118].

In the early 90s, serum TNF levels were reported to be high in AS and subsequently TNF-producing mononuclear cells were demonstrated in the sacroiliac joints [119,120]. It is also expressed in high concentration in the gut of both ulcerative colitis and Crohn’s disease patients [121]. Up to 60% of patients with AS have pathological changes in the GI tract and the lesions are strikingly similar to those in Crohn’s disease [51,52,55,122]. TNF blockers lead to significant improvement in inflammatory bowel disease and reduce spinal inflammation. Sensitive techniques like MRI have shown a decrease in inflammation following the use of anti-TNF agents [64].
Among the several polymorphisms of the TNFα gene, those at positions -238 and -308 are well studied [123]. These polymorphisms lead to nucleotide changes in the putative transcription regulatory region and could possibly alter TNFα production. TNF transgenic mice with a 3'-modified human TNF gene have overexpression of TNF. These mice develop sacroiliitis like AS in addition to polyarthritis [124].

A metanalysis of 8 separate studies showed no association of AS with TNF promoter polymorphisms [125]. This negative association was replicated by another metanalysis of 14 studies comprising 1,766 AS patients and 2,114 controls [126]. The recent genome-wide scans have suggested an association with TNFRSF1A [127,128]. It is already known that rheumatic disease patients with a TNFα -308 G/G genotype respond better to anti-TNF treatment than those with A/A or A/G genotypes independent of the diagnosis (rheumatoid arthritis, psoriatic arthritis or AS) [129]. Why a TNF gene promoter polymorphism would affect anti-TNF response is not known. Interestingly, the TNFα -308 G/G genotype has been shown to be associated with decreased TNFα production [130,131]. It is possible that patients with a tendency for lower production of TNFα respond even better with anti-TNF agents.

Other cytokines implicated in the pathogenesis of AS are IL1 and IL23. There is increasing evidence of the involvement of the Th17 pathway in AS pathogenesis. In a study on PBMC from AS and RA patients, IL-17 and IL-22 positive CD4 T cells were increased with higher secretion of IL-17 compared with healthy controls [132]. The frequency of IL-17 positive cells was higher in the facet joints of AS patients compared to osteoarthritis patients [133]. There is exciting new evidence which shows that HLA-
B27 dimer expressing Antigen Presenting Cells (APC) can stimulate the proliferation of KIR3DL2 expressing CD4 T cells. This subset of T cells is enriched for IL-17 production [134]. The unfolded protein response (UPR) resulting from unfolding of HLA-B27 in HLA-B27 transgenic rats induces the production of IL-23 a th17 cytokine [135].

A study of serum levels of IL-17, IL-23, IL-6, IL-12 and TGFβ did not differentiate active and inactive AS patients [136]. The serum levels of these cytokines were, however, higher than healthy controls. IL23R polymorphisms are associated with AS and this has been well replicated in genome wide scans [127,128,137]. A common association with Crohn’s disease again raises the issue of common pathogenic mechanisms in IBD and AS. The level of IL-23p19 expression in the cecum and distal ileum of AS patients’ is comparable to Crohn’s disease [138].

The IL1 gene cluster contains nine genes homologous to IL1A and IL1B or IL1RN. The IL1A, IL1B and IL1RN code respectively for IL1α, IL1β and IL1Ra. There are reports of association of AS with IL1RN and other IL1 gene cluster members [139-142]. These findings have been corroborated by a meta-analysis [143]. The recent genome wide scans have shown evidence of association of AS with the IL1RII [127].

In summary, cytokines are important in AS. The remarkable efficacy of anti-TNF agents in controlling inflammation but inability to control progression points towards an uncoupling of inflammation and new bone formation. It is not entirely clear if the cytokine defects are the primary event or secondary to other pathogenic events. Emerging evidence on the induction of Th17 cytokines with UPR and the KIR3DL2-B27
dimer interaction further shows the potential involvement of cytokines in AS pathogenesis.

1.7.3 Genetics of ankylosing spondylitis

1.7.3.1 Lessons from Family and Twin Studies

While the pathogenesis of AS is not well understood, a strong genetic basis for the disease is clear from the results of family and twin studies. If the risk of a disease is higher among family members compared to the general population the possibility of a genetic risk factor should be considered. This suspicion is strengthened if the risk increases with the degree of familial relationship. In a disease with strong genetic risk, the highest concordance rate will be in identical twins followed by first degree relatives. The sibling recurrence risk ($\lambda_s$) is calculated as the ratio of the risk of disease developing in siblings compared to the risk in the general population.

The $\lambda_s$ in AS has been reported to be as high as 82 [144]. Thus the recurrence risk of AS in siblings is 8.2% compared to 0.1% in the general population [144]. Similarly in twin studies, the concordance of AS in monozygotic twins is 50% compared to 25% in dizygotic twins [103]. High rates of concordance have been reported in other studies from Europe [104,145]. The heritability of AS is estimated to be over 95% thereby bringing the role of environmental factors to a minimum.

1.7.3.2 Role of HLA-B27 in Ankylosing Spondylitis

Tremendous hope for an imminent resolution of the pathogenesis of AS was raised when $HLA-B27$ was found to be present in 96% of patients with AS, compared
with 4% of controls [146]. There were suspicions initially whether HLA-B27 itself was pathogenic rather than a locus in strong linkage with the pathogenic gene. These concerns were dispelled when the HLA-B27 transgenic rats showed features quite similar to human spondyloarthritis [147]. Although the exact role of HLA-B27 in the pathogenesis of AS is not clear, several hypotheses have been put forward (Table 1.3). Many of these themes will be alluded to again in the discussion on HLA-B27 transgenic rats.

In the ‘arthritogenic peptide hypothesis’ HLA-B27 in its classical role as an MHC-I molecule presents special peptides to CD8+ T cells resulting in an arthritogenic immune response. The presentation of distinctive peptides could be defined by the unique peptide binding specificity of HLA-B27, which eventually results in a T-cell-cascade, thereby leading to joint inflammation [148,149]. In a recent mass spectrometry (MS) study, 1,268 HLA–B27 peptides derived from human proteins were identified from HLA-B*2705 expressing C1R cells [150]. Out of these peptides 569 were identified as certain based on at least 3 to 4 MS verifications. Another 582 peptides were considered probable with 1 or 2 MS verifications. Most of the identified peptides had 8–10 amino acids. Around 211 peptides were extended with 11–15 amino acids. Around 94 peptides were present in varying lengths and the variation was most commonly seen at the C- terminal end. Interestingly several peptides were derived from cartilage or bone related proteins. There were peptides with similarity to sequences from several bacteria including Salmonella, Shigella, Campylobader and Yersinia.
### Table 1.3 Hypothesis of the role of HLA-B27 in the pathogenesis of Ankylosing Spondylitis

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Arthritogenic Peptide Hypothesis</strong></td>
<td>HLA-B27 binds and presents specific peptides that can lead to arthritis</td>
</tr>
<tr>
<td><strong>Unfolded Protein Response</strong></td>
<td>HLA-B27 has a propensity to misfold and accumulate in the ER leading to an ER stress response and release of pro-inflammatory cytokines</td>
</tr>
<tr>
<td><strong>HLA-B27 dimer mediated immune response</strong></td>
<td>HLA-B27 heavy chains can form dimers that can be recognized by atypical immune receptors like KIRs and LIRs. This can lead to abnormal immune responses</td>
</tr>
<tr>
<td><strong>Molecular Mimicry</strong></td>
<td>HLA-B27 and some bacteria share peptide sequence similarities that can lead to molecular mimicry and abnormal immune response to HLA-B27</td>
</tr>
<tr>
<td><strong>Thymic Selection of T cell repertoire</strong></td>
<td>The presence of HLA-B27 can affect the selection of the T-cell repertoire and this may influence the development of T cells that are more ‘arthritogenic’</td>
</tr>
<tr>
<td><strong>HLA-B27 as a source of peptides</strong></td>
<td>HLA-B27 itself can be degraded and form peptides that can lead to inflammatory arthritis</td>
</tr>
</tbody>
</table>

ER: Endoplasmic Reticulum; KIR: Killer Cell Immunoglobulin-like Receptors; LIR: Leukocyte Immunoglobulin-Like Receptors
As Jose Lopez de Castro elegantly wrote in an accompanying editorial, peptide searches alone will not help us solve the mystery of HLA-B27 [151]. Nevertheless, these efforts help us better understand the HLA-B27 binding motif and its variations. We may be able to identify the target proteins from which the HLA-B27 bound peptides are derived. Despite all the advances in technology and sincere efforts, we have no evidence that a specific peptide bound to HLA-B27 is the cause of AS.

Crystallographic analysis of HLA-B27 revealed a typical heterodimeric structure of MHC-I, composed of a heavy chain non-covalently linked to beta 2-microglobulin (β2m) [152]. The heavy chain and β2 microglobulin are encoded in chromosomes 6 and 15 respectively. The peptide binding groove is formed by the α1 and α2 domains of the heavy chain. Typically a nonamer peptide with a single main chain conformation is bound to HLA-B27 with anchoring of the N- and C-terminal ends.

The HLA-B27 peptide binding groove has six pockets (A to F) that accommodate the side chains of the residues of the peptides. The second amino acid is almost always arginine with the side chain occupying Pocket B. The side chains of residues on position P1, P4 and P8 are oriented away from the cavity and are important in T cell responses [153]. There are unique amino acid positions in the structure of HLA-B27 including Arg2, His9, Thr24, Glu45 and Cys67 [154]. Cys67 is particularly interesting as this amino acid can form disulphide bonds with another Cys67 leading to B27 heavy chain dimers that can stabilize without β2 microglobulin [155]. The Cys67 residue is a cardinal part of pocket B [156] and it is also known to affect TCR recognition [157].

HLA-B27 has a propensity to misfold during assembly in the ER, and these misfolded B27 heavy chains may accumulate in the ER leading to an unfolded protein
response (UPR) [158]. The UPR theory brings a different perspective to the pathogenesis of AS, making it an auto-inflammatory disease. HLA-B*2705 has a slow folding rate, with misfolding and ER retention [159,160]. The resulting ER stress leads to the release of pro-inflammatory cytokines [161]. There is a correlation between the misfolded HC and the degree of UPR [160].

HLA-B27 has a greater tendency for surface expression of homodimers of free heavy chains than other class I alleles [162-165]. There are opposing studies showing that HLA-B27 dimers are recognized by KIR receptors in a peptide dependent [166] and non-dependent manner [167]. It was not known previously that KIRs are highly polymorphic. With this new knowledge, it is possible that different KIR subtypes could recognize MHC in a peptide dependent manner.

A recent study by Bowness et al. indicated possible involvement of HLA-B27 dimers in the pathogenesis of AS [134]. The study showed that antigen presenting cells expressing HLA-B27 dimers stimulate the survival and proliferation of KIR3DL2 expressing CD4+ T cells[134]. These T cells were elevated in number in AS patients and enriched for IL-17 production. A provocative finding was that 70% of the Th17 producing cells were KIR3DL2 positive, despite being a minority among all PBMC [134].

1.7.3.3 HLA-B27 subtypes and Ankylosing Spondylitis

There are more than 35 HLA-B27 subtypes described so far of which HLA-B*2705 is present in almost all populations world-wide and is the most common allele in Caucasians. HLA-B*2704 is common in Asians and HLA-B*2702 in middle-eastern
and North African population. \textit{HLA-B’2703} is most commonly seen in western Africans and differs from \textit{HLA-B’2705} by a substitution of tyrosine for histidine at position 59.

The \textit{HLA-B27} subtypes most commonly associated with AS are \textit{HLA-B*2705}, \textit{B*2702}, \textit{B*2704} and \textit{B*2707}. Interesting \textit{HLA-B*2706} that forms half of all \textit{HLA-B27} subtypes in Indonesia, Malaysia and Thailand, has not been reported in a clinically documented case of AS [168]. Similarly, \textit{HLA-B*2709} comprises 20\% of all B27 alleles in Sardinia and yet no single patient of AS has been identified with \textit{HLA-B*2709} here.

There are only minor differences between the AS-associated and non-associated \textit{HLA-B27} subtypes. \textit{HLA-B*2706} differs from \textit{HLA-B*2704} at amino acids 114 and 116. However, these amino-acids normally occupy the floor of the F pocket, within the peptide-binding groove of \textit{HLA-B27} [168]. Similarly, \textit{HLA-B*2709} differs from \textit{B*2705} at amino acid position 116 with Asp in B*2705 and His in B*2709 [28]. The peptide repertoire of the AS-associated and non-associated subtypes differs and this could be due to the changes in the peptide binding groove [168]. Similarly there are differences in the way peptides are accommodated in the groove.

\textit{HLA-B*2705}, in comparison to B*2709, is known to bind at least three different peptides (examples being derived from Vasoactive Intestinal Polypeptide Receptor (pVIPR), Latent Membrane Protein 2 of EBV (pLMP2) and Glucagon Receptor (pGR) respectively) in two different conformations [169-171]. In another study using the peptide GRFAAAIAK, it was found that the bulkier histidine in position 116 induces a movement of the peptide C-terminal Lys to form a novel salt bridge with aspartate in position 77 [172]. His in position 116 in contrast to Asp can adopt two different conformations and overall, the peptide biding groove of \textit{HLA-B*2709} is more flexible.
than that of B*2705 [172]. The differences mentioned above can induce subtle differences in the B27-peptide conformation resulting in differential immune responses in the different HLA-B27 subtypes.

Apart from studies looking at differences in the context of the arthritogenic peptide hypothesis in AS, the rate of folding of the different subtypes has been studied. When the subtypes B*2702, B*2704, B*2705, B*2706, B*2707 and B*2709 were compared, B*2702, B*2704 and B*2705 folded much more slowly than the AS-non-associated B*2706 and B*2709 subtypes [173]. However, B*2707 which is known to be associated with AS was discordant with a high folding rate comparable to the AS-neutral subtypes [173]. Similarly the accumulation of Free Heavy Chain (FHC) in the ER was discordant in only B*2707. There was increased FHC accumulation in HLA-B*2702, B*2704 and B*2705 but not in B*2706 and B*2709.

Thus several differences exist between the AS-associated and non-associated subtypes, but no definitive explanation is yet available as to why certain subtypes alone increase the risk of AS.

1.7.3.4 Lessons from the HLA-B27 transgenic rats

The HLA-B27 transgenic rats was generated by Joel D. Taurog and reported first in 1990. These rats were transgenic for human HLA-B27 and human β2-microglobulin. More than 90% of male and female rats develop disease manifestations similar to human spondyloarthritis including sacro-iliitis, peripheral arthritis, colitis, psoriasiform skin lesions and nail changes. Although the incidence in both genders is the same, male rats develop disease much earlier than females.
The level of expression of HLA-B27 correlated with the incidence of disease. Thus two strains of B27 transgenic rats 33-3 and 21-4H had high copy numbers of HLA-B27 and developed spontaneous disease. The strains 21-4L and 21-2, 25-1 and 25-6 with low B27 copy numbers did not develop any disease manifestation. The rats with high B27 copy numbers also had higher β2m expression. Almost two-thirds of newly formed HLA-B27 molecules are misfolded in the ER and are present as monomers or dimers [159]. Misfolding of HLA-B27 and stimulation of the unfolded protein response (UPR) is seen in the HLA-B27 transgenic rats [161]. The UPR response in HLA-B27 rats was found to be associated with increased IL-23 production linking the pathogenesis of manifestation in rats to that of human AS and the Th-17 pathway [135].

There is a possibility that disease state itself induced a higher expression of HLA-B27. However, if inflammation is cause by inducing adjuvant induced arthritis in early days of the B27 transgenic rats, there is no increase in B27 expression. Similarly, in rats with both HLA-A2 and HLA-B27, only HLA-B27 expression is higher with disease onset and HLA-A2 levels remain the same. Thus a generalized increase in MHC expression as a response to inflammation is unlikely. A threshold effect has been suggested to explain the requirement of higher B27 expression, such that a certain level of B27-peptide complexes is required to induce a pathological immune response.

If unfolding of HLA-B27 is the central pathogenic event, the disease should be preventable by interrupting B27 from unfolding. For this purpose, β2m expression was increased in the rat but paradoxically the arthritis worsened and colitis improved [174]. This was despite the fact that HLA-B27 misfolding and UPR was less. A similar
discordance between arthritis and colitis is also seen with regards to peptide specificity. A specific influenza peptide when over expressed in HLA-B27 transgenic rats, led to more arthritis but no effect on colitis [175]. When the two findings above are seen together there is a possibility that the pathogenesis of arthritis could be peptide mediated while that of colitis could be due to the UPR.

Although the HLA-B27 transgenic rats appear to provide solid evidence for HLA-B27 being sufficient in the pathogenesis of AS, complexities soon became evident. These rats were asymptomatic if bred in a germ free state [111]. Based on the classical function of MHC-I, HLA-B27 would be expected to present pathogenic peptides to CD8 T cells. However the HLA B27 transgenic rats continued to have no amelioration of the disease despite a complete absence of CD8 T cells, in the HLA-B27 CD8-/− rats [176]. This however does not exclude the possibility of a CD8 negative MHC-I restricted T cells response as has been reported elsewhere [177]. Transfer of CD4 cells to transgenic nude rats can lead to reconstitution of disease manifestations of the HLA-B27 transgenic rats [178]. Another layer of confusion was added when β2 microglobulin deposition led to arthritis independent of HLA-B27 heavy chains [179].

1.7.4 Inflammation and New Bone Formation in AS

Even though there is a significant inflammatory process in AS, what differentiates it from other forms of inflammatory arthritis is new bone formation. The pathogenesis of new bone formation in AS is still not understood. Although it is tempting to state that inflammation is followed by new bone formation, the evidence for this is scanty. Due to the inaccessibility of tissue and the long duration of follow up
required to answer some of these compelling questions, there is still no definite answer regarding the relationship of inflammation to bone formation. With the introduction of highly sensitive imaging techniques like MRI it is possible to identify inflammation non-invasively and then follow up these patients for the development of syndesmophytes. There are at least 2 such studies done on patients treated with anti-TNF agents that have given us some clues.

In the first study on 39 AS patients followed up for 2 years, out of 922 vertebral edges analyzed, syndesmophytes developed in 26 [180,181]. Inflammation was identified in 153 corners at baseline and significantly more syndesmophytes developed in these corners than in those without inflammation (6.5% vs 2.%; p=0.002). This would seem to be a clear indication that inflammation precedes syndesmophyte formation. However, when we consider all syndesmophytes at follow up and look at baseline, there were more syndesmophytes in corners which had no inflammation at baseline compared to those with (62% vs 38%; P = 0.03). There was no clear cut explanation for these data, but based on the evidence that corners with inflammation prospectively led to new bone formation, Baraliakos et al. concluded that suppression of inflammation could help prevent syndesmophyte formation.

MRI is very sensitive in picking up inflammation, but the sensitivity is not 100%. Spondyloarthritis is well known to have a fluctuating course and an MRI evaluation at one point may not reflect inflammation that was present in a corner on some other time. Moreover, there are issues of inter-reader reliability. Hence it is possible that the
corners with syndesmophyte formation had at some point of time inflammation, but this has not been proven.

In the second study on AS patients, syndesmophytes developed more frequently in vertebral corners with baseline inflammation than in those without (20% vs 5.1%)[181]. Contrary to the proposal of the previous study, syndesmophytes developed more frequently at the corners in which inflammation resolved (14.3%) than at those where inflammation persisted (2.9%). This finding has led the lead author Walter Maksymowych to propose a ‘TNF-brake hypothesis’ according to which inflammation inhibits new bone formation.

1.7.5 DKK-1 in the Pathogenesis of Ankylosing Spondylitis

The term Wnt is derived from a combination of Wg (wingless) and Int1 (Integration-1), two Wnt genes discovered initially. Mutations of a gene in Drosophila were found in wingless flies, and it was called ‘wingless’ originally[182]. Similarly Int1 (integration 1) was activated in mice with mouse mammary tumor virus (MMTV) induced breast tumors[183]. From the early studies, the importance of Wnt pathway in limb development, antero-posterior axial patterning and eye formation was established[182,184,185]. It is now known that the Wnt signaling pathway is critical in the regulation of cell growth, differentiation and death[182,184-186].

The canonical Wnt-signalling pathway involves signaling through the surface receptors LDL-receptor like protein (LRP) -5 and 6 along with Frizzled. In the inactive state (Figure 1.1 A), β-catenin, the key protein that mediates the Wnt signaling pathway is trapped in a complex and programmed for ubiquitination and degradation. When the
pathway is activated (Figure 1.1 B) by Wnt binding to LRP5/6 and Frizzled, Dishevelled is activated which inhibits Glycogen Synthase Kinase-3β (GSK-3β). Axin moves out of the complex and can secondarily attach to LRP5/6. As Axin moves away from the complex the ability of GSK-3β to phosphorylate β-catenin decreases. β-catenin accumulates and is transported into the nucleus where it binds with the transcription factor T cell Factor 4 (TCF). This leads to the synthesis of chondrogenic and osteoblastic factors.

The canonical Wnt-signaling pathway is now well recognized to play a role in bone homeostasis[184]. Increased β-catenin is seen in committed osteoblast lineage cells. A loss of β-catenin in osteoblast precursor cells results in reduced bone deposition[187]. The relevance of Wnt has been noted in conditions like osteoporosis, RA, AS, osteoarthritis, and Paget’s disease [188-195]

It is puzzling that bone resorption is the phenotype in RA while bone formation is the hallmark of AS, even though the same cytokines appear to mediate both diseases. This riddle seemed to be solved with the identification of divergent DKK-1 levels in the two conditions[188]. DKK-1 is a circulating inhibitor of the Wnt-signaling pathway. It binds to the same receptors as Wnt and prevents the signalling. Thus in AS, where there is more bone formation, DKK-1 levels are low, resulting in excess osteoblastic factors forming. This has been replicated in Diffuse Idiopathic Skeletal Hyperostosis (DISH), another condition characterised by excess bone formation in the spine [196].
Figure 1.1 The Canonical Wnt-β Catenin Pathway

LRP: LDL-receptor Like Protein; GSK-3β: Glycogen Synthase Kinase-3β; β-Cat: β Catenin; Dsh: Dishevelled; TCF: Transcription Factor T cell Factor 4
Legend to Figure 1.1

The Canonical Wnt-β Catenin Pathway

(A) In the inactive state, β-catenin is trapped in a complex and programmed for ubiquitination and degradation.

(B) When the pathway is activated by Wnt binding to LRP5/6 and Frizzled, Dishevelled is activated which inhibits Glycogen Synthase Kinase-3β (GSK-3β). Axin moves out of the complex and can secondarily attach to LRP5/6. As Axin moves away from the complex the ability of GSK-3β to phosphorylate β-catenin decreases. β-catenin accumulates and is transported into the nucleus where it binds to the transcription factor T cell Factor 4 (TCF).
On the other hand, in RA, the DKK-1 level is higher with the balance tilting in favour of bone resorption. Even lupus patients with erosive arthritis have elevated DKK-1 levels[197]. DKK-1 has been shown to be low in another AS cohort [192] and when tested in a Jurkat T-cell model, was found to be dysfunctional in AS patients at activating the Wnt pathway [198]. Thus not only is DKK-1 lower in AS, it is also less effective in activating the Wnt signalling pathway.

The TNF-transgenic mouse has inflammation affecting the sacroiliac joints with no spontaneous fusion. TNF blockade reduced inflammation, erosions and osteoclasts in the sacroiliac joints [199]. On the other hand, DKK1 blockade reduced erosions and promoted ankylosis of the sacroiliac joints [199]. This further shows the relevance of the Wnt-signalling pathway in AS.

1.8 Endoplasmic Reticulum Aminopeptidase 1

Ever since it was realized that HLA-B27 was only one actor in a large dramatis personae, the hunt for other candidate genes started. It was recognized that AS is likely to be an oligogenic disease [144] and several attempts have been made to look for non-major histocompatibility complex (MHC) genes that might contribute to increased susceptibility to the disease [200]. A genome wide association study (GWAS) of nonsynonymous single nucleotide polymorphisms (SNPs) identified a novel association of AS with ERAP1 [137].

The name ERAP1 is approved by the Human Genome Organization (HUGO) Nomenclature Committee. ERAP1 is known by various aliases, such as endoplasmic reticulum aminopeptidase associated with antigen processing (ERAAP), adipocyte-
derived leucine aminopeptidase (A-LAP), puromycin-insensitive leucyl aminopeptidase (PILS-AP) and aminopeptidase regulating tumor necrosis factor receptor I (TNFRI) shedding (ARTS-1). ERAP1 consists of 930 amino acids and was identified in mice following solubilization and fractionation of microsomes derived from liver and spleen [201]. At approximately the same time, human ERAP1, which has 941 amino acids and 86% sequence similarity to murine ERAP1 (ERAAP), was isolated from HeLa S cells and found to be the same enzyme as A-LAP and PILS-AP [202]. The human ERAP1 gene spans 54.61 kb of chromosome 5 from 96149849 to 96095244, on the reverse strand. The ERAP1 gene lies between the ERAP2 and Calpastatin (CAST) genes (Figure 1.2).

The human ERAP1 mRNA has 2826 nucleotides (Figure 1.3) and the protein has 941 amino acids (Figure 1.4). The two SNPs that have been reported to be strongly associated with ERAP1 are rs30187 and rs27044 that lead to R528K and E730Q residue changes respectively (Figures 1.3, 1.4 and Table 1.4).

The discovery of the ERAP1 association has brought new insights and renewed excitement into the field of AS research. After HLA-B27, which has an attributable risk of 50%, ERAP1 has the second strongest association, with an attributable risk of 26% [137]. Thus after 35 years of research since discovering the B27-AS genetic association, for the first time a strong and replicated genetic association in AS has been identified. The known functions of ERAP1 have further triggered the expectations and imagination of the AS research community.

ERAP1 has two known functions. First, it aids the shedding of the membrane-bound cytokine receptors TNFRI, IL-1 receptor II (IL-1RII) and IL-6Rα. In fact ERAP1
was independently discovered as a molecule that bound cytokine receptors while searching for receptor shedding enzymes. This function of ERAP1 will be further described and the implications to AS pathogenesis discussed later.

The second known function of ERAP1 is the trimming of peptides within the ER for loading onto MHC-I molecules. ERAP1 has ERAP2, a structurally functionally related ER aminopeptidase, as a close ally in this function. Delineating the functional aspects of the AS-associated ERAP1 and ERAP2 polymorphisms has become a major focus of research. This novel connection could be the long sought bridge between HLA-B27 and AS.

1.8.1 Structure of ERAP1

ERAP1 and ERAP2 belong to the M1 family of zinc metallopeptidase enzymes, which share a HEXXH(X)18E zinc-binding motif [203]. Over the last 3 years there was a rigorous attempt to obtain a 3-dimensional model of ERAP1. Crystallographic structures can aid in predicting the possible functional effects of genetic polymorphisms on molecules.

The first attempt was to develop a molecular homology model of ERAP1 based on the crystallographic structure of a closely related protein, tricorn-interacting factor F3 (TIFF3), with which ERAP1 shares 32% amino acid sequence identity (residues 63–665) [204]. The sequence similarity increases to 82% when considering the active catalytic site (residues 280–486) of ERAP1, in particular the zinc-binding domain. Based on the crystallographic structure of TIFF3 sequence alignment was performed and a homology model of ERAP1 was generated.
Figure 1.2 The Chromosome 5q15 locus with ERAP1 and other surrounding genes
Legend to Figure 1.2

Chromosome 5 with *ERAP1*, *ERAP2*, *LNPEP* and *CAST* genes

The human *ERAP1* gene spans 54.61 kb of chromosome 5 from from 96149849 to 96095244, on the reverse strand (note direction of arrow). The *ERAP1* gene lies between the *ERAP2* and *Calpastatin (CAST)* genes.
Nucleotides in Blue are exonic. The nucleotide change associated with \textit{rs30187} and \textit{rs27044} SNPs are highlighted in yellow and green respectively. Thus the two major polymorphisms associated with ankylosing spondylitis are:

\begin{center}
\textbf{Figure 1.3 Nucleotide Change in the ERAP1 Polymorphisms Studied}
\end{center}

\textbf{ERAP1 mRNA Nucleotide Sequence (2826 nt):}

```
ATGGTGTTTCTGCCCCTCAAAATGGTCCCTGGAACCATGTCATTTCCTACTTTCTCACGTTGGCTCTCT
TAACGTGTCAGCCAATCTCTTTCTATGGTGCTAGACACGTCTAACATCTCACAACACCAAGCTGATGGAC
TCCTGGAAATAAAAATACGCTTCTGAGATGCTACCTCCAGTCTATTATCTTCTTGATCTGGAAGAAAC
CTTACACGCTGACTCTTTTCGGAGAACAGAAGAATAGAAATACACGGACTGACCACACAGTGGAAGG
TCCTGCTAGATCGACCTGCGATATCTAGGGGCACCCCTCAGGAAAGGAGCTGAGAGGCTATCGGA
AGAAACCCCTGCAAGGGCCGGACAGGAAATATGACACTGCTGGCTCCAGGCCACCTGCTT
GTCCGGCTCCCGTACACAGTTGTTCATTGACTATGCGGCAATTTCTTTCGGAGACTTTTACGA
AAAGCACTACAAGAACAGGAAGGGAACTGAGGATACTAGCTCAACACAAATCTGGAACCCACTGCAC
TAGAAATGGCCTTTCTGATGAAACCTGCTTCTAAAGCAAGGTTTCTCTCAAAATAATGAGAGAG
CACAAGGCACCTACATCCATTTCAATGCGATTGTAATCTGCTAGTTGCAAGATGACCTGATAGAG
ACCATTGGATGTACGTAGAGAAGCTACTATCTGCTGAGGCTTCACTATTGAGATTTTGGCTGTG
CAGCAAGATACAAAGAGTCAGCTCAAGTTTCTGTATGCTGTCAGACAAAGATTAATGAAAGAT
TACTACGTAGGATGCGCCAGTACTTCTTACTAGATATTTATGGAGATTATTTTACAGGATACGG
TACCCCTGTTAATCTGCTACCTAGCTGGGAGATTTCGTTATTTATGCTGACAGGAACTTACGG
ACACCCTGTTAAAATCCTGGCTAGATCCGGGAGATTTTGGATATTTATGCTGACAGGAACTTAC
GAAAAGGGAGGAATGACATCAGAGCAAGAAGCCTACAGTGAGAAGGCTGACGGGAAGCTTGAG
GGTGAGATGGAAGAACATGATGAAACCTGGCAACTGCAAGACGGTTGTTTCTCTAAATGAGAG
GAAATAGCTGAAAGGATATCTTATGCTGACAGGAAATGAGAGGGCTCTGCTGAAAGGAACTTAC
GTCTGGCTGCTTCTGAGGAAATCTTCTCTCTTCTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
```

Nucleotides in Blue are exonic. The nucleotide change associated with \textit{rs30187} and \textit{rs27044} SNPs are highlighted in yellow and green respectively. Thus the two major polymorphisms associated with ankylosing spondylitis are associated with ankylosing spondylitis are.
The residue changes resulting from the rs30187 and rs27044 polymorphisms are highlighted in yellow and green respectively. Thus the two major polymorphisms of ERAP1 associated with AS lead to change in protein structure.

<table>
<thead>
<tr>
<th>SNP</th>
<th>mRNA Position</th>
<th>Allele Change</th>
<th>Protein Position</th>
<th>Residue Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs30187</td>
<td>1930</td>
<td>AAG - AGG</td>
<td>528</td>
<td>K (Lys) – R (Arg)</td>
</tr>
<tr>
<td>rs27044</td>
<td>2535</td>
<td>CAA - CAA</td>
<td>730</td>
<td>Q (Gln) – E (Glu)</td>
</tr>
</tbody>
</table>
The model of ERAP1 contains a large cavity that could accommodate 9-mer peptides. Based on this structure, none of the SNPs associated with AS are near the catalytic site of the enzyme.

There is still an explanation as to how these SNPs can affect the function of ERAP1. ERAP1 has a preference for leucine at the N-terminal end and in fact was characterized first as a Leucine Aminopeptidase [205]. Other hydrophobic residues like methionine, phenylalanine and alanine, when present at the N-terminal end still results in reasonably fast peptide trimming. The enzymatic activity decreases when hydrophilic residues like arginine, lysine, glutamic acid or threonine are present [204]. However, this change in enzymatic activity has been shown with variations in in the C-terminal end with a preference for hydrophobic residues [206]. It is now known that changes in the internal sequence of substrate peptides—and not just in the N- or C-terminal end—affect the enzymatic activity of ERAP1 [204]. Thus for internal sequences to affect the enzymatic activity, a close interaction between the entire peptide substrate and ERAP1 has to be considered. This close association is required for efficient enzymatic activity. Polymorphisms at sites remote from the catalytic sites, and closer to the substrate binding areas, thus can affect this close association.

Two groups have now independently solved the crystal structure of ERAP1. Varying conformations have been identified and clearly ERAP1 appears to take an ‘open’ (Figure 1.5) and ‘closed’ (Figure 1.6) conformation [207]. The final crystal structure is quite similar to the original computed model with 4 domains. The N-terminal end is formed by the first domain extending up to residue 254. This domain is
Figure 1.5 ERAP1 Crystal Structure: Open Configuration
The N-terminal end is formed by the saddle shaped first domain extending up to residue 254, which forms the binding site for the N-terminal end of the peptide. Domain II (residues 255–527), like all aminopeptidases, houses the catalytic domain, the GAMEN and the HEXXHX$_{18}$E motifs. Domain II is connected to the C-terminal Domain IV by a bridging domain III (resides 528-613). In the closed state of ERAP1, domain IV arches over domain II forming a large central cavity enclosing the catalytic site. The zinc ion in the catalytic domain as well as the positions of the ERAP1 SNPs rs30187 (K528R) and rs27044 (Q730E) are depicted here. The surface carbohydrate moiety is composed of Mannose and N-Acetyl Glucosamine.

Quite clearly, from the open structure, these polymorphisms are away from the catalytic site. The figure of the crystal structure of ERAP1 [207] was generated using the Jmol software [208].
Figure 1.6 ERAP1 Crystal Structure: Closed Configuration
Legend to Figure 1.6

The closed conformation of ERAP1

In the closed conformation of ERAP1, domain IV closes the catalytic site and forms a large cavity that accommodates large peptides. It is evident that the K528R variation affects the outer surface of domain II and is not near the catalytic site. The K528R variation however could affect the conformational change from open to closed state by altering the interaction with surrounding residues.

The Q730E variation cannot be seen here as it is located within the cavity in the closed state. Thus there is a potential for interaction with peptides in the closed state. The figure of the crystal structure of ERAP1 [207] was generated using the Jmol software [208].
saddle shaped and consists of eight β-sheets and forms a binding site for the N-terminal end of the peptide. Domain II (residues 255–527), like all aminopeptidases, houses the catalytic domain, the GAMEN and the HEXXHX$_{18}$E motifs [207,209].

In contrast to some other aminopeptidases, domain II is connected to the C-terminal Domain IV by a bridging domain III (resides 528-613). As domain IV forms a cover over the catalytic site on domain II, the presence of domain III increases the size of the cavity. This could be the reason why ERAP1 can accommodate longer peptides for processing. Domain IV (residues 614–941) is made of 16 alpha-helices arranged like a cup. In the closed state of ERAP1, domain IV arches over domain II forming a large central cavity enclosing the catalytic site [207,209].

1.8.2 Genetic association of \textit{ERAP1} with AS

The ERAP1-AS association is well established now and has been replicated in multiple cohorts and ethnicities. The study by the Wellcome Trust Case Control Consortium and the Australo-Anglo-American Spondylitis Consortium (WTCCC-TASC study) was the first to establish this link. Single Nucleotide Polymorphisms (SNPs) are changes in the DNA sequence by one nucleotide and the lowest allele frequency at that particular locus in the population is called minor allele frequency (MAF). When the frequency of this allele is less than 1% it is called a mutation and not as SNP. By looking for overexpression of a particular allele, usually by comparing the MAF, in the disease cohort vs. controls, possibly pathogenic DNA variations can be identified. The WTCCC-TASC group used 14,500 nonsynonymous SNPs and found that, in British Caucasian patients with AS, the minor allele frequency (MAF) of the \textit{ERAP1} SNPs
rs30187 and rs27044 was considerably higher than in controls [137]. The study also showed the MAF of the SNPs rs10050860, rs17482078 and rs2287987 was lower than in controls [137].

In a multicenter study in Canada, which included 992 cases of AS and 1,437 controls, all of Northern European Caucasian ancestry, five genes involved in the antigen-processing pathway were assessed [210]. ERAP1 was found to be significantly associated with AS and the haplotype rs27044/10050860/30187-CCT was strongly associated with an increased risk of AS (odds ratio [OR] 1.8, 95% confidence interval [CI] 1.5–2.2; \( P=7\times10^{-8} \)). The ERAP1 haplotype rs30187/26618/26653-CTG was protective (OR 0.8, 95% CI 0.7–0.9; \( P=9\times10^{-5} \)). The association between ERAP1 and AS has been replicated in Caucasian and non-Caucasian populations, including Han Chinese, Korean and Portuguese individuals [211-214].

In family studies there are no confounding issues of population stratification unlike larger population studies. We studied multiplex AS families in the North American Spondylitis Consortium (NASC) and found a novel haplotype of ERAP1/ERAP2 to be associated with AS [215]. This study was the first to report an association of ERAP2 with AS. This is a particularly interesting finding with implications to the pathogenesis of AS. Unlike ERAP1, the only known function of ERAP2 is peptide trimming for MHC class I molecules. ERAP1 and ERAP2 act in concert for efficient antigen presentation. Hence an association with a haplotype of ERAP1 and ERAP2 would indicate that antigen presentation might be the key abnormality in the pathogenesis of AS. This is corroborated by the recent finding that ERAP1 is associated
with AS only in HLA-B27 positive patients [128]. This was not the case for the other well known AS associated genes like *IL-23R* and *IL-12B*. Similarly ERAP1 is associated with psoriasis only in patients who are HLA-C positive [216]. Thus a strong genetic interaction between *ERAP1* and *HLA-B27* could implicate functional interaction as well considering the known functions of the two genes and the fact that the genes are not in linkage.

### 1.8.3 Normal functions of ERAP1

During antigen processing, peptides generated by the proteasome are transported into the ER with the help of the transporter associated with antigen processing (TAP) protein. TAP can transport peptides up to 16 amino acids in length. Hence these peptides have to be trimmed before presentation on MHC-I. Moreover, peptides with proline in any of the first three positions of the peptide are not efficiently transported by TAP. Analysis of MHC–peptide complexes has determined that proline is commonly encountered in these positions [217,218]. This result suggested that peptides are transported with N-terminal extensions into the ER and then trimmed to appropriate lengths before MHC class I uploading. Following a search for such ER resident enzymes, ERAP1 was discovered [202]. Thus, N-terminally extended peptides generated by the proteasome enter the ER, assisted by TAP. These peptides are, in turn, cleaved by ERAP1 and prepared for loading onto MHC class I molecules (*Figure 1.7*).

ERAP1 is peculiar among aminopeptidases in its ability to cleave long substrates [206,219]. Although ERAP1 prefers 8-10 amino acid long peptides, it can
Figure 1.7 Schematic Diagram to show the role of endoplasmic reticulum aminopeptidases in antigen processing and presentation
Legend to Figure 1.7

Schematic Diagram to show the role of endoplasmic reticulum aminopeptidases in antigen processing and presentation

Proteins targeted for degradation are processed by the proteasome. The resulting peptides are N-terminally extended, relative to the ideal length of MHC-I binding peptides. They are transported into the Endoplasmic Reticulum (ER) by Transporter Associated with Antigen Processing (TAP). These N-terminally extended peptides have to be trimmed by ER-resident peptidases before they can be finally presented on MHC-I.

Abnormal ERAP1 and ERAP2 activity can lead to the generation of abnormal peptides. This can lead to unstable peptide-MHC complexes that can misfold. These can accumulate in the ER or be presented as Free Heavy Chains (FHC) on the surface. Abnormal peptide MHC complexes can also be presented on the surface with resulting altered immune responses.
accommodate peptides with up to 16 amino acids [206,219]. A molecular ruler mechanism was proposed initially to explain the length specificity of ERAP1 [206]. Not only the N-terminal residues of the peptide that bind to the catalytic site of ERAP1, but also the C-terminal residues and other residues along the length of the peptide can affect the enzymatic activity of ERAP1 [204,206]. Peptides with both N- and C-terminal ends bound respectively to the catalytic site and hydrophobic pockets are better processed. There are some clues in the new crystal structure of ERAP1. ERAP1 assumes the closed conformation after peptide binding. ERAP1 has maximal activity in this conformation. Thus a properly fitting, ideal peptide could promote the closed conformation of ERAP1 and thereby improve enzymatic activity. Peptides shorter than eight amino acids do not extend from the catalytic site up to the hydrophobic pocket of ERAP1 (Figure 1.8) and could cause ERAP1 to remain in an open state.

Proline in the first 3 positions and hydrophilic C-terminal residues make the peptides less favored (Figure 1.8). The inability to process peptides with N-terminal proline is thought to be due to the absence of a free N-terminal group to be recognized by the GAMEN motif of ERAP1 [220]. The catalytic site and the groove extending from there are hydrophobic [209]. This could be the reason why very short hydrophilic side chains are not preferred [221]. Valine, isoleucine and threonine with β-branched side chains are poorly tolerated [221].

The subcellular localization of ERAP1 and ERAP2 has been a matter of debate. In mice, ERAP1 has been shown to be present in the ER only. By contrast, the
Figure 1.8 ERAP1 Peptide specificity

- Very Long Peptide
- Short Peptide
- Proline in N-Terminus
- Hydrophyllic C-terminus
- Ideal Peptide
- ERAP1 Peptide Binding Groove
Legend to Figure 1.8

Peptide binding Specificity of ERAP1

ERAP1 prefers residues between 8-10 residues but can accommodate peptides up to 16 residues due to the large size of its peptide binding groove. Longer peptides cannot be accommodated and are poor substrates for ERAP1. Peptides shorter than eight amino acids do not extend from the catalytic site up to the hydrophobic pocket of ERAP1. Peptides with both N- and C-terminal ends bound respectively to the catalytic site and hydrophobic pockets are better processed. Proline in the first 3 positions and hydrophilic C-terminal residues make the peptides less favored.
presence of ERAP1 has been demonstrated on the cell surface as well as the ER in humans. Interestingly, in human cell lines, both ERAP1 and ERAP2 are secreted when overexpressed [203]. The postulated explanation for this phenomenon is that ERAP1 and ERAP2 are normally retained in the ER by certain binding proteins but can escape when these retention proteins become saturated. It should, however, be borne in mind that evidence of ERAP1 and ERAP2 secretion exists only in cell lines, and the current understanding is that these aminopeptidases in human tissue are predominantly in the ER but might be present on the membrane as well. ERAP1 and ERAP2 polymorphisms associated with AS have been shown to affect the levels of expression of these genes [222]. Whether any subsequent change in protein levels results in secreted aminopeptidase activity needs to be resolved. Currently, the possible role of ERAP1 and ERAP2 in the pathogenesis of AS remains speculative, based on well-established functions.

1.8.3.1 ERAP1 in peptide presentation

In HeLa cells that stably expressed an MHC class I molecule (H-2K\textsuperscript{b}) and transfected with an H-2K\textsuperscript{b} binding epitope (MSIINFEKL), ERAP1 was suppressed by small interfering RNA (siRNA). This led to a reduction in the processing of the MSIINFEKL epitope, and the levels of peptide–MHC complexes decreased to 20% of those of controls [219]. Cells from ERAP1-deficient mice show decreased levels of MHC class I peptide expression on their surface [223]. Interestingly, a novel set of MHC–peptide complexes were seen in the ERAP1-deficient mice, which were not present in the wild-type mice. This observation indicates that some peptides that normally get
eliminated can be presented on MHC-I when ERAP1 is lacking. Thus, a change in ERAP1 levels caused a total change in the peptide–MHC repertoire [217]. A follow-up study showed that ERAP1 can also remove peptides that are not avidly bound to MHC class I molecules in the ER, adding to the potential mechanisms by which ERAPs might be associated with AS [224].

The loss of ERAP1 decreased the trimming of the aminopeptidase substrate leucine-7-amino-4-methylcoumarin (L-Amc) but not Arginine-Amc (R-Amc), which indicated the presence of other aminopeptidase enzymes within the ER with different specificities.[202] Following a search for this aminopeptidase activity, ERAP2 was cloned from a human leukocyte cDNA library (and initially known as leukocyte-derived arginine aminopeptidase [L-RAP]) [203]. As mentioned previously, ERAP2 has not been demonstrated in mice or rats and ERAP2 has been identified as a pseudogene by genome sequencing [225]. The expression of ERAP1 and ERAP2 is upregulated by interferon gamma (IFN-γ) in human cells; the proteins co-localize in the ER and can form heterodimers [226]. ERAP1 and ERAP2 have unique peptide specificities, which results in efficient and coordinated presentation of a wide variety of peptides when these aminopeptidases act in concert. Apart from the study of multiplex AS families, no other reports so far link ERAP2 to AS. Tsui et al. reported that polymorphisms in both ERAP1 and ERAP2 could be involved in the pathogenesis of AS by altering peptide processing [227].

ERAP2 was not found to be linked to AS in the initial GWAS, and there was strong Linkage Disequilibrium (LD) between ERAP1 and ERAP2 [137]. ERAP2 is now
known to be associated with Crohn’s disease which has numerous similarities in arthritis manifestations and pathogenesis to AS [228]. Individuals with the rs2248374 polymorphism that affects the 5’ canonical splice site of exon 10 of ERAP2 have altered mRNA splicing, a truncated protein and nonsense mediated decay [229]. Almost 25% of the general population have this SNP and thus lack ERAP2 naturally [229]. The lack of ERAP2 is associated with reduced MHC-I expression as assessed by the W6/32 antibody [229]. Thus a closer look at this phenomenon was warranted in AS. In a study of 470 AS patients and 420 ethnically matched normal controls, there was no association between the ERAP2 rs2248374 SNP and AS [230].

Abnormal peptide processing could lead to unstable peptide–MHC complexes that are prone to misfold. HLA B27 molecules that are not properly folded tend to accumulate in the ER (see previous Figure 1.7) and cause ER stress leading to UPR, a pro-inflammatory cellular response [135,160,161]. Alternatively, or in addition to, the presentation on MHC class I molecules of abnormally-generated peptides as a result of ERAP polymorphisms could lead to an abnormal immune response (see previous Figure 1.7). It has been shown that cells from ERAP-deficient mice are immunogenic in wild-type mice,[223] which could mean that ERAP variants can contribute to the pathogenesis of AS by conferring anomalous peptide presentation and T-cell recognition. Peptide-MHC instability leading to FHC formation can also cause aberrant immune responses by binding to nontraditional immunoreceptors such as killer cell immunoglobulin-like receptors (KIR) and leukocyte immunoglobulin-like receptors (LIR).[231-233] The KIR receptors are found on lymphocytes and T cells while LIR are seen predominantly in myelomonocytic lineage.[164] HLA B27 homodimers are
recognized by the KIR receptors KIR3DL1 and KIR3DL2 as well as by the LIR receptor LIRB2.[163,167,234] It needs to be seen if the FHC interaction with these immunoreceptors can play a role in the pathogenesis of AS.

1.8.3.2 ERAP1 and cytokine receptor shedding

ERAP1 has been reported to be involved in the cleavage of IL-1RII, IL-6Rα and TNFRI [235-237]. In addition to the discovery of ERAP1 following a search for ER-resident aminopeptidases, ERAP1 was independently identified through a yeast two-hybrid approach while searching for molecules that bind to the extracellular portion of TNFRI [237]. ERAP1 was also shown on bronchial airway cells by confocal microscopy to be a type II transmembrane protein that co-localizes with TNFRI [237].

In the first study examining a link between ERAP1 and TNFRI shedding, an inverse correlation between the surface expression of TNFRI and ERAP1 was shown by immunoblots, indicating low levels of TNFRI on the surface of cells with high ERAP1 expression [237]. In the same study, culture supernatants of cells with high expression of ERAP1 showed increased soluble TNFR-I level [237]. By site-directed mutagenesis, it was also shown that ERAP1 does not function as the actual cleaving enzyme but, rather, aids other metalloproteinases in TNFRI shedding [237]. It has subsequently been demonstrated that ERAP1 must be present in a complex with nucleobindin-2 and a 43-kDa heterogeneous nuclear ribonucleoprotein (hnRNP) G coded by the gene RBMX (RNA-binding motif gene, X chromosome), for TNFRI shedding [238,239]. By co-immunoprecipitation the association between ARTS-1 and RBMX was identified and attenuating RBMX expression by RNA interference reduced the constitutive release of
TNFR1 exosome-like vesicles and IL-1beta-mediated cleavage of soluble TNFR1 ectodomains. Over-expression of RBMX increased TNFR1 exosome-like vesicle release [239]. As is known for hnRNPs, hnRNP G could shuttle between the cytoplasm and nucleus and bind to the cytoplasmic portion of the TNFRI-ERAP1 complex. ERAP1 has also been shown to aid the cleavage of IL-1RII and IL-6Rα [235,236].

On the basis of this known function, I hypothesize that abnormal cytokine receptor shedding (as a result of ERAP polymorphisms) could lead to decreased levels of circulating cytokine receptors, which, in turn, might result in higher levels of cytokine biological activity. Interestingly, the corresponding cytokines that bind to the three receptors known to be affected by ERAP1 are directly or indirectly linked to the pathogenesis of AS. The IL-1 gene complex has been found to be associated with AS in a number of studies, and IL-6 is a cardinal cytokine involved in the generation of T helper 17 (Th17), cells which have been found to be important in the pathogenesis of AS [143,240].

1.8.4 ERAP1 polymorphisms and AS phenotype

There are no published reports on the effect of ERAP1 polymorphisms on the clinical phenotype of AS. Clinical heterogeneity in AS is perplexing, as most patients have the HLA-B27 gene and HLA-B27 cannot, therefore, be a determinant in variable disease expression. Hence, other modifying factors have been considered, and ERAP polymorphisms might be an important contributor in this regard.
1.8.5 ERAP1 and HLA-B27 pathogenesis

The hypothesis for HLA-B27 conferring susceptibility to AS has been outlined above. It would be very interesting to see whether ERAP1 could contribute to any of these mechanisms. As seen from the evidence so far, variations in ERAP1 can affect the peptide trimming efficacy and support an altered peptide hypothesis. MHC-I FHC might be formed or a UPR response evoked as a result of abnormal peptide presentation. These FHC can exist as monomers or dimers and cause abnormal immune interaction with KIR and LIR receptors.[134,163,233,241] Finally, abnormally processed microbial peptides can result in molecular mimicry. Thus, almost all of the aforementioned hypotheses could be explained by the ERAP1 association with AS.

1.8.6 ERAP1 in other diseases

Often the initial suggestions for a functional role of a novel gene come from other conditions with which it may be associated. Both ERAP1 and ERAP2 are associated with other diseases and thereby provide a source of clues to their biological functions. The ERAP1 association in HLA-C positive patients with psoriasis has been mentioned above [216]. Interestingly in Crohn’s disease, which has several common genetic links, there was an association not with *ERAP1* but with *ERAP2* [228]. The reported association was strongest with rs27524, an intronic SNP on ERAP1. These authors did note that the *ERAPI* rs30187 SNP also was significantly associated with *HLA-C* positive psoriasis.

The ERAP1 K528R variant has been reported to be associated with systemic arterial hypertension, diabetes mellitus and hemolytic uremic syndrome [242-244].
Investigating the functional role of \textit{ERAP1} in hypertension, the K528R variant of ERAP1 has been shown to be less active than the ancestral wild-type variant, which leads to less bradykinin formation or lower inactivation of angiotensin II, or both [243]. ERAP1 is an enzyme that cleaves and inactivates angiotensin II[205]. ERAP1 also converts kallidin to bradykinin[205]. Thus an efficiently functioning ERAP1 is essential for decreasing angiotensin II (a vasopressor) and increasing bradykinin (a vasodilator). The K528R variant would thus lead to hypertension by less vasodilator and more vasoconstrictor level in circulation.

By site directed mutagenesis, ERAP1 variant cDNAs were synthesized and expressed in insect Sf9 cells. ERAP1 was extracted from the supernatant of Sf9 cell cultures and used for testing enzymatic activity. Using a fluorogenic substrate, L-leucyl-\(\beta\)-naphtylamide, it was shown that replacing Leu at position 528 with any other amino acid led to decreased enzymatic activity. The K528R variant in particular had the lowest enzymatic activity. This was the first report to show an effect of ERAP1 polymorphisms on the catalytic function of ERAP1.

A similar hypothesis of decreased enzymatic activity has been proposed to explain the ERAP1 association with hemolytic uremic syndrome. Abnormal cleavage can lead to excess of the Shiga toxin, which is responsible for this disease [242].

ERAP1 and ERAP2 have been associated with pre-eclampsia.[245] Interestingly, an exaggerated inflammatory response to the fetal allograft is a proposed mechanism for pre-eclampsia [246]. HLA-G expression on the placenta is important for maintaining the immune-privileged status of the fetus. Inhibitory KIR receptors bind to HLA-G
molecules, thereby suppressing the activity of natural killer cells [247,248]. The expression of HLA-G can be decreased by blocking ERAP1 expression using inhibitory RNA, but whether the association seen with ERAP1 and ERAP2 in pregnancy relates to changes in HLA-G expression has not been resolved [249].

Variations in the antigen-presenting function of ERAP1 have been correlated with several infectious and malignant diseases. An impaired ability to present pathogen-derived peptides could lead to inadequate immune responses and increased pathogenicity. ERAP1-deficient mice are not able to process the immunodominant decapeptide HF10 of *Toxoplasma gondii* and die from overwhelming infection when challenged with this pathogen [250]. An Ala146Pro mutation in the Gag protein of HIV inhibited its ability to be cleaved by ERAP1 for MHC presentation, thereby resulting in a decreased cytotoxic T cell response. This is corollary proof of the importance of ERAP1 specificity and immune response. Similarly, increased cervical cancer metastasis and decreased survival have been reported to be associated with polymorphisms in *ERAPI* [252-254]. Variations in the levels of ERAP1 and ERAP2 along with variations in the expression of surface MHC have been reported in various non-lymphoid cancer cell lines [255,256]. *ERAP* polymorphisms could reduce the MHC expression in tumor cell lines and result in a decrease in immune surveillance.

The ERAP1 K528R polymorphic variant is associated with pre-menopausal osteoporosis in Japanese women. As this variant is the same as that found to be associated with decreased radiographic progression of AS [258], it seems that ERAP1
activity is important in the catabolic–anabolic balance of bone metabolism. Further studies in this regard are warranted.

1.9 Aim

The broad aim of the study was to identify the functional implications of the newly identified ERAP1-AS association. Specifically I had the following aims:

1. Does ERAP1 polymorphism lead to changes in the surface expression of HLA-B27 in AS patients? Are there differences in FHC expression?

2. Can in vitro suppression of ERAP1 lead to changes in HLA-B27, surface FHC and intracellular FHC expression? Can ERAP1 suppression lead to an increase in HLA-B27 presenting abnormal and long peptides?

3. Are there differences in the interaction of ERAP1 with the different HLA-B27 subtypes?

4. Can ERAP1 polymorphism affect the serum cytokine receptor level in AS patients?

5. In addition to being a susceptibility marker, is ERAP1 also a marker of severity of disease?
1.10 Hypothesis

ERAP1 variants have differential peptide trimming capabilities. This leads to altered HLA-B27 peptide loading and an abnormal immune response. Abnormal peptide-B27 complexes can be unstable and lead to increased FHC formation that can accumulate in the ER or be presented on the surface either as monomers or FHC dimers. This functional interaction of ERAP1 with HLA-B27 may depend on the subtype and can lead to pathogenic changes only in the AS-associated HLA-B2704 and B*2705 subtypes.

AS-associated ERAP1 variants can have abnormal clipping capability, resulting in lower levels of circulating cytokine receptors in circulation. This could result in an imbalance in the receptor-cytokine ratio with more cytokine causing inflammation.

As ERAP1 is a strong susceptibility marker it could impact the radiographic severity of AS and the rate of radiographic progression.
Chapter 2

This chapter was published in the Annals of Rheumatic Disease: Haroon N, Tsui FWL, Uchanska-Ziegler B, Ziegler A and Inman RD. Endoplasmic Reticulum Aminopeptidase 1 (ERAP1) has functionally significant interaction with HLA B27 and dictates subtype specificity in ankylosing spondylitis. Ann Rheum Dis 2011; In Press).

FWLT aided in genotyping patients in the first part of the study. BUZ and AZ provided the MARB4 antibody. RDI and FWT aided in planning the experiments. The draft was approved by all co-authors. I recruited patients for the study, collected samples, separated PBMC and did flow cytometry. I did tissue culture of C1R cells, siRNA experiments, western blot and flow cytometry. I performed all statistical analysis and prepared the draft of the manuscript.

2 Functional interaction of ERAP1 and HLA-B27 is subtype specific and could be pathogenic in ankylosing spondylitis

2.1 Introduction

The pathogenesis of AS is not well understood. The significance of genetics is evident from twin and family studies showing an increased risk of recurrence of AS in first degree relatives and twins[103,104,144,145]. The first genetic risk factor identified to be strongly associated with AS was HLA-B27 [146]. Despite thirty-five years of research we still do not know how exactly HLA-B27 contributes to the pathogenesis of AS. As HLA-B27 is a classical MHC-I molecule, an abnormality of antigen presentation has been suspected for some time. Despite arduous efforts to identify ‘arthritogenic peptides’ in AS, they have not met with great success [259].

Recurrence risk modeling studies have hinted at the possibility of AS being an oligogenic disease with contributions from at least 6 important genes[144]. The recent
discovery of *ERAP1* as an important genetic risk factor in AS invigorates the hope of resolving the pathogenesis of AS. ERAP1 is an ER resident peptide cleaving enzyme that can functionally interact with HLA-B27 by affecting antigen presentation [260]. Thus variations in ERAP1 can lead to changes in the peptide-B27 repertoire with effects on immune response. Altered peptide-B27 complexes can have reduced stability and lead to the formation of MHC-I free heavy chains (FHC). These FHC can be retained in the ER leading to an ER stress response, or presented on the surface resulting in abnormal immune interactions[158,164,167,233].

Antigens are processed and loaded on MHC-I molecules in the ER and transported to the surface for presentation. Thus the surface expression of MHC-I is a surrogate marker for adequacy of peptide processing. A screening approach to assess the effects of ERAP1 variants on antigen presentation would be to study the level of MHC-I expression on PBMC in AS patients. Considering the potential effects on the pathogenesis of AS, it is important to look for the level of HLA-B27 expression as well as MHC-I FHC.

The *ERAP1* SNPs *rs30187* and *rs27044* lead to a K528R and Q730E change respectively in the ERAP1 protein. These are the most widely replicated and strongest *ERAP1* SNPs associated with AS. It is now well established that ERAP1 variants have altered aminopeptidase activity [207,261]. The *rs27044* SNP has less enzyme activity than the ancestral / wild type ERAP1 variant depending on the presence of other substrates in the reaction [261].
As a first step, I compared the surface expression of HLA-B27 and FHC on the peripheral blood mononuclear cells (PBMC) of AS patients with different genotypes of these ERAP1 SNPs. Then I addressed the question whether reducing the level of ERAP1 would have similar results *in vitro*. I developed an *in vitro* system with C1R cells and used siRNA for ERAP1 suppression. In addition to FHC and HLA-B27 expression, I studied the effect of ERAP1 suppression on the surface expression of HLA-B27 that reacts with the mAb MARB4 and intracellular FHC (IC-FHC).

Taking this one step forward, I also addressed if the effects seen with ERAP1 suppression on C1R cells differ with the subtype of HLA-B27 that the cells express. I did this by using different C1R cell lines that have been stably transfected with the HLA-B27 subtypes: B*2704, B*2705, B*2706 and B*2709. As detailed in the introductory chapter, HLA-B*2704 and B*2705 are associated with AS while B*2706 and B*2709 are not. Hence if the result of the ERAP1 suppression is different in the AS-associated HLA-B27 subtypes compared to the AS-neutral subtypes, it would further strengthen the hypothesis that ERAP1-B27 interaction is pathogenic in AS.
2.2 Methods

2.2.1 Patients

We enrolled Caucasian AS Patients attending the Spondylitis Clinic at Toronto Western Hospital in the study if they met the following criteria

(a) Satisfied the modified New York criteria for the diagnosis of AS

(b) Consented to participate in the study

(c) Were HLA B27-positive

(d) Were not on any anti-TNF agents at the time of study

All patients underwent a comprehensive clinical examination. Patients were assessed for disease activity by the Bath AS Disease Activity Index (BASDAI), Erythrocyte Sedimentation Rate (ESR) and C Reactive Protein (CRP). Functional capacity was evaluated using Bath AS Functional Index (BASFI). Peripheral blood was collected at the same time as clinical assessment for flow assisted cell sorting (FACS).

2.2.2 Genotyping

DNA was prepared by Basil Chu in the Inman Lab from peripheral blood using a standard DNA isolation kit (Gentra Systems, Minneapolis). The principle of cDNA isolation from peripheral blood involves lysing RBCs which are non-nucleated and lacking genomic DNA first, followed by lysing WBCs in the presence of a DNA stabilizer to protect the DNA from the action of DNAase. Briefly, RBCs and WBCs were lysed with
a proprietary cell lysis solution from Gentra systems containing an anionic detergent. Salt precipitation was used to precipitate protein and alcohol precipitation was used to precipitate DNA from the supernatant. DNA was mixed with 100% isopropanol. The precipitated DNA was subsequently washed with 70% ethanol and hydrated with a DNA hydration solution consisting of 1 mM EDTA and 10 mM Tris.Cl at pH 7.5. DNA concentrations were measured and samples aliquoted.

Genotyping was performed by Hing Wo Tsui in the Tsui Lab at the Toronto Western Hospital [215]. Two SNPs in the ERAP1 gene known to be associated with AS (rs27044 and rs30187) were typed by the allelic discrimination assay [262]. The rs27044 SNP is at exon 16 of ERAP1 and leads to a C/G substitution at position 96118852 on chromosome 5. The rs30187 SNP is on exon 12 with a C/T substitution at position 96124330 of chromosome 5. Optimized allelic discrimination assays for SNPs were purchased from Applied Biosystems (Foster City, CA, USA). The Applied Biosystems 7900HT fast system was used for the PCR.

Briefly, allelic discrimination assays are based on the ability to directly detect the polymerase chain reaction (PCR) product by measuring the increase in fluorescence of dye-labeled DNA probes. Two probes are used, one for each allele. Each probe consists of an oligonucleotide with a 5´-reporter dye and a 3´-quencher dye. The reporter dye is different in the two alleles while the quencher is common. Once the PCR reaction progresses in a 5´-3´ direction, the 5´ reporter dye gets released from the quenching effects of the quencher on the 3´ end of the probe leading to a signal that can be detected by the PCR machine. In our system TET (6-carboxy-4,7,2´,7´-
tetrachlorofluorescein) is covalently linked to the 5´ end of the probe for Allele 1. FAM (6-carboxyfluorescein) is covalently linked to the 5´ end of the probe for Allele 2. TAMRA (6-carboxy-N,N,N´,N´-tetramethylrhodamine) is the common quencher. Thus if signals from TET and FAM are detected, the patient is heterozygous with both allele present. Single signals would indicate a homozygous state for the corresponding allele.

2.2.3 Cell Lines

Hmy2.C1R (C1R) is a human lymphoid cell line with low expression of its endogenous HLA class I molecules [263]. The cells have been stably transfected with different HLA-B27 subtypes (B*27:04, B*27:05, B*27:06, and B*27:09) [264]. These cells (a kind gift of Dr. José López de Castro, Madrid) were cultured in IMDM with 10% Fetal Bovine Serum (FBS) to 70 to 90% confluence. The culture medium was changed once every three days or earlier if required. The nucleofection protocol for C1R cells was optimized using the cell line optimization nucleofector kit from AMAXA (Lonza, MD, USA).

2.2.4 Optimization of Nucleofection

The Nucleofector™ Technology is a significantly better technique than electroporation technology for gene delivery. The Amaxa Nucleofector™ Technology involves a combination of optimized electrical parameters and specific Nucleofector™ Solutions that are ideal for different types of cells. By combining the ideal nucleofector solution with the proper delivery pulse, delivery of the DNA of interest straight into the
nucleus can be achieved. As the DNA quickly reaches the nucleus, expression can begin without delay.

The C1R cells were cultured to confluence and the cell line optimization nucleofector kit (Amaxa) was used to identify the ideal combination of nucleofector solution and nucleofection program that gave the best ratio of efficacy to cell death. In brief, the cell line optimization kit has 2 nucleofection solutions L and V. A set of 7 programs were tried with both solutions bringing the total possible combinations to 14 (Table 2.1). The C1R cells were removed from culture and mixed with the nucleofector solution and green fluorescent protein (GFP) DNA in an Amaxa-certified cuvette. The number of cells that could be processed at one time ranged between $1 \times 10^6$ to $5 \times 10^6$ cells. The cells were mixed with 2 μg pmaxGFP and 100 μl Nucleofector Solution. pmaxGFP encodes the GFP from copepod *Pontellina p*. FACS was used to monitor the efficiency of transfection. Two control samples were performed for assessing the effect of DNA alone and nucleofection alone on cell survival. In the first control experiment, cells were mixed with Nucleofector Solution and pmaxGFP but no nucleofection program was used. In the second control experiment, cells were mixed with nucleofector solution, but without pmaxGFP and the nucleofection program was applied.

Following nucleofection, the C1R cells were suspended in pre-warmed supplemented IMDM media and incubated in a humidified 5% CO2 incubator at 37°C. After 24 hours, cells were removed from culture and re-suspended in FACS buffer with propidium iodide (PI). FACS was used to estimate the percentage of cells with GFP
expression and dead cells. Based on these experiments, the X-001 program with the best efficacy:toxicity ratio was selected for further experiments.

Table 2.1 The Amaxa cell line optimization nucleofector kit

<table>
<thead>
<tr>
<th>Program</th>
<th>Solution</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-020</td>
<td>L</td>
<td>V</td>
</tr>
<tr>
<td>T-020</td>
<td>L</td>
<td>V</td>
</tr>
<tr>
<td>T-030</td>
<td>L</td>
<td>V</td>
</tr>
<tr>
<td>X-001</td>
<td>L</td>
<td>V</td>
</tr>
<tr>
<td>X-005</td>
<td>L</td>
<td>V</td>
</tr>
<tr>
<td>L-029</td>
<td>L</td>
<td>V</td>
</tr>
<tr>
<td>D-023</td>
<td>L</td>
<td>V</td>
</tr>
</tbody>
</table>

There are 14 possible combinations of nucleofection solutions and programs. These are tried in various combinations to identify the ideal solution and program that gives the best efficacy:toxicity ratio.
2.2.5 ERAP1 Suppression by RNA Interference

The Stealth RNAi™ (Invitrogen, Carlsbad, CA, USA) technology was used for the siRNA experiments. Pools of three specific RNA duplexes for ERAP1 and a pool of non-targeting duplex as negative control (NC) were used. The NC siRNA (Invitrogen, USA) is a stealth control siRNA that does not induce any stress response and is not homologous to anything in the vertebrate transcriptome. The HSS numbers of the oligos are HSS122605, HSS122607 and HSS182221 and were delivered to the cells by nucleofection using Amaxa® Cell Line Nucleofector Kit® V (Lonza, MD, USA). ERAP1 suppression was measured by Western blot after 96 hours. Five separate experiments were done. Compared to the NC siRNA, ERAP1 siRNA (Figure 2.1) achieved a median protein suppression of 74% (range: 69-79%; n=5).

2.2.6 Western Blot Assay

Western Blot (WB) was done to determine ERAP1 protein level in the C1R cells. The level of ERAP1 after siRNA was compared to the value after NC to calculate the level of suppression. Cell lysates were prepared from the different cell lines before and 96 hours after specific ERAP1 or negative control siRNA nucleofection. Cells were washed and lysed with a buffered saline containing 1% Nonidet P-40 in the presence of a mixture of protease inhibitors (Sigma-Aldrich). Following SDS-PAGE on an 8% slab gel, the separated components were electroblotted onto a nitrocellulose membrane at 25 V overnight. The blots were incubated with peroxidase-coupled secondary Ab and
Legend: ERAP1 (E1) siRNA resulted in reduction of ERAP1 by 75 to 87% when compared to the negative control (NC) siRNA. There was no difference between the AS-associated and non-associated subtypes in level of suppression.
developed using an ECL detection kit. Films were analyzed with the Scion Image Analysis software.

2.2.7 Fluorescence Activated Cell Sorting (FACS)

(1) Flow cytometry of patient samples: Peripheral blood was drawn and mononuclear cells (PBMC) were isolated by Lympholyte® Cell Separation Media (Cedarlane, Canada) followed by staining for flow cytometry. Intact HLA B27 was identified by the conformation-specific mAb ME1 [265] and free MHC class I heavy chains were identified using the HC10 mAb [266]. Secondary staining was done using FITC-labeled goat anti-mouse IgG (Sigma-Aldrich, USA). As monocytes and B cells are important antigen presenting cells, they were separately gated and studied using anti-CD19 PE (BD Biosciences, USA) and anti-CD14 APC (AbD Serotec, USA) antibodies respectively. Unstained cells and cells stained only with secondary antibodies were used as controls for FACS. A total of 10,000 PBMCs were acquired and the mean fluorescence intensities (MFI) of surface HLA-B27 and FHC on PBMCs as a whole and separately on monocytes and B cells were measured (Figure 2.2).

(2) Flow Cytometry of C1R cells: C1R cell lines (transfected with B*27:04, B*27:05, B*27:06 or B*27:09) were removed from culture 96 hours after siRNA treatment. The cells were stained with the following antibodies: ME1, HC10 and MARB4. The cells were also submitted to fixation and permeabilization by the BD Cytofix/Cytoperm™ kit (BD Biosciences, USA) followed by HC10 staining for intracellular FHC (IC-HC10). Comparison was made to flow cytometry results on cells
Figure 2.2 Flow Assisted Cell Sorting of Peripheral Blood Mononuclear Cells
Legend to Figure 2.2

Flow Assisted Cell Sorting of Peripheral Blood Mononuclear Cells

(A) Scatter plot of peripheral blood cells with forward scatter (FSC) and side scatter (SSC) showing the gating for PBMC.

(B) The mean fluorescence intensities (MFI) of HC10 and ME1 staining followed by FITC tagged secondary antibody on PBMC are compared here against negative control staining (FITC tagged secondary antibody alone)

(C) Scatter plot with FL1 against FL2 showing negative control staining (FITC tagged secondary antibody alone)

(D) Scatter plot with single staining for HC10

(E) Scatter plot with single staining for ME1

(F) Histograms showing anti-CD14 APC positive monocyte population and (G) anti-CD19 PE positive B cell population compared to unstained cells as controls in the gray shaded area.

(H) Double staining of PBMC with anti-CD19 PE and anti-CD14 APC clearly showing the distinct B cell and monocyte population.

(I) Double staining of PBMC with anti-CD14 and either ME1 or (J) HC10.

(K) Double staining of PBMC with anti-CD19 and either ME1 or (L) HC10.
treated similarly with control siRNA. MARB4 has been reported to bind to HLA B27 with extended peptides [267].

2.2.8 Statistics

The MFI of HLA-B27 and FHC were compared between the genotypic groups of the ERAP1 polymorphisms tested. In the dominant model, the MFI in the patients with the minor allele only (homozygous for the minor allele) was compared to that in patients with the major allele (homozygous for the major allele or heterozygous) using the Mann-Whitney U test. For the additive model, the MFI were compared across the three genotypes of each SNP by the Kruskal Wallis Test. The comparisons were made in the whole PBMC population as well as separately in the monocyte and B cell population. The Wilcoxon Matched-Pair Signed Rank test was used to compare MFI before and after siRNA use in the B*27:05 cell lines. The Mann-Whitney U test was used to compare the change in MFI of ME1, HC10, MARB4 and IC-HC10 between C1R cells transfected with AS susceptible and non-susceptible HLA-B27 subtypes.
2.3 Results

A total of 54 patients (13 females) were enrolled in the study. These patients were attending the spondylitis clinic in the Toronto Western Hospital with a diagnosis of AS, based on the modified New York criteria. The modified New York criteria is satisfied if the patient has long standing inflammatory back pain or spinal mobility restriction with the presence of a set minimum level of damage in the sacroiliac joints. Thus these patients had unequivocal evidence of AS. Anti-TNF medications are prescribed for symptom control in AS patients and this can alter the immunological profile of patients. Hence patients on anti-TNF medications were excluded. The mean (+ SD) age of the patients was 41.2 (± 13.7) years and disease duration was 10 (± 9.8) years.

Standardized and widely accepted disease outcome measures were used to assess the symptoms of patients and to compare disease activity between groups of patients with different genotypes. This was done to ensure that changes in disease activity did not contribute to any differences noted in HLA-B27 expression on the PBMC of these patients. The best tool for assessing disease activity in AS is the BASDAI, a disease activity score derived from patient reported symptom rating. A BASDAI of ≤ 4 in a scale of 0 to 10 is considered an acceptable level of control of symptoms while a level above 4 is the threshold for starting anti-TNF medications. The mean BASDAI values in the spondylitis clinic vary between 4 and 5 and it is higher in the anti-TNF trials. C-Reactive Protein (CRP) and Erythrocyte Sedimentation Rate (ESR) are serum markers of inflammation and are widely used in arthritis assessment. However, in AS
both CRP and ESR are neither sensitive tests for inflammation nor are they correlated to patients symptoms [268].

The functional ability, or the ability to do routine and leisure activity is captured in a questionnaire and the average score gives the functional index BASFI. Spinal mobility is assessed by clinical measurements and an average of several measurements can be represented in a 0 to 10 scale called the Bath AS Metrology Index (BASMI). The mean BASDAI, BASFI and BASMI were 4.2 (± 2.4), 3.0 (± 2.6) and 2.2 (± 1.5) respectively. Thus the patients included in this study had disease activity at par with routine clinic patients. The number of patients with rs27044 genotypes CC, CG, GG were 22, 26 and 6 respectively; the number with rs30187 genotypes CC, CT and TT were 24, 23 and 7 respectively. There were no significant differences in the age, BASDAI, BASFI, hemoglobin, total leukocyte count, platelet count, ESR and CRP between the patients with different genotypes of ERAP1 rs27044 (Table 2.2) or rs30187 (Table 2.3).

2.3.1 Monocytes have the highest level of surface expression of HLA B27 and free heavy chains.

After gating on PBMC based on forward (FSC) and side scatter (SSC), monocytes and B cells were separately identified by antiCD14-APC and antiCD19-PE antibodies. The conformation specific ME1 antibody was used for estimating the intact HLA-B27 expression and HC10 was used for FHC expression on whole PBMC and separately on B cell and monocytes.
<table>
<thead>
<tr>
<th>ERAP1 SNP</th>
<th>rs27044</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CC</td>
</tr>
<tr>
<td>ERAP1 Variant</td>
<td>E730</td>
</tr>
<tr>
<td>N</td>
<td>22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ERAP1 SNP</th>
<th>rs27044</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CC</td>
</tr>
<tr>
<td>Age</td>
<td>41.5 ± 14.1</td>
</tr>
<tr>
<td>BASDAI</td>
<td>3.34 ± 2.3</td>
</tr>
<tr>
<td>BASFI</td>
<td>2.35 ± 2.2</td>
</tr>
<tr>
<td>Hemoglobin (g/L)</td>
<td>144.3 ± 15.1</td>
</tr>
<tr>
<td>Leukocytes (x 10^3 cells/dL)</td>
<td>7.39 ± 2.0</td>
</tr>
<tr>
<td>Platelets (x 10^3 cells/dL)</td>
<td>284.0 ± 58.0</td>
</tr>
<tr>
<td>ESR (mm/Hr)</td>
<td>14.0 ± 27.9</td>
</tr>
<tr>
<td>CRP (mg/L)</td>
<td>9.6 ± 13.6</td>
</tr>
</tbody>
</table>
Table 2.3 Clinical and laboratory features of patients with different ERAP1 rs30187 genotypes

<table>
<thead>
<tr>
<th>ERAP1 SNP</th>
<th>rs30187</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CC</td>
</tr>
<tr>
<td>ERAP1 Variant</td>
<td>R528</td>
</tr>
<tr>
<td>N</td>
<td>24</td>
</tr>
<tr>
<td>Age</td>
<td>41.04 ± 13.6</td>
</tr>
<tr>
<td>BASDAI</td>
<td>3.57 ± 2.3</td>
</tr>
<tr>
<td>BASFI</td>
<td>2.4 ± 2.1</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>145.7 ± 15.0</td>
</tr>
<tr>
<td>Leukocytes</td>
<td>7.5 ± 2.2</td>
</tr>
<tr>
<td>Platelets</td>
<td>283.6 ± 56.5</td>
</tr>
<tr>
<td>ESR</td>
<td>13.1 ± 26.7</td>
</tr>
<tr>
<td>CRP</td>
<td>9.3 ± 13.2</td>
</tr>
</tbody>
</table>
The median (interquartile range – IQR) values of the MFI of HLA-B27 ME1 staining were 807 (619; 1150) on monocytes, 588 (403; 775) on B cells and 445 (323; 599) on whole PBMC (Figure 2.3 A). The median (IQR) values of the mean fluorescence intensity (MFI) of FHC HC10 staining were 57.8 (44; 124) on monocytes, 22.8 (13; 44) on B cells and 21.9 (14; 31) on whole PBMC (Figure 2.3 B).

2.3.2 ERAP1 rs27044 polymorphisms affect the expression of free heavy chains on the surface of monocytes

The monocytes, B cells and whole PBMCs were separately analyzed for total MHC-I and FHC expression and this was compared between groups of patients with different genotypes of the ERAP1 SNPs tested. Significant differences were noted in the FHC expression on monocytes between groups of patients with the rs27044 SNP. Patients with the major C allele, coding for the E730 variant of ERAP1 rs27044 had significantly higher (92.1 vs 27.9; p<0.01) FHC on monocytes compared to those homozygous for the minor G allele, coding for the Q730 variant (Fig. 2.4 A,B). The Q730 variant increases the risk for AS in genetic studies. Hence patients who had only the AS-risk allele of the ERAP1 rs27044 SNP had lower FHC on the surface of monocytes.

The expression of FHC needs to be corrected for the total surface HLA-B27 expression and this was done by taking a ratio of the ME1 staining to the HC10 staining intensity. Even after correction for the ME1 staining level, the difference was significantly different between the genetic groups with higher ME1:HC10 ratio (26.7 vs
Figure 2.3 Total HLA-B*27 and Free Heavy Chain Expression was highest in the monocyte population among the peripheral blood mononuclear cells.
Legend to Figure 2.3

Total HLA-B*27 and Free Heavy Chain Expression was highest in the monocyte population among the peripheral blood mononuclear cells.

Among the PBMCs, monocytes and B cells were separately identified by antiCD14-APC and antiCD19-PE antibodies. The conformation specific ME1 antibody was used for estimating the intact HLA-B27 expression and HC10 was used for FHC expression on whole PBMC and separately on B cells and monocytes.

The Mean Fluorescence Intensity (MFI) of intact surface HLA-B27 by ME1 antibody (A) and MHC-I free heavy chains by HC10 staining (B) were highest in the monocyte population. B-cells followed by total peripheral blood mononuclear cells including T-cells had progressively lower MFI of ME1 and HC10.
Figure 2.4 Flow Cytometry Analysis of MHC-I Free Heavy Chain Expression on Peripheral Blood Mononuclear Cells from Patients with Ankylosing Spondylitis
Legend to Figure 2.4

Flow Cytometry Analysis of MHC-I Free Heavy Chain Expression on Peripheral Blood Mononuclear Cells from Patients with Ankylosing Spondylitis

Flow cytometry analysis of MHC class I free heavy chain (FHC) surface expression (HC10+ staining) on peripheral blood mononuclear cells of patients (N=54) with ankylosing spondylitis (AS). Shown are the differences in FHC surface expression on monocytes (A) between patients with different genotypes of the rs27044 polymorphism of \textit{ERAP1}. A representative histogram plot (B) from patients with median HC10 MFI from the two homozygous genotypic groups of the \textit{rs27044 ERAP1} SNP (CC vs GG) shows the difference in FHC on monocytes. The histograms representing the CC and CG group are labeled in the figure with the shaded area representing the GG group and having lower FHC than the CC group. FITC-tagged secondary antibody alone was used as negative control for staining. The difference remained statistically significant (C) even after correcting for the expression of HLA-B27 complexes (ME1 MFI) on monocytes.
12.9; p<0.01) in the group of patients with the disease susceptible Q730 variant (Fig. 2.4C).

The FHC expression changes were significant only in the monocytes (Figure 2.5 A) and not in B cells (Figure 2.5 C) or whole PBMC (Figure 2.5 E) with the ERAP1 rs27044 SNPs. There was no significant difference in the intact HLA-B27 expression as well on any of the cell populations with the rs27044 SNP (Figure 2.5 B, D, F). The ERAP1 SNP rs30187 did not affect HLA-B27 or FHC expression on any of the cell populations (Figure 2.6).

2.3.3 Surface expression of HLA B27 and FHC was not related to disease activity:

BASDAI, a marker for inflammation, was comparable in patients across the different ERAP1 genotypes (Tables 2.2, 2.3). The CRP and ESR appeared to be higher in the group of patients with the E730 and R528 variants of ERAP1, but this was not statistically significant. BASDAI, a patient reported form, is the gold standard for assessing inflammation in AS while ESR and CRP are not good markers unlike in other arthritides like RA [269]. There was no correlation of the surface expression of HLA-B27 or FHC expression in any cell population with the inflammatory markers ESR or CRP, nor with the clinical variables BASDAI or BASFI. The correlation between FHC expression on monocytes and inflammatory markers (Figure2.7) BASDAI (A), ESR (B) and CRP (C) are shown. Scatter plot with FHC MFI on y-axis and markers of inflammation on x-axis are shown (Figure 2.7) and the linear regression lines with 95% confidence intervals are also depicted.
Figure 2.5 Comparing the effects of ERAP1 rs27044 SNP on free heavy chain and Intact HLA-B27 expression on different cell populations

[Graphs showing mean fluorescence intensity for HC10 and ME1: Monocytes, B Cells, and PBMC with CC, CG, and GG genotypes]
Legend to Figure 2.5

Comparing the effects of ERAP1 rs27044 SNP on free heavy chain and Intact B27 on different cell populations

The effect of the rs27044 polymorphism of ERAP1 on the HLA-B27 and free heavy chain expression on all cell populations is shown here. Out of all comparisons done, statistically significant difference (* = p<0.01) was noted in only the FHC expression on the surface of monocytes (A, same as shown in the previous figure). There was a trend mirroring this change with lower FHC expression in the total PBMC population (E). In the monocytes (B) and B cells (D) the intact HLA-B27 expression, as identified by ME1 staining, was almost identical in patients with different rs27044 polymorphisms. In the whole PBMCs, though not statistically significant, the ME1 expression showed a trend towards lower levels in patients with the GG genotype.
Figure 2.6 Comparing the effects of \textit{ERAP1 rs30187 SNP} on free heavy chain and Intact HLA-B27 expression on different cell populations.
Legend to Figure 2.6

Comparing the effects of \textit{ERAP1 rs30187} SNP on free heavy chain and Intact HLA-B27 expression on different cell populations

There was no significant effect of the \textit{ERAP1 rs30187} SNP on FHC or intact HLA-B27 expression on Monocytes (A,B), B cells (C,D) and whole PBMC (E,F). Statistical analysis for the differences between patients with different genotypes was done using the Kruskal Wallis One Way Anova test. For the recessive model, comparing the patient with the major allele (CC and CT) to those patient with only the minor allele (TT) I used the Mann-Whitney U Test. These are standard statistical analysis test available in the statistical package of Predictive Analysis SoftWare (PASW). The Kruskal Wallis test is used when comparing more than 2 unrelated groups and Mann-Whitney test is used for 2 groups (CC and CT are grouped as one in the recessive model).
Figure 2.7 Comparison of inflammatory markers across the different genotypes

Legend: As seen from the figures, there was no correlation of free heavy chain expression on monocytes to inflammatory markers like BASDAI (A), ESR (B) or CRP(C). Thus changes seen in FHC expression cannot be explained by change in the level of inflammation. Moreover, there was no significant difference in the inflammatory markers, between groups of patients with different ERAP1 genotypes.

X-Y scatter plots are used to show the relationship between two variables. Linear fit lines are drawn to show graphically the relationship based on linear regression of the data. The fit lines were drawn by the PASW 18 program based on a linear fit method and the 95% confidence intervals are shown.
2.3.5 ERAP1 suppression of HLA-B*2705 C1R cells causes changes in intracellular free heavy chains and surface MHC-peptide complexes

In this first round of investigations on C1R cells expressing HLA-B*2705, by FACS I measured intact HLA-B27 and surface FHC as well as the level of IC-FHC and a subpopulation of HLA-B27 reacting with the mAb MARB4 which has been reported to bind HLA-B27 presenting unusually long peptides [267]. The C1R cells were subjected to ERAP1 suppression by siRNA and negative controlled siRNA (non-targeting siRNA) was used as control. Data from five siRNA experiments showed a significant increase in the IC-HC10 (55.3 vs. 32.9; Figure 2.8 A) and MARB4 staining (8.2 vs. 6.5; Figure 2.8 C) after ERAP1 silencing. Representative histograms are shown for IC-HC10 increase (Figure 2.8 B) and MARB4 increase (Figure 2.8 D) following ERAP1 siRNA compared to negative control siRNA. There was no significant difference in the HLA-B27 complexes identified by ME1 (Figure 2.8 E) and surface FHC staining by HC10 antibody (Figure 2.8 F).

2.3.6 C1R cells expressing different HLA-B*27 subtypes have variable level of total MHC-I expression

To establish the baseline level of total MHC-I expression in the C1R cells with different HLA-B27 subtypes, I did western blot analysis of the cell lysates and stained with the HC10 antibodies that detects HLA-B and HLA-C free heavy chains. As C1R cells express HLA-Cw4, HC10 would detect the respective HLA-B27 subtype as well as HLA-Cw4 in the western blots. HLA-B*2705 expressing cells had the highest level of MHC-I
Figure 2.8 Effect of siRNA induced ERAP1 suppression on HLA-B*27 expression in C1R cells expressing HLA-B*2705
Legend to Figure 2.8

Effect of siRNA induced ERAP1 suppression on HLA-B*27 expression in C1R cells expressing HLA-B*2705. Figures are representative of 5 experiments.

ERAP1 suppression led to consistent changes in IC-HC10 and MARB4 positive HLA-B27 complexes on C1R cells expressing HLA-B*2705. (A) Compared to control, nontargeting siRNA, when ERAP1 siRNA was used to suppress ERAP1, C1R cells with HLA-B*2705 were subjected ERAP1 suppression led to an increase in the intracellular free heavy chain (FHC) expression as assessed by HC10 staining after cell permeabilization (IC-HC10). (B) Representative plot showing increase in IC-HCq0 staining after ERAP1 siRNA (red) compared to control (blue). Negative control for staining was done by secondary antibody alone (shaded area)

(C) ERAP1 suppression led to an increase in MARB4 positive surface HLA-B*2705 molecules, compared to cells treated with control siRNA. MARB4 antibody recognizes HLA-B27 presenting long peptides. Representative plot (D) shows an increase in MARB4 staining with ERAP1 siRNA compared to control siRNA. There was no significant change in ME1 (E) or surface HC10 (F) staining with ERAP1 suppression. The figures represent five experiments.
expression detected by HC10 while B2704, 06 and 09 expressing cells had similar levels (Figure 2.9 A).

2.3.7 C1R cells expressing AS-susceptible HLA-B*27 subtypes have high MARB4 staining.

I tested the level of HLA-B*27 and FHC expression in the different HLA-B*27 subtype-expressing cells. A clear distinction between the AS-associated and non-associated subtypes was seen only with respect to HLA-B27 complexes recognized by the MARB4 antibody (Figure 2.9 E), which was higher in HLA-B*2704 and B*2705 cells compared to those expressing B*2706 and B*2709 (p=0.004). Compared to the other HLA-B*27 subtypes, surface expression of intact HLA-B*27 (ME1) was higher in the B*2704 cells (Figure 2.9 B) while intra-cellular FHC (HC10) was higher in B*2705 expressing cells (Figure 2.9 D). There was no difference in the surface FHC expression between the cell lines (Figure 2.9 C). Hence a pattern differentiating the AS-associated and AS-neutral B*27 subtypes was not present in any other baseline characteristic.

2.3.8 ERAP1 suppression leads to an increase in intracellular free heavy chains only in C1R cells expressing AS associated HLA-B*27 subtypes

I then tested whether ERAP1 interacted differentially with distinct HLA-B27 subtypes. I compared the changes induced by ERAP1 siRNA in C1R cells transfected either with the AS-associated B*27:04 and B*27:05 or with the non-associated B*27:06 and B*27:09 subtypes. After ERAP1 siRNA treatment, the IC-FHC was significantly increased in C1R cells expressing either B*27:04 (mean 2.4 ± 1.6 fold) or B*27:05 (mean 1.8 ± 0.4 fold), but not in B*27:06 (mean 1.1 ± 0.3 fold) or B*27:09 (mean 1.1
Figure 2.9 Western Blot and FACS analysis of C1R cells at baseline.

Western blot analysis shows total HLA-B expression in C1R cells. FACS shows HLA-B27, surface and intracellular free heavy chains and MARB4 positive HLA-B27 levels in C1R cells at baseline

The y-axes in all graphs are mean fluorescence intensities of the corresponding antibody
Legend to Figure 2.9

Baseline expression of HLA-B*27 and MHC-I Free Heavy Chains in C1R cells with different HLA-B27 subtypes

(A) Western blot analysis of cell lysates from C1R cells expressing the different HLA-B27 subtypes. HC10 antibody, which binds to HLA-B and HLA-C, was used to estimate their total baseline level in the cells. As C1R cells express HLA-Cw4, HC10 would detect HLA-B27 and HLA-Cw4 in the western blots. HLA-B*2705 expressing cells had the highest level of MHC-I expression detected by HC10 while B2704, 06 and 09 expressing cells had similar levels.

(B) Surface ME1 staining recognizing intact HLA-B27 was high in C1R cells expressing HLA*B2704. The level of ME1 in HLA B*2705 expressing cells was similar to those expressing the AS-neutral B27 subtypes.

(C) There was no pattern of surface MHC-I FHC expression (HC10), distinguishing the AS-associated and neutral subtypes

(D) Intracellular (IC) HC10, which identifies FHC sequestered within the cells, was higher in cells expressing HLA-B2705. But there was no pattern distinguishing AS-associated and neutral subtypes.

(E) The C1R cells expressing AS-associated subtypes had higher MARB4 staining, which recognizes HLA-B27 presenting distinct peptides that can be longer. However, it must be noted that this closely parallels the total MHC-I content as seen by HC10 staining of western blots (A). Hence this difference in MARB4 positive complexes could just be a reflection of the variation of the total MHC-I content. However, as shown such a variation was not evident by flow cytometry using the ME1 or HC10 antibodies.
± 0.1 fold), resulting in a significant difference (p=0.02) between the AS-associated and non-associated B*27 subtypes (Figure 2.10 A).

Representative plots show results of ERAP1 siRNA inhibition in C1R cells expressing HLA-B*2704 (Figure 2.10 B), HLA-B*2705 (Figure 2.10 C), HLA-B*2706 (Figure 2.10 D) and HLA-B*2709 (Figure 2.10 E).

The percentage of cells expressing high IC-FHC levels was consistently and similarly increased following ERAP1 suppression only in the AS-associated subtypes (Figure 2.11 A and B) but not in the AS-neutral subtypes (Figure 2.11 C and D). This analysis was done by gating the cells with IC-HC10 staining beyond a threshold value (Figure 3.5 A and B) set based on negative staining with secondary antibody alone: FITC-labeled goat anti-mouse IgG (Sigma-Aldrich, USA).

2.3.9 ERAP1 suppression leads to increase in MARB4 staining only in C1R cells expressing AS associated HLA-B*27 subtypes

We found an increase (p<0.001) in surface MARB4 staining following ERAP1 siRNA treatment in the C1R cells expressing B*27:04 (Figure 2.12 A) and B*27:05 (Figure 2.12 B), but no significant difference in the case of the AS non-associated subtypes B*27:06 (Figure 2.12 C) and B*27:09 (Figure 2.12 D). The increase in MARB4 MFI in the cells expressing the B*27:04 and B*27:05 subtypes was significantly higher (1.25 ± 0.2 vs 0.9 ± 0.1; p<0.01) compared to those with B*27:06 and B*27:09 (Figure 2.12 I). Similarly, the proportion of cells with high MARB4 MFI were significantly increased only in the C1R transfectants expressing the AS-associated subtypes (Figure 2.12 E, F vs G, H).
Figure 2.10 The effect of ERAP1 Suppression on Intracellular Free Heavy Chain Expression in C1R cells expressing different HLA-B*27 subtypes
Legend to Figure 2.10

The effect of ERAP1 Suppression on Intracellular Free Heavy Chain Expression in C1R cells expressing different HLA-B*27 subtypes.

Following ERAP1 suppression in C1R cells using siRNA, flow cytometry was used to study changes in intracellular FHC (IC-FHC). The change in IC-FHC was significantly higher (A) in C1R cells expressing the AS- associated subtypes (B*27:04 and B*27:05 vs B*27:06 and B*27:09).

Representative histograms show that the increase in IC-FHC is seen in C1R cells with B*27:04 (B) and B*27:05 (C) and not in those with B*27:06 (D) and B*27:09 (E). Histograms in blue and red indicate intracellular HC10 staining after negative control (NC) siRNA and ERAP1 siRNA respectively. The figures are representative of 5 separate siRNA experiments.
Figure 2.11 Percent of cells with high expression of intracellular free heavy chains

Control siRNA | ERAP1 siRNA
---|---
A | C1R-B2704 | C1R-B2704
   | 6.23 | 16.3
B | C1R-B2705 | C1R-B2705
   | 16.1 | 27.9
C | C1R-B2706 | C1R-B2706
   | 2.89 | 2.94
D | C1R-B2709 | C1R-B2709
   | 31.1 | 32.9

Intracellular HC10
Legend to Figure 2.11

Histograms showing percent of cells with high expression of intracellular free heavy chains

Based on negative control staining (shaded area in the plots) obtained after using only secondary antibody (FITC-labeled goat anti-mouse IgG) a threshold for high positive staining was set and the percentage of cells with high positive intracellular free heavy chain (IC-FHC) staining was estimated.

As seen with the mean fluorescence intensities, the percentage of cells with high positive staining for IC-FHC increased in C1R cells expressing HLA-B27 subtypes associated with AS (B2704 in A and B2705 in B) but not in AS neutral subtypes (B2706 in C and B2709 in D). The control siRNA is a non-targeting negative control siRNA.
Figure 2.12 The effect of ERAP1 suppression on HLA-B*27 complexes recognized by MARB4 antibody on the surface of C1R cells expressing different HLA-B*27 subtypes
Legend to Figure 3.5

The effect of ERAP1 suppression on HLA-B*27 complexes recognized by MARB4 antibody on the surface of C1R cells expressing different HLA-B*27 subtypes.

In the representative histograms shown, compared to negative control siRNA, following ERAP1 suppression with siRNA the surface expression of MARB4-reactive HLA-B27 molecules increased in the C1R cells expressing B*27:04 (A) and B*27:05 (B) but not in those expressing B*27:06 (C) and B*27:09 (D). Histograms in blue and red indicate MARB4 staining after negative control (NC) siRNA and ERAP1 siRNA respectively.

There was a similar observation with the number of cells that were strongly positive for MARB4 staining. Based on the negative control staining (solid area in histogram) with secondary antibody alone (FITC-labeled goat anti-mouse IgG), a threshold for high staining MARB4 positive cells was decided and the number of cells was estimated. In the representative histograms shown with control and ERAP1 siRNA, an increase in the number of MARB4 high positive cells after ERAP1 inhibition was seen only in C1R cells transfected with HLA-B*2704 (E) and B*2705 (F) but not in B*2706 (G) and B*2709 (H).

There was a significant difference (average of 5 experiments) in the change in MARB4 MFI (I) between the cell lines expressing the AS-susceptible subtypes (B*27:04 and B*27:05) compared to those expressing the non-susceptible subtypes (B*27:06 and B*27:09).
2.4 Discussion

The possibility of a distinctive interaction between peptide processing pathways and HLA B27 has been considered for some time [270]. The recent discovery of a genetic association of ER aminopeptidases and AS has raised expectations of finally identifying the missing link in the pathogenesis of AS. This is the first study to show that ERAP1 polymorphisms can affect MHC-I expression on the surface of human peripheral blood cells. We have shown that patients with the E730 variant of ERAP1 have significantly higher FHC on the surface of monocytes. In addition, the expression of intact HLA B27 is not different in patients with different polymorphic forms of \textit{ERAP1}.

The differences noted in the FHC expression appear to be due primarily to differences in the genotype of \textit{rs27044} and not to disease activity variations as there was no correlation of the surface expression of FHC and intact HLA B27 with any of the disease activity markers. Moreover, there was no difference in the clinical variables across the different genotypic groups. All patients included in the study were Caucasians, and none were on any anti-TNF medications. The fact that the difference remained significant after adjusting for the intact HLA B27 expression on the surface of monocytes further strengthens the observation.

This novel finding is significant as it establishes a biological relevance and link between the reported \textit{ERAP1} polymorphisms and \textit{HLA-B27}. It has to be established however, if the results would be the same in patients and controls who are \textit{HLA-B27-}
negative. The antibody HC10 is not specific for HLA-B27 and FHC expression may be different in patients who are HLA-B27 negative.

There was a differential effect of the rs27044 ERAP1 polymorphisms in different peripheral blood cells. We and others [165] have observed the presence of significant FHC expression only on the surface of monocytes and it is possible that the level was not sufficient to show significant differences in the other cell populations. It is also possible that peptides are handled differently or the types of peptides generated are different in different antigen presenting cells. ERAP1 levels vary widely between different tissues and it remains to be seen if the levels of ERAP1 vary between different subpopulations of white blood cells. If the baseline levels are different in monocytes and lymphocytes, a decrease in ERAP1 activity or expression might affect only the cell population with lower levels to begin with. These questions and possibilities need to be addressed in future studies.

With respect to possible relevance to disease pathogenesis, it is interesting that the polymorphisms associated with decreased gene expression of ERAP1 are the same polymorphisms strongly associated with AS [214]. Thus, if these polymorphisms lead to a decreased level of enzyme expression, a direct effect on the catalytic site may not be necessary to affect the overall enzymatic activity. The crystallographic structure of ERAP1 shows that the polymorphisms reported to date are not in the catalytic site of the enzyme [207,209]. These polymorphisms however may affect the substrate-ERAP1 interaction. A close binding of the substrate to ERAP1 might be cardinal to optimal enzymatic activity of ERAP1. The fact that peptidase activity is affected by changing
the internal amino acid sequence of the peptides and not just the amino terminal end cleaved by ERAP1 further supports this hypothesis [204].

It is also interesting that the minor allele which is seen in a higher frequency in patients with AS results in lower FHC expression on monocytes. It is compelling to think that, if FHC expression is important in the pathogenesis of AS, why should the AS-associated minor allele of the tested polymorphism lead to a decrease in FHC. The purpose of this study was to see if there is an interaction between ERAP1 and HLA B27 and that appears to be the case. It is possible that the Q730 variant of ERAP1 leads to abnormal FHC that are retained in the ER leading to ER stress. ER retention would result in lower surface expression of FHC with the net result that patients with the E730 variants have higher FHC surface expression. A drawback of the studies on PBMC is that, IC-FHC was not analyzed. However, this was done in the C1R cells and a reduction in ERAP1 levels did result in increased IC-FHC. FHC can also be ligands for the immunosuppressive receptors like the LIR receptors [233,241]. Thus a lower expression of FHC could lead to less activation of the immunosuppressive receptors and result in inflammation.

The new crystal structure shows that ERAP1 exists in both an open and closed conformation. The closed conformation is enzymatically more active than the closed form. This closing action happens after the peptide is accommodated in the catalytic groove. The K528R variant is on the outer surface of domain II of ERAP1 and is unlikely to directly affect the catalytic effect. But this variation can lead to abnormal interactions with neighboring residues and inhibit the conformational change of ERAP1,
indirectly reducing enzymatic activity. The Q730E variant which we have seen to be affecting FHC expression in our studies has a change in the inner surface of domain IV of ERAP1. Domain 4 has a concave surface and arches over the catalytic site on domain II forming a large cavity. As Q730E is on the inner surface, in the closed state, this approximates the active site with the bound peptide. Thus this residue change could impact the interaction of ERAP1 with the substrate.

It is now known that both the R528 (ERAP1 rs30187) and E730 (ERAP1 rs27044) variants have altered peptide trimming activity [261]. However, it is difficult to predict the effect of the variation without knowing the substrate involved. There is significant variability in the enzyme activity of the Q730E and K528R variants depending on the substrate and its concentrations. Hence the effects of these ERAP1 variants may change with cell type and milieu. In the presence of a non-substrate peptide activator of ERAP1 (SIINFEKL) both R528 and E730 had lower enzymatic activity than the wild type. In the absence of the activator, the substrate reaction consisted of leucine - amido methyl coumarin (L-AMC) alone wherein the R528 variant had lower peptide trimming effect compared to the wild type enzyme while the E730 variant had increased peptide trimming effect. In low substrate concentration levels the E730 variant had 2 fold lower peptide trimming action compared to the wild type variant. Evnouchidou et al. reported that both the E730 and R528 variants decreased HLA-B27 expression on cells with a greater effect conferred by the E730 variant [261]. Evans et al. recently reported decreased trimming by the R528 variant (rs30187) associated with AS [128]. However, the Q730E variants were not studied.
Thus it is quite clear from the studies in hypertension and the recent functional studies that the ERAP1 variants associated with AS can lead to a change of enzymatic function. However, the studies by Evanouchido et al. outlined above clearly show the complexity of the ERAP1 enzymatic reaction kinetics which appears to follow a substrate inhibition model [261]. It can be appreciated that both gain and loss of function can affect normal antigen processing. FHC can decrease on the surface as a result of either an increase in function leading to the generation of more B27-binding peptides. It could also reflect a loss of ERAP1 function with less degradation of the B27-binding peptides. Hence our results cannot definitively answer the question of whether a loss of function or gain of function underlies the observations. What we have demonstrated is that polymorphisms of rs27044 are functionally relevant and are associated with significant changes in the surface HC10-staining population (FHC). The Q730 variant was associated with lower surface FHC. This by itself shows that the polymorphism has a functional impact on MHC ± MHC-peptide expression.

Unlike the studies on PBMC, there was no significant difference in the surface FHC expression. Probably, the peptide-HLA complexes resulting from ERAP1 suppression are so unstable after significant reduction in ERAP1 activity, that they cannot be transported to the surface. Hence all FHC would be retained in the ER resulting in minimal variation of the surface FHC. We did see a strong increase in the IC-FHC in the AS-associated HLA-B27 subtypes and the relevance of this finding to AS is further strengthened by the variation observed between AS-associated and non-associated subtypes. There are differences between the conformation and the flexibility of the peptide binding grooves of these HLA-B27 subtypes [259]. Changes in
FHC expression have been implicated in the pathogenesis of AS [163,167]. FHC dimers can be recognized by leukocyte receptors such as Killer Cell Immunoglobulin-like Receptors (KIR) and Leukocyte Immunoglobulin-like Receptors (LILR) leading to abnormal immune responses [163,167]. FHC accumulation in the ER can lead to an unfolded protein response and ER stress with production of inflammatory cytokines [160].

The results above should be interpreted in light of the fact that cellular activation can affect MHC-I expression. I have not corrected the FHC and HLA-B27 expression for the level of activation of the cells. However, the clinical markers of inflammation like CRP and ESR were not different in the different genotypic groups. It should also be pointed out that HC10 is not specific for FHC of HLA-B27 origin and could reflect FHC from any HLA-B or HLA-C molecule. The exact specificity of MARB4 has not yet been defined.

Mass spectrometry determined that the MARB4-reactive population presented peptides ranging from 8 to 33 amino acids [267]. It was subsequently shown to detect HLA B27 with very low or absent β2 microglobulin but was conformation-dependent, like mAb ME1 and unlike mAb HC10 [271]. The MARB4-positive complexes seem to be different with respect to the peptides presented and possibly the conformation of the B27-peptide complex. Thus changes in MARB4 MFI in the current results are relevant for modeling pathogenesis, because this could ultimately affect immune recognition and the response generated.
Dr. Andreas Ziegler and Dr. Barbara Uchanska-Ziegler from the Institut für Immungenetik, Charité-Universitätsmedizin Berlin, Freie Universität Berlin, Berlin, Germany, are our active collaborators and they addressed the specificity of MARB4 in our combined manuscript [272]. Dr. Ziegler carried out binding experiments in BM36.1 cells transfected with B*27:05. MARB4 reacted only with B*27:05 presenting peptides with Gly in the N-terminus. Up to six amino acids could be added to the N-terminal end without abolishing MARB4 reactivity, provided the Gly-Arg motif (Gly and Arg in the A and B pockets, respectively) was present in the peptide. These results show that MARB4 reacts with B*27:05 molecules in a peptide-dependent fashion which is additionally restricted to peptides that are presented such that a Gly residue occupies the A pocket. Peptides with an N-terminal Gly and a length of nine or more residues are known to be poor substrates for ERAP1 [273]. Reducing the activity of ERAP1 could further increase their availability for presentation by HLA-B27 molecules. This could explain why MARB4 staining complexes increased after ERAP1 suppression.

The presence of Gly at the N-terminus leads to a conformational rearrangement of several heavy chain residues at one end of the binding groove (Arg62, Glu163, Trp167) that appear to be part of the MARB4 epitope on HLA-B27 molecules[274]. Hence MARB4, by detecting HLA-B27 complexes displaying a peptide with Gly at the N-terminus (a negative for ERAP-1 activity) and Arg at P2 (ideal for HLA-B27 binding), could be seen as a unique antibody that identifies a peptide-B27 subpopulation resulting from insufficient cleavage by an aminopeptidase which may be of paramount importance in antigen presentation.
2.5 Conclusions:

ERAP1 variants can lead to changes in MHC-I FHC expression on PBMC. The MHC-I FHC expression on monocytes is significantly higher in AS patients with the E730 variant of ERAP1. Alteration of ERAP1 levels can affect both the conformation and the peptide display by HLA-B27 molecules. Furthermore, these properties appear to be differentially affected in cells expressing AS-associated as opposed to non AS-associated HLA-B27 subtypes. The ERAP1-B*27 functional interaction could be a key missing link in the pathogenesis of AS.
Chapter 3

This chapter was published in the Journal of Rheumatology: Haroon N, Tsui FWL, Chiu B, Tsui HW, Inman RD. Serum cytokine receptors in ankylosing spondylitis: Relationship to inflammatory markers and endoplasmic reticulum aminopeptidase polymorphisms. J Rheum 2010;37(9):1907-10.

FWLT and HWT aided in genotyping the patients. BC aided in extracting DNA from peripheral blood and storing serum. RDI guided planning the experiments. All authors approved the draft. I sorted the serum, performed all ELISAs, did statistical analysis and prepared the draft.

3 Serum Cytokine Receptors in Ankylosing Spondylitis: Relationship to inflammatory markers and Endoplasmic Reticulum Aminopeptidase Polymorphisms

3.1 Background

HLA-B27, the major susceptibility gene, has an estimated attributable risk of 16-50% [28]. The role of HLA-B27 remains unknown and recently other risk factors including non-MHC genes such as IL1R2, IL12B, TNFRSF1A and IL23R have been identified [127,128,137,275]. As discussed above, the strongest non-MHC gene associated with AS is ERAP1 (endoplasmic reticulum aminopeptidases 1) with a population attributable risk of 26% [137]. ERAP1 is also known as ARTS-I (aminopeptidase regulating tumor necrosis factor receptor [TNFR] shedding 1).

ERAP1 plays a cardinal role in peptide trimming within the ER for MHC-I antigen presentation [260]. Mice lacking ERAP1 show reduced surface expression of MHC-I and altered presentation of MHC-I associated antigens [223]. I have presented evidence in the previous chapters that ERAP1 polymorphisms and varying ERAP1 levels can affect HLA-B27 expression with increase in MARB4 positive complexes on the cell surface and
more IC-FHC accumulation. In this section I address another known function of ERAP1.

As the alternate name ‘ARTS-1’ suggests, ERAP1 was first established as an enzyme that plays a role in clipping of the membrane bound receptor of TNFα (TNFRI) [237]. Subsequently it was found that ERAP1 also helps in clipping the receptors for interleukin 1 receptor II (IL1RII) and IL6Ra [235,236]. Abnormal clipping could result in lower levels of circulating cytokine receptors in circulation. This could result in an imbalance in the receptor-cytokine ratio with more cytokine causing inflammation.

I formulated a hypothesis based on the receptor shedding function of ERAP1 as this could provide a basis for the association of ERAP1 with AS. The AS associated ERAP1 variants are known to have less enzymatic activity. AS patients could have less ERAP1-mediated cytokine receptor shedding and more circulating TNFα, IL-1 and IL-6 leading to pronounced inflammation. If so, patients with evidence of active inflammation should have lower serum cytokine receptors. Moreover, patient with different variants of ERAP1 should have different levels of cytokine receptors in circulation.

3.2 PATIENTS AND METHODS

3.2.1 Patients

Patients satisfying the modified New York criteria for AS, attending the Spondylitis Clinic at Toronto Western Hospital and consenting to participate were included in the study. Sequential patients from the clinic were invited to participate, irrespective of disease activity at the time. Patients on anti-TNF treatments or other
biological agents were excluded. All patients underwent a comprehensive clinical and radiographic examination. Patients were assessed for disease activity by BASDAI, ESR and CRP. Functional capacity was evaluated using the BASFI.

3.2.2 Genotyping

DNA was prepared by Basil Chu in the Inman Lab, from peripheral blood using a standard DNA isolation kit (Gentra Systems, Minneapolis) as described in the previous chapter. Briefly, the principle of cDNA isolation from peripheral blood involves lysing RBCs which are non-nucleated and lacking genomic DNA first, followed by lysing WBCs in the presence of a DNA stabilizer to protect the DNA from the action of DNAase. Briefly, RBCs and WBCs were lysed with a proprietary cell lysis solution from Gentra systems containing an anionic detergent. Salt precipitation was used to precipitate protein and alcohol precipitation was used to precipitate DNA from the supernatant. DNA was mixed with 100% isopropanol. The precipitated DNA was subsequently washed with 70% ethanol and hydrated with a DNA hydration solution consisting of 1 mM EDTA and 10 mM Tris.Cl at pH 7.5. DNA concentrations were measured and samples aliquoted.

Two SNPs that were strongly associated with AS (rs27044 and rs30187), included in the immunological studies described above, and one that was found to be protective for AS in the GWAS studies (rs10050860) were selected for this study. The ERAP1 SNP rs10050860 leads to an amino acid change D575N where aspartic acid in position 575 is changed to asparagine. This change happens in Domain III of ERAP1 away from the catalytic site. Optimized allelic discrimination assays for SNPs from
Applied Biosystems (Foster City, CA, USA) and the Applied Biosystems 7900HT fast system was used for the PCR [262].

Briefly, allelic discrimination assays are based on the ability to directly detect the polymerase chain reaction (PCR) product by measuring the increase in fluorescence of dye-labeled DNA probes. Two probes are used, one for each allele. Each probe consists of an oligonucleotide with a 5´-reporter dye and a 3´-quencher dye. The reporter dye is different in the two alleles while the quencher is common. Once the PCR reaction progresses in a 5´-3´ direction, the 5´ reporter dye gets released from the quenching effects of the quencher on the 3´ end of the probe leading to a signal that can be detected by the PCR machine. In our system TET (6-carboxy-4,7,2´,7´-tetrachlorofluorescein) is covalently linked to the 5´ end of the probe for Allele 1. FAM (6-carboxyfluorescein) is covalently linked to the 5´ end of the probe for Allele 2. TAMRA (6-carboxy-N,N,N´,N´-tetramethylrhodamine) is the common quencher. Thus if signals from TET and FAM are detected, the patient is heterozygous with both allele present. Single signals would indicate a homozygous state for the corresponding allele.

3.2.3 Serum cytokine and soluble cytokine receptor levels

Sera were collected from AS patients at the time of clinical evaluation and stored in aliquots at -70°C until analyzed. ELISA was used to determine the serum levels of the cytokines IL1, IL6 and TNFα (BD, San Jose, CA) and the soluble receptors of IL1RII (sIL1RII), IL6R (sIL6R) and TNFRI (sTNFRI) (R & D, Minneapolis, MN), performed in duplicates according to the manufacturer’s directions. The minimum detectable cytokine serum levels by the ELISA respectively for IL1, IL6 and TNFα were
0.8, 2.2 and 2 pg/ml. The minimum detectable serum cytokine receptor levels were 10, 6.5 and 0.77 pg/ml for sIL1RII, sIL6R and sTNFRI, respectively.

### 3.2.4 Statistical analysis

Patients were stratified according to the genotypes of the *ERAP1* SNPs (*rs30187 [CC, CT, and TT]; rs27044 [CC, CG, and GG]*) and *rs10050860 [CC, CT, and TT]*. For small sample numbers, parametric tests should be used for analysis. Comparison across 3 or more groups can be done by the Kruskal-Wallis H test [276], while comparison of two groups is done by the Mann-Whitney U test [277] in parametric statistical testing. The soluble cytokine receptor levels in the different genotype groups were compared using the Kruskal-Wallis H Test. To rule out recessive effects, the Mann-Whitney U test was used to compare soluble receptor levels in patients with the major allele of the three SNPs to those homozygous for the respective minor allele. Thus comparisons were made between the following groups: patients with the Q730 variant: E730 variant alone (*rs27044*), K528:R528 alone (*rs30187*) and D575:N575 alone (*rs10050860*).

As inflammation could affect the cytokine receptor levels, ‘partial correlations’ was used to correlate the serum level of each cytokine receptor to other receptors [278]. The partial correlations test allows for correction for other parameters and we corrected for the objective markers of inflammation (ESR and CRP). The soluble receptor level was also correlated with BASDAI, BASFI, ESR and CRP, using Spearman’s Rank correlation.
3.3 RESULTS

3.3.1 Patients

Eighty AS patients (21 women and 59 men) with a mean (± SD) age of 42.3 (± 10.6) years and mean (± SD) disease duration of 18.5 (± 11.4) years were enrolled in the study. Sixty seven patients (83.7%) were HLA-B27-positive. The mean ESR and CRP were 18 ± 15.6 mm/hr and 14.8 ± 18.8 mg/L, respectively. Both ESR and CRP are measures of inflammation and are tested using blood drawn from patients. The normal value of ESR and CRP are below 20 mm/Hr and 10 mg/L, respectively with slight variations between labs, depending on the kit used. Although they are recognized as very good markers of systemic inflammation it is not a particularly good marker in AS [269]. This may be because inflammation is more localized in AS.

BASDAI, a disease activity score, on the other hand better reflects disease activity in AS [269]. It is calculated from a patient reported questionnaire of symptoms and a value below 4 is considered an acceptable level of disease activity. The BASFI is a functional index and reflects any functional limitations the patients may experience. The mean BASDAI and BASFI scores were 5.3 ± 2.4 and 4.4 ± 2.6 respectively. Thus the patients included in the study had active disease and this is the typical range seen in AS patients attending rheumatic disease clinics. There was no significant difference between the patients in the different genotypic subsets with regards to age, disease duration, BASDAI, ESR or CRP.
3.3.2  Genotyping

Two AS-associated and one AS-protective ERAP1 SNP was genotyped as detailed in the methods. The AS-associated ERAP1 SNPs rs30187 and rs27044 SNPs increase the risk of AS while rs10050860 protects against AS. The use of the terms ‘associated’ and ‘protective’ simply means that, compared to the general population, the minor allele of the AS-associated SNPs are overrepresented while those of the protective SNPs are underrepresented in AS patients. The allele frequencies of ERAP1 SNPs were 29, 39 and 12 (GG, AG, AA) for rs30187; 37, 39 and 4 (CC, CG and GG) for rs27044 and 53, 25 and 2 (CC, TC, TT) for rs10050860 respectively.

3.3.3  Soluble cytokine receptor levels

The mean (±SD) serum soluble cytokine receptor levels in the AS patient cohort were 1.38 ± 0.42 ng/ml (sTNFRI), 11.89 ± 4.3 ng/ml (sIL1RII) and 29.63 ± 11.8 ng/ml (sIL6R). X-Y scatter plots were used to show the relationship between two variables. Linear fit lines were drawn to show graphically the relationship based on linear regression of the data. The fit lines were drawn by the PASW 18 program based on a linear fit method and the 95% confidence intervals are shown in blue. The linear fit method uses the formula \( y = a + bx \). Thus based on the data available, the corresponding value of \( Y \) for any value of \( X \) will depend on ‘\( a \)’ (constant) and ‘\( b \)’ (the slope of the fit line).

Contrary to the expectation that lower cytokine receptors would be associated with higher inflammation, we observed (Figure 3.1 A,B) a significant positive (rather than negative) correlation of sTNFRI with the markers of inflammation CRP (\( R=0.43; \)
Figure 3.1 Correlation of TNFRI with CRP and ESR

A

R = 0.43 p < 0.001

B

R = 0.30 p = 0.01
Legend to Figure 3.1

Correlation of TNFRI with CRP and ESR

The scatter plots above show TNFRI level on the y-axis and CRP (A) or ESR (B) on the x-axis. There was significant positive correlation of serum soluble TNFRI level with CRP (A) and ESR (B). The correlation between TNFRI and CRP was stronger ($R=0.43$, $p < 0.001$) compared to that with ESR ($R = 0.30$; $p = 0.01$).

X-Y scatter plots are used to show the relationship between two variables. Linear fit lines are drawn to show graphically the relationship based on linear regression of the data. The fit lines were drawn by the PASW 18 program based on a linear fit method and the 95% confidence intervals are shown in blue. The linear fit method uses the formula $y = a + bx$. Thus based on the data available, the corresponding value of $Y$ for any value of $X$ will depend on ‘$a$’ (constant) and ‘$b$’ (the slope of the fit line). When the fit line is at 45 degrees between the $X$ and $Y$ axes, it indicates perfect correlation of $X$ and $Y$ values with an $R$ value (correlation co-efficient) of 1. Correlation is significant if the $p$ value is $< 0.05$. However, the strength of correlation depends on the $R$ value. Traditionally $R > 0.9$ indicates very strong correlation, $0.7-0.9$ is strong, $0.4-0.7$ is moderate and $0.2-0.4$ is weak correlation.
p<0.001) and ESR (R=0.30; P=0.01). There was no correlation of sTNFRI with BASDAI. There was also no correlation of sIL1RII and sIL6R with ESR, CRP or BASDAI (Figure 3.2).

3.3.4 No significant effect of ERAP1 polymorphism on serum cytokine receptor level

There was no significant difference in the soluble cytokine receptor serum levels between the different genotype groups of ERAP1 rs27044, rs30187 and rs10050860 polymorphisms (Figure 3.3). A dominant model was tested where the patients who have the major allele of a polymorphism were grouped together and the other group consisted of patients with only the minor allele. Thus comparisons were made between the following groups: patients with the Q730 variant: E730 variant alone (rs27044), K528:R528 alone (rs30187) and D575:N575 alone (rs10050860).

Thus in the Kruskal-Wallis H test, there was no difference in the serum cytokine receptor levels across the three genotypic groups of each SNP tested. Similarly, with the Mann-Whitney U test comparing patients with the major and minor variants mentioned above, there was no significant influence of the ERAP1 polymorphisms on the cytokine receptor levels.

3.3.5 Serum cytokine levels

Serum cytokine levels are difficult to measure and were detectable only in a minority of patients that I enrolled. Serum levels were detectable in 21. 29 and 22
Figure 3.2 No correlation of CRP, ESR and BASDAI with serum IL-6Rα and IL-1RII
Legend to Figure 3.2

Correlation of CRP, ESR and BASDAI with serum IL-6Rα and IL-1RII

In the scatter plots, IL-6R (blue) and IL1RII (gray) are shown on the y-axis with markers of inflammation CRP (A), ESR (B) and BASDAI (C) on the x-axis. The linear fit method was used to draw the fit lines and there was no significant correlation between the level of soluble IL-6Rα and IL-1RII with CRP, ESR or BASDAI.

X-Y scatter plots are used to show the relationship between two variables. Linear fit lines are drawn to show graphically the relationship based on linear regression of the data. The linear regression fit lines are shown in blue for IL-6R and gray for IL1RII. The linear fit method uses the formula \( y = a + bx \).
Figure 3.3 Serum Cytokine Receptors in patients with different ERAP1 polymorphisms

- **A**: TNFRI (ng/ml)
- **B**: IL-6Ra (ng/ml)
- **C**: IL-1RI (ng/ml)

For each receptor, there are three genotypes indicated:

- **C**: CC, CG, GG
- **F**: AA, AG, GG
- **I**: CT, CC, TT

**Genotypes and Labels:**
- **ERAP1 rs27044**
- **ERAP1 rs30187**
- **ERAP1 rs10050860**
Legend to Figure 4.3

Serum Cytokine Receptors in patients with different ERAP1 polymorphisms

This figure shows box plots or box and whisker plots which are ideal for showing the distribution of the values across the groups. These plots do not assume a parametric nature for the data. The central line stand for the median value, the upper and lower edges of the box represent the 25th and 75th quartile values while the whiskers represents 2 standard errors of the mean value.

As seen in the figure, the distribution of the values was similar across the different groups and the serum cytokine receptors of TNFRI, IL-6Rα and IL-1RII, respectively were not significantly different in patients with different genotypes of ERAP1 rs27044 (A, B, C), rs30187 (D, E, F) and rs10050860 (G, H, I).
patients respectively for IL1, IL6 and TNFα assays. The mean (±SD) serum levels were 28.4 ± 27.1 (IL1), 44.4 ± 71.7 (IL6) and 104.5 ± 90.2 (TNFα).

There is a possibility that cytokine receptors in circulation can affect the serum levels of cytokines. However there was no difference in the soluble cytokine receptor levels between patients with and without detectable corresponding cytokine levels. The ratio of serum cytokine to the soluble receptor levels was not different in patients with different genotypes of the *ERAP1* polymorphisms tested.

### 3.4 Discussion

This is the first reported study of the role of the cytokine receptor shedding function of ERAP1 in the pathogenesis of AS. I found no influence of the respective *ERAP1* SNPs (*rs30187*, *rs27044* and *rs10050860*) on the serum levels of sTNFRI, sIL1RII and sIL6R. I found a significant positive correlation of sTNFRI with the inflammatory markers ESR and CRP. There was no correlation of serum cytokine receptors with BASDAI.

Thus the reported *ERAP1*-AS association is unlikely to be related to cytokine receptor shedding function of ERAP1. There was no difference in the serum levels of the cytokine receptors in the different *ERAP1* genotypic groups. I started with the hypothesis that decreased receptor shedding can lead to more inflammation as a result of less blocking receptors in circulation. On the contrary, I found that TNFRI was higher in patients with higher CRP and ESR. Thus the original hypothesis was rejected. This observation could be due to more receptor shedding secondary to inflammation,
resulting in higher sTNFRI in serum. The strong correlation seen in the current study between TNFRI and CRP could indicate that TNF receptor shedding is more sensitive to inflammation than IL1 and IL6 receptors.

Serum cytokine measurements are not straightforward and previous attempts found these not to be informative in AS [279,280]. It was nevertheless important to attempt this in the present study. My hypothesis involved an altered balance of cytokine to its corresponding receptors that lead to inflammation. Hence I tested if the ratio of serum TNFα, IL6 and IL1 to their corresponding receptors was affected by the polymorphisms in ERAP1. This also was negative, corroborating my earlier findings.

Unlike humans, mouse ERAP1 has been found only in the ER and has not been shown to be important in cytokine receptor shedding. The only known function of ERAP1 in mice is ER peptide processing for MHC loading and as discussed before, the ERAP1 KO mouse has an abnormal peptide-MHC repertoire at the cell surface [223]. I have already shown that ERAP1 variants can result in changes in FHC expression in monocytes of patients with AS. ERAP1 suppression also led to accumulation of FHC in the intracellular compartment. Peculiar B27-peptide complexes recognized by the MARB4 antibody were increased following ERAP1 suppression. Thus there is more evidence that altered peptide processing can explain the AS-ERAP1 association better than receptor shedding.

Another strong indicator of this possibility is the fact that a haplotype of ERAP1/ERAP2 is associated with AS [215]. To date ERAP2 has not been shown to be involved in cytokine receptor shedding. Thus the association of AS with a haplotype of
ERAP1/ERAP2 is a strong indicator that ERAP1 influences the pathogenesis of AS by its peptide processing function.

3.5 Conclusion

The serum soluble TNFRI level correlates with the inflammatory markers ESR and CRP. ERAP1 polymorphisms, reported previously to be associated with AS, do not influence the serum cytokine receptor levels in AS patients.
Chapter 4

This chapter was published in Arthritis and Rheumatism: Haroon N, Maksymowych WP, Rahman P, Tsui FWL, O’shea FD and Inman RD. Radiographic severity in ankylosing spondylitis is associated with polymorphism in Large Multifunctional Peptidase 2 (LMP2) in the SPARCC cohort. Arthritis Rheum 2011 (Epub).

WPM, RDI and O’Shea FD helped in recruiting patients. PR helped with genotyping the patients. RDI guided planning the study. All authors approved the draft. I recruited patients, scored the X-rays for damage assessment, performed all statistical analysis and prepared the draft manuscript.

4 Radiographic severity in ankylosing spondylitis: The effect of genetic polymorphisms in ERAP1 and other elements of the antigen presenting machinery

4.1 Background

Ankylosing Spondylitis (AS) is a chronic inflammatory joint disease predominantly affecting the spine leading to pain and disability [281]. In addition to inflammation of the spine, there is progressive new bone formation often leading to ankylosis of the spine. New bone formation is a hallmark of AS but pathogenesis is not well understood.

A disconnect between pain and radiographic manifestations is well known. Patients with severe pain may not have any x-ray changes while those with extensive ankylosis may be asymptomatic. Although it is proposed that inflammation sets in first followed by new bone formation, the link between inflammation and new bone formation is not clear[180,181,282].

Previous attempts at identifying predictors of spinal fusion and radiographic progression have met with variable success. DKK-1 and a dysregulated Wnt signaling pathway is considered to be key to new bone formation in AS[188,192,199,283]. Low
serum sclerostin levels have been reported to be linked to structural damage [284]. Urinary C-terminal cross-linking telopeptide of type II (CTX-II) is elevated in patients with more severe radiographic damage at baseline and more progression on follow up [285,286]. Serum matrix metalloproteinase 3 (MMP3) was found to be an independent predictor of radiographic progression in one study, especially in patients with pre-existing radiographic damage [287]. The best predictor of progression in AS so far is not a biomarker but rather the presence of syndesmophytes at baseline [288].

The strong association of AS with \textit{HLA B27} was identified in 1973. Despite the strong genetic association with AS, there is no definite evidence that HLA-B27 affects the severity of AS as assessed by radiographic progression [146]. However, these studies are confounded by the high prevalence of HLA-B27 in the AS population. Thus other genetic markers are required to predict radiographic severity. Identification of such a genetic marker would not only aid in identifying patients at higher risk of radiographic progression, but also shed light on the pathogenesis of new bone formation in AS. \textit{HLA B27} has an attributable risk (AR) of only 16-50\% in AS and clearly other genes are involved in the pathogenesis of the disease [144,260]. Recurrence risk modeling studies have shown that AS is probably an oligogenic disease [144]. As discussed above, several recent genetic studies on AS have reported an association with the gene \textit{ERAP1} and this has been replicated in Canadian AS patients as well as in North American multiplex families [127,137,210-213,215,289]. With an attributable risk of 26\%, \textit{ERAP1} is second only to \textit{HLA B27} among genetic risk factors reported to be associated with AS [260].
Polymorphisms of other genes of the antigen presentation pathway including the proteasomes and the transporter associated with antigen processing (TAP) have been variably reported to be associated with AS or with extra articular manifestations of AS [290-305]. Genes strongly associated with a disease could also be a marker of severity. This study is an attempt to look at the effect of ERAP1 and other genes involved in antigen processing on radiographic severity and progression in AS. For studies involving multiple genes, the size of the cohort is important. I have collaborated with the Edmonton Spondylitis Clinic under Dr. Walter Maksymowych in this study.

4.2 Methods

4.2.1 Patients

Caucasian AS patients, diagnosed by the modified New York criteria, attending the Spondylitis clinics in Toronto and Edmonton were included in the study [68]. All AS patients were followed annually with a standardized protocol and clinical information was recorded in a common web-based database. Patients included in the progression study should have had at least two full sets of x-rays (lumbar and cervical) at a minimum gap of 1.5 years that could be scored for radiographic severity. Since it is difficult to detect progression in patients with complete or near-complete fusion of the entire spine, patients with a baseline mSASSS score of 65 or higher were excluded from the study.
4.2.2 Radiographic Scoring

Radiographic severity of AS was scored using the modified Stoke’s ankylosing spondylitis spine score (mSASSS) [306]. In brief, all patients had x-ray of the lumbar and cervical spine in the lateral view for scoring. Each vertebral corner from C2 lower end to T1 upper end in the cervical and T12 lower end and S1 upper end in the lumbar spine were scored. With a total of 12 vertebral edges in the cervical and lumbar spine each, 24 corners could be scored, each ranging from 0 to 3 points. Thus the range of the total mSASSS score is 0 to 72 for all 24 corners. The scoring is done as given in the box below and shown in figure 4.1.

<table>
<thead>
<tr>
<th>Vertebral Corner</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0</td>
</tr>
<tr>
<td>Squaring or Sclerosis or Erosion</td>
<td>1</td>
</tr>
<tr>
<td>Syndesmophyte</td>
<td>2</td>
</tr>
<tr>
<td>Bridging syndesmophytes</td>
<td>3</td>
</tr>
</tbody>
</table>

In each center, paired X-rays in a known temporal profile were scored by two blinded readers, by a consensus method. The inter-class correlation coefficients (ICC) between the centers were 0.94 and 0.57 for the status and change mSASSS scores. The smallest detectable change (SDC) was 1.95 for Edmonton and 2.73 for Toronto. Radiographic progression was assessed by noting the change in mSASSS scores ($\Delta$mSASSS) between the first and last available sets of X-rays. A Damage Duration
The scoring method for the Modified Stoke’s Ankylosing Spondylitis Spine Score (mSASSS) is shown with the major abnormalities that are scored. Squaring is defined as an absence of any gap between the anterior border of the vertebral body and a line drawn between the upper and lower anterior corners of the vertebral body. Sclerosis is the hyperdensity seen as white patches at the vertebral corners. Erosion is a break in the cortical bone at the corner. A syndesmophytes is a bony spur (new bone) developing from the vertebral corner and progressing parallel to the axis of the spine. On the contrary, an osteophyte is not scored which is also a bony spur, but progressing at right angles to the axis of the spine. When the syndesmophytes bridges from one vertebral corner to the next, the highest score of 3 points is given to each of the corners involved.
Index (DDI) was calculated as the ratio of ΔmSASSS to the gap in years between the x-rays, which gives an average increase per year. Progressors were identified as those patients who had an increase of at least 1 mSASSS unit per year (DDI ≥1) as this is the average rate of mSASSS progression reported.

4.2.3 Genotyping

DNA was prepared from the peripheral blood of AS patients and shipped to the Memorial University, Newfoundland, Canada. Genotyping for a panel of 13 coding-region SNPs in the ERAP1 (N=6), Large Multifunctional Peptidase 2 (LMP2; N=1), LMP7 (N=1), TAP2 (N=4) and TAP7 (N=1) genes (Table 4.1) was done here. In previous studies, the SNPs rs27044, rs30187 and rs10050860 in the ERAP1 gene were shown to be significantly associated with AS [137,289]. The SNP rs26653 was found to be associated with AS subsequently in Canadian patients and this was not reported in the initial genome wide studies. It was added to the analysis in this study which included patients from two Canadian centers and was not included in the previous studies on PBMCs or cytokine receptors. The remaining SNPs (rs26618, rs3734016, rs241447, rs4148876, rs2228396, rs1800454, rs2071543, rs1057141, and rs17587) were chosen based on previous data demonstrating haplotype interaction between the ERAP1 and TAP/LMP loci [307].

Minor allele frequency (MAF) refers to the frequency of the minor allele in the cohort and if the frequency is very low, statistical analysis is not reliable. Alleles with MAF lower than 5% are not included in statistical analysis. The minor allele frequency at all loci in this study was above 5%. Another quality control measure required for
Table 4.1 Genotype frequency of the tested SNPs in the study patients

<table>
<thead>
<tr>
<th>Single Nucleotide Polymorphism (SNP)</th>
<th>Gene</th>
<th>Genotype</th>
<th>Frequency of Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs26618</td>
<td>ERAP1</td>
<td>T/TC/C</td>
<td>0.58/0.36/0.06</td>
</tr>
<tr>
<td>rs26653</td>
<td>ERAP1</td>
<td>G/GC/C</td>
<td>0.40/0.52/0.08</td>
</tr>
<tr>
<td>rs27044</td>
<td>ERAP1</td>
<td>C/CG/G</td>
<td>0.44/0.48/0.08</td>
</tr>
<tr>
<td>rs30187</td>
<td>ERAP1</td>
<td>C/CT/T</td>
<td>0.30/0.57/0.13</td>
</tr>
<tr>
<td>rs10050860</td>
<td>ERAP1</td>
<td>C/CT/T</td>
<td>0.69/0.28/0.03</td>
</tr>
<tr>
<td>rs3734016</td>
<td>ERAP1</td>
<td>G/GA</td>
<td>0.94/0.06</td>
</tr>
<tr>
<td>rs17587</td>
<td>LMP2</td>
<td>G/GA/A</td>
<td>0.36/0.57/0.07</td>
</tr>
<tr>
<td>rs2071543</td>
<td>LMP7</td>
<td>C/CA/A</td>
<td>0.70/0.27/0.03</td>
</tr>
<tr>
<td>rs1057141</td>
<td>TAP1</td>
<td>A/AG/G</td>
<td>0.72/0.27/0.01</td>
</tr>
<tr>
<td>rs1800454</td>
<td>TAP2</td>
<td>G/GA/A</td>
<td>0.78/0.19/0.03</td>
</tr>
<tr>
<td>rs2228396</td>
<td>TAP2</td>
<td>G/GA/A</td>
<td>0.86/0.13/0.01</td>
</tr>
<tr>
<td>rs4148876</td>
<td>TAP2</td>
<td>C/CT/T</td>
<td>0.89/0.10/0.01</td>
</tr>
<tr>
<td>rs241447</td>
<td>TAP2</td>
<td>A/AG/G</td>
<td>0.48/0.43/0.09</td>
</tr>
</tbody>
</table>
genetic studies is to ensure that alleles comply with the Hardy-Weinberg Equilibrium (HWE). The Hardy-Weinberg principle states that frequencies of alleles remain constant in the absence of major genetic changes like mutation, migration or genetic drifts seen with epidemics. This can be tested by looking at the genotype frequency and the allele frequency. Based on the allele frequency, we can estimate the genotype frequency by a formula. If ‘A’ is the major allele and ‘a’ is the minor allele with frequencies of x and y respectively, the expected genotype frequencies for [AA], [Aa] and [aa] are $x^2$, $2xy$ and $y^2$, respectively. This expected frequency is then compared to the observed frequency in the study cohort to see if the alleles are in HWE. No marker in our study deviated significantly from the Hardy-Weinberg Equilibrium at the 1% significance level. If the HWE is not satisfied, it could refer to some form of error like systematic genotyping error or the presence of homologous regions in the genome.

The MassARRAY system (Sequenom, San Diego, CA) was used for genotyping and the genotypes were determined using MassARRAY Typer software, version 4.0.

4.2.4 Statistical Analysis

All information on genotypes and clinical data on patients included were compiled in Toronto. I managed this database and performed the statistical analysis. Regression analysis is the standard statistical analysis used to predict an outcome (dependent variable) based on several independent variables. For continuous dependent variables like mSASSS scores, a linear regression analysis is done while a binary logistic regression analysis is done for dichotomous dependent variables. Linear regression analysis was done to identify the predictors of baseline mSASSS scores.
Binary logistic regression analysis was done for predictors of progression. Univariate followed by multivariate regression analysis was done. This is the standard methodology adopted for regression analysis where the interesting independent variables are individually tested to see if they predict the outcome. Those variables which independently predict the outcome are subsequently included in a multivariate analysis where each variable included is corrected for the other included variables. Thus variables that are significantly predicting outcome in a multivariate analysis would correct for all the variables that are individually predicting outcome.

Variables significant at $p \leq 0.1$ in the univariate analysis were included in the multivariate regression model and the final model was obtained after forward method conditioned on including variables significant at $p \leq 0.05$. A forward conditional method of multivariate regression analysis is a method in which the most significantly associated variable is included in the regression equation first. Subsequently variables are added to the equation only if they are significantly contributing to the prediction equation at a significance level that we decide (0.05 or 5.00 % is standard and was used in this case).

Apart from the SNPs tested, gender, *HLA B27*, mSASSS at baseline and duration of disease were included in the univariate regression models. Considering the high correlation between age of patients and disease duration ($R=0.78; p<1 \times 10^{-10}$), age was not included in regression models. Receiver operating characteristic (ROC) curve and area under curve (AUC) analysis was done with the logistic regression model obtained by entering baseline mSASSS alone.
ROC curves are graphical X-Y plots representing the true positive rate (sensitivity) against the false positive rate (1 – specificity) for a binary outcome. ROC curves can be used to identify the best cut off values to discriminate the two outcomes by plotting the sensitivity and 1-specificity rates for the different cutoffs. Thus a curve is generated with the points yielding highest sensitivity being higher on the Y-axis and values with the highest specificity being on the left hand side of the X-axis. Thus the value that lies to the left upper corner would yield the highest sensitivity and specificity for distinguishing the binary outcome. When a statistical model is tested in this way, the AUC can be calculated. When the values are leaning toward the left upper corner, the AUC will be higher while models yielding curves that are shifted to the lower right side will have lower AUC. Thus a regression model with higher AUC indicates better prediction of the outcome (radiographic progression in this case). For interpreting the value of the model, $R^2$ is the coefficient of determination used in the context of statistical models. The value indicates how well the prediction of future outcomes can be done with the information included in the model. Naegelkerke $R^2$ indicates the proportion of the variation explained by the regression model.

In a dominant model of genetic statistical testing, the minor variant has a dominant allele effect with the group heterozygous for the minor allele differing significantly in the outcome from the group homozygous for the major allele. In the recessive model only the group homozygous for the minor allele differs from the other groups. Dominant and recessive models were initially tested and only recessive models were significant. For the SNP rs3734016, no patients were homozygous for the minor allele and so the dominant model was used in the initial univariate analysis.
Contingency 2X2 tables were analyzed with Chi-squared tests ($\chi^2$) and odds ratio (OR) of the relevant SNPs for progression were calculated. Student’s t test was used for comparison of means. The PASW Statistics 18 software was used for the analysis.

4.3 Results

4.3.1 Patients

A total of 241 AS patients (81% males and 82% HLA B27-positive) were enrolled from the two centers and were evaluated for predictors of baseline radiographic severity. Out of this, 210 patients had follow up X-rays at a minimum gap of 18 months and were included for studying predictors of radiographic progression. The patients were followed up for a mean (± SD) duration of 2.4 (± 0.8) years. There were 158 patients enrolled from Edmonton and 83 patients from Toronto (Table 4.2).

The Edmonton cohort was older (43.0 ± 12.7 vs 37.1 ± 12.5; p=0.001), with longer duration of disease (15 ± 9.7 vs 19 ± 12; p=0.01) and the proportion of HLA B27 positive patients was higher (87% vs 76%; p=0.04). The baseline mSASSS, rate of progression and the proportion of progressors was not different between the centers.
Table 4.2 Demographic features of the two AS cohorts

<table>
<thead>
<tr>
<th>Variable</th>
<th>Combined Cohort</th>
<th>Toronto</th>
<th>Edmonton</th>
<th>p*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients (N)</td>
<td>241</td>
<td>83</td>
<td>158</td>
<td></td>
</tr>
<tr>
<td>Age at baseline</td>
<td>41.0 ± 13.0</td>
<td>37.1 ± 12.5</td>
<td>43.0 ± 12.8</td>
<td>0.001</td>
</tr>
<tr>
<td>Females (%)</td>
<td>18.7</td>
<td>14.5</td>
<td>20.9</td>
<td>ns</td>
</tr>
<tr>
<td>HLA B27 (%)</td>
<td>82.3</td>
<td>75.6</td>
<td>86.5</td>
<td>0.04</td>
</tr>
<tr>
<td>Baseline mSASSS</td>
<td>16.1 ± 17.6</td>
<td>13.6 ± 18.2</td>
<td>17.74 ± 17.2</td>
<td>ns</td>
</tr>
<tr>
<td>Disease duration at baseline</td>
<td>17.6 ± 11.5</td>
<td>14.9 ± 9.7</td>
<td>19.1 ± 12.2</td>
<td>0.01</td>
</tr>
<tr>
<td>DDI</td>
<td>0.81 ± 1.3</td>
<td>0.58 ± 0.9</td>
<td>0.93 ± 1.5</td>
<td>ns</td>
</tr>
<tr>
<td>Progressors (%)</td>
<td>29.0</td>
<td>22.0</td>
<td>32.0</td>
<td>ns</td>
</tr>
</tbody>
</table>

Values are mean ± SD unless otherwise indicated in brackets.

mSASSS: modified Stoke’s ankylosing spondylitis spine score

DDI: damage duration index

ns: not significant.

DDI was calculated by dividing the change in mSASSS (ΔmSASSS) by the gap (in years) between the x-rays, which gives an average increase in mSASSS per year.

Progressors were identified as those patients who had an increase of at least 1 mSASSS unit per year (DDI ≥1) as this is the average rate of mSASSS progression reported.

* P values are reported for Student’s T test in case of continuous variables and for Chi-squared test with categorical variables.
4.3.2 Baseline radiographic severity can be influenced by polymorphisms in LMP2

Detailed regression analysis was done to assess the predictors of baseline radiographic severity in the cohort. Univariate followed by multivariate regression analysis was done. As age at baseline and duration of disease are correlated highly (R=0.78; p< 1 x 10^{-10}), age was not included in regression models. Univariate regression analysis showed that baseline mSASSS scores were associated with gender, duration of disease, the ERAP1 SNP rs30187 and the LMP2 SNP rs17587 (Table 4.3). Although, by T-test, HLA B27–positive patients had significantly higher mSASSS scores than HLA B27-negative AS patients (20.4 ± 22.2 vs 11.9 ± 17.8; p=0.01), HLA B27 was not statistically significant in the regression model (B = 5.21;p=0.09). Baseline mSASSS scores were higher by T-test in males than females (21.6 ± 22.69 vs 12.23 ± 16.44, p=0.002). Patients homozygous for the minor allele of the LMP2 SNP rs17587 (AA) had significantly higher mSASSS at baseline (27.2 vs 15.7; p=0.025). AS patients with the major allele of the ERAP1 SNP rs30187 (CC/CT) had significantly higher baseline radiographic severity (16.7 vs 9.4; p=0.04).

In multivariate analysis with the forward conditional method, duration of disease (B=0.74; p = 1 x 10^{-11}), gender (B = 12.1; p= 5x10^{-5}) and the LMP2 SNP rs17587 (B=6.2; p=0.01) were significantly associated with baseline mSASSS (Table 4.3). A forward conditional method of multivariate regression analysis is a method in which the most significantly associated variable is included in the regression equation first. Subsequently variables are added to the equation only if they are significantly contributing to the prediction equation at a significance level that we decide (0.05 or
Table 4.3 Univariate analysis of predictors of baseline radiographic severity

<table>
<thead>
<tr>
<th>Variables</th>
<th>Linear Regression Analysis</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Univariate Analysis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>95% C.I. of B</td>
<td>Beta</td>
<td>Sig.</td>
<td>B</td>
<td>95% C.I. of B</td>
</tr>
<tr>
<td>Duration of disease</td>
<td>0.67</td>
<td>0.50 to 0.85</td>
<td>0.45</td>
<td>&lt;1x10^{-12}</td>
<td>0.74</td>
<td>0.53 to 0.95</td>
</tr>
<tr>
<td>HLA B27</td>
<td>5.21</td>
<td>-0.9 to 11.4</td>
<td>0.12</td>
<td>0.09</td>
<td>12.1</td>
<td>6.4 to 17.8</td>
</tr>
<tr>
<td>Sex</td>
<td>6.28</td>
<td>0.4 to 12.1</td>
<td>0.14</td>
<td>0.027</td>
<td>6.2</td>
<td>1.5 to 10.8</td>
</tr>
<tr>
<td>LMP2: rs17587</td>
<td>5.74</td>
<td>0.7 to 10.7</td>
<td>0.16</td>
<td>0.025</td>
<td>-6.5</td>
<td>-22.3 to 9.4</td>
</tr>
<tr>
<td>LMP7: rs2071543</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.89</td>
<td>-3.1 to 6.9</td>
</tr>
<tr>
<td>ERAP1: rs26618</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.17</td>
<td>-4.4 to 4.7</td>
</tr>
<tr>
<td>rs26653</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.5</td>
<td>-5.8 to 2.9</td>
</tr>
<tr>
<td>rs27044</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-3.63</td>
<td>-7.0 to -0.2</td>
</tr>
<tr>
<td>rs30187</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-5.2</td>
<td>-16.2 to 5.9</td>
</tr>
<tr>
<td>rs374016</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.6</td>
<td>-2.0 to 7.2</td>
</tr>
<tr>
<td>rs10050860</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.6</td>
<td>-6.9 to 3.7</td>
</tr>
<tr>
<td>TAP1: rs1057141</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.05</td>
<td>-1.2 to 7.3</td>
</tr>
<tr>
<td>TAP2: rs241447</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.56</td>
<td>-14.9 to 13.8</td>
</tr>
<tr>
<td>rs1800454</td>
<td>15.2</td>
<td>-9.9 to 40.3</td>
<td>0.09</td>
<td>0.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs2228396</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-22.83</td>
<td>-10.1 to 14.5</td>
</tr>
<tr>
<td>rs4148876</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6.72</td>
<td>-1.10 to 14.5</td>
</tr>
</tbody>
</table>

* Multivariate analysis with forward conditional method, where only significant variables in the regression model are included in the equation.
5.00 % is standard and was used in this case). Increased disease duration, male gender and the presence of the major allele of the LMP2 SNP rs17587 made it more likely to have higher baseline mSASSS scores. Disease duration alone accounted for 20% of the variability in baseline mSASSS. The sequential addition of gender and the rs17587 SNP data independently improved the model to account for 29% and 32% of the variability respectively.

4.3.3 Baseline mSASSS is the strongest predictor of progression

The mean increase in mSASSS scores in the entire cohort was 1.77 ± 2.9 mSASSS units. A minimum gap of 1.5 years between X rays was present in 210 patients who were included in the progression study. The mean increase in mSASSS scores in this cohort was 1.95 ± 3.1 mSASSS units and 60 (29%) patients had progression of mSASSS scores at a minimum rate of 1 mSASSS unit per year during the follow up period.

The patients in the baseline and progression studies were similar in demographics and baseline mSASSS (data not shown). As evident from the cumulative probability plot (Figure 4.2), mSASSS change of 1, 2 and 4 mSASSS units were seen in 98 (46.6%), 84 (40%) and 43 (20.5%) patients respectively. In univariate analysis only baseline mSASSS (OR=1.03; p=0.001) and duration of disease (OR=1.03; p=0.02) predicted progressors (Table 4.4). In multivariate analysis, only the baseline mSASSS was significant in the final model (Table 4.4). The odds of progression with every unit increase in baseline mSASSS, was 1.03 (CI: 1.01 to 1.06; p=0.003). For
Cumulative probability plot showing the progression of mSASSS scores in 210 patients included in the progression study. On the y-axis, the change in x-ray score over the follow up period is shown while the x-axis shows the percentage of patients with the corresponding change. The dotted lines traced from the y-axis correspond to progression of 1, 2 and 4 mSASSS units. The corresponding perpendicular dashed lines to the x-axis indicate the percentage of patients with the respective mSASSS progression. Thus 53% of patients had no change in mSASSS scores and 80% had less than 4 mSASSS units change. Increase in radiographic score of 1, 2 and 4 mSASSS units were seen in 98 (46.6%), 84 (40%) and 43 (20.5%) patients respectively.
### Table 4.4 Predicting patients who are likely to progress on follow up

<table>
<thead>
<tr>
<th>Variables</th>
<th>Univariate Analysis</th>
<th>Multivariate Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Sig.</td>
</tr>
<tr>
<td>Duration of disease</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>mSASSS_baseline</td>
<td>0.03</td>
<td>0.001</td>
</tr>
</tbody>
</table>

B is the co-efficient of the constant in the model (also known as intercept). Beta gives the odds ratio for the variable in the model. Thus a Beta of 1.03 means for every increase of 1 unit in the mSASSS score at baseline, the odds of progression increases by 1.03 times. Thus compared to a patient with an mSASSS score of zero at baseline, patients with scores of 10 and 20 respectively have a 1.3 and 1.8 times increased chance of progressing at a rate of 1 mSASSS unit per year.

Sig. (significance) is the p value for the variable in the equation.

Only duration of disease and mSASSS at baseline were significant in the univariate analysis at p \( \leq 0.1 \) and included in the multivariate analysis.
increasing the threshold to identify progressors, they were defined variably as those with an increase of at least 2 or 4 mSASSS units. Even with this change in definition, baseline mSASSS continued to be the only predictor of progression in both cases.

Receiver operating characteristic (ROC) curve and area under curve (AUC) analysis was done with the logistic regression model obtained by entering baseline mSASSS alone. ROC curves are graphical X-Y plots representing the true positive rate (sensitivity) against the false positive rate (1 – specificity) for a binary outcome. ROC curves can be used to identify the best cut off values to discriminate the two outcomes by plotting the sensitivity and 1-specificity rates for the different cutoffs. We can use the ROC curve to test the value of our statistical model. If the AUC is high (models yielding curves that are shifted to the left) the regression model is good at predicting the outcome (radiographic progression in this case). Naegelkerke $R^2$ is calculated and it indicates the proportion of the variation explained by the regression model.

ROC curve analysis (Figure 4.3) showed that the regression model with baseline mSASSS alone had a modest AUC of 0.68 ($p=1\times10^{-4}$) and the Naegelkerke $R^2$ square estimates showed that baseline mSASSS alone explained only 7% of the variability in progression. Hence there are several important variables yet to be identified that can influence radiographic progression. $R^2$ is the coefficient of determination used in the context of statistical models. The value indicates how well the prediction of future outcomes can be done with the information included in the model. Thus in this case, other factors that explain 93% of the variability are yet to be identified.
The Receiver Operating Characteristic (ROC) curve analysis shows that the AUC for the regression model with baseline mSASSS alone is 0.68 (C.I. 0.60-0.76; p = 1x 10^-4). As the AUC increases, the ROC curve would shift more toward the upper left corner, thereby indicating a model with more variables that predict the outcome. Naegelkerke R square (R^2) which estimates the variability of the outcome (radiographic progression) showed that baseline mSASSS alone explained only 7% of the variability in progression. Hence other important predictor variables for progression are yet to be identified.
4.4 Discussion

The identification of ERAP1 as one of the strongest genetic associations in AS, second to only HLA B27, has triggered renewed interest in studies on the biologic basis of AS. Although a strong genetic association with AS has been established for ERAP1 any influence on the severity of the disease is unknown. This is the first study looking at the effect of ERAP1 and other genes involved in the antigen presentation pathway on radiographic severity and progression in AS patients. The ERAP1 SNP rs30187 significantly predicted the baseline radiographic severity in univariate analysis, but failed to do so in a multivariate model. I found that the LMP2 SNP rs17587, duration of disease and gender are independently associated with baseline radiographic severity of AS patients. The baseline mSASSS was found to be the best predictor of radiographic progression.

It is evident from the initial analysis that the two cohorts are quite distinct. The Edmonton cohort is larger, older and with higher HLA B27 prevalence. Though not statistically significant, there were more progressors in the Toronto cohort and more rapid progressors in the Edmonton cohort. Combining two cohorts with different baseline characteristics can negate the effects of genetic variants on an outcome. It should also be noted that the two cohorts are from two distinct geographic locations in Canada and may be subjected to varying gene-environment interactions further affecting the outcome. Despite these problems, the combined analysis of the cohorts indeed showed remarkable consistency with the cohort specific analysis. The best predictors of baseline radiographic severity in univariate analysis of the combined
cohorts were disease duration and gender. Indeed these same variables have remained important in the two cohorts separately. More over the SNPs which predicted baseline radiographic severity in the combined cohort were rs30187 and rs17587, the same SNPs predicting progressors in the Toronto and Edmonton cohorts respectively. HLA-B27 was the next most important variable in the regression analysis but had a p-value of 0.09 and was included in the multivariate analysis.

For severity and long term damage assessment, mSASSS scoring for x-ray changes is the preferred method. Baseline radiographic severity is a marker of long term damage in patients with AS. Disease duration was the strongest predictor of radiographic severity and by itself accounted for 20% variability in the data. It is known that females tend to have less radiographic severity [308]. This was reflected in this study as well. The gender of the patients explained an additional 9% variability in mSASSS scores over and above disease duration.

After controlling for disease duration, HLA B27 status and gender, the LMP2 SNP rs17587 was independently associated with radiographic severity at baseline. LMP2 has been variably reported to be associated with AS and/or extra articular manifestations of AS especially uveitis [290-295,297,298]. The recent genome wide scans on AS however failed to identify a link to LMP2.

In line with previous reports on predictors of progression, baseline radiographic severity stands out as the best predictor of future progression. With an increase of 1 unit in the mSASSS score at baseline, the odds of progression increased by 1.03 times. Thus compared to a patient with an mSASSS score of zero at baseline, patients with
scores of 10 and 20 respectively have a 1.3 and 1.8 times increased chance of progressing at a rate of 1 mSASSS unit per year.

The rs17587 LMP2 SNP though significantly affecting baseline radiographic severity explained only an additional 3% of variability in the mSASSS scores and it did not significantly affect radiographic progression. It is quite evident that the effect of individual SNPs in radiographic progression is small and a number of genetic factors remain to be discovered [309]. The small effects also mean that larger cohorts and longer follow up is required to see significant influences.

Changes in antigen presentation can affect immune responses and thereby cause inflammation. Inflammation may be associated with new bone formation [180,181,282] but the treatment of AS patients with anti-TNF agents have not consistently shown inhibition of radiographic progression [310-312]. The SNP rs17587 results in a change in the structure of the immunoproteasome Large Multifunctional Peptidase 2 (LMP2) with an R60H substitution. There is no published data on the effect of this change on function of proteasomes. The proteasome, like ERAP1 is important in antigen presentation and is upregulated by IFN-γ [313,314]. Apart from the role in antigen presentation and thereby on inflammation, proteasomes have a well-established direct link to bone metabolism. LMP2 has been shown to regulate NF-κB levels and thus can affect osteoclastic activity [315-318].

Proteasomes are important in osteoblast activity by modulating the Wnt/β Catenin pathway[319]. The Wnt/β Catenin pathway is considered to be overactive in AS due to the low level and/or dysfunctional activity of Dickkopf 1 (DKK-1), a natural
inhibitor of this pathway[188,283]. Blockade of DKK-1 results in fusion of the sacroiliac joints of TNF transgenic mice [320]. When the Wnt signaling pathway is activated, β-catenin, a transcriptional co-activator, accumulates and is translocated to the nucleus[321]. This β-catenin, which is normally degraded by the proteasome, can accumulate in excess when the proteasome is abnormal resulting in excess osteoblastic activity and new bone formation [319,321]. A hypothesis on the role of LMP2 in radiographic progression is shown in figure 4.4. ERAP1 variants can have an indirect effect on NFκB levels by affecting the unfolded protein response. UPR has been shown to affect osteoclast differentiation via NFκB activation [322]. NFκB binding regions are seen in the promoter of ERAP1 helping in regulating ERAP1 expression [323].

In general AS is a slowly progressive disease. It is estimated that within a period of 10 years, radiographic progression can be linear [324]. However this rate of progression varies with follow up. It has been reported that 30% of radiographic progression occurs during the first decade of disease followed by 40% in the next decade and finally 35% in the 3rd decade [324]. Two things are quite evident here. Studies of radiographic progression ideally should have long follow up, and depending on the duration of disease, the rate of radiographic progression may vary. A larger cohort followed over longer periods of time could help us identify additional factors with subtler effects on radiological progression in AS. Studies including haplotype analysis and gene-gene interactions with other AS-associated genes like IL-23R are warranted.
Abnormal proteasome LMP2 variant could affect osteoclast function by affecting NF-kB levels. There is no published data on the effect of the LMP2 variant studied (R60H) on the function of proteasomes. LMP2 variants can decrease degradation of β-catenin in osteoblasts and result in osteoproliferation. The increased levels of β-catenin in turn can decrease the expression of RANKL on osteoblasts and indirectly further decrease osteoclast function.

LMP2: Large Multifunctional Peptidase 2; NF-kB: Nuclear Factor kappa B; RANK: Receptor activator of NF-kB; RANKL: Receptor activator of NF-kB Ligand; LRP: LDL Receptor Protein 5
4.5 Conclusions

This is the first study to demonstrate that LMP2 variants can affect radiographic severity in AS. ERAP1 could affect radiographic severity, but larger studies are needed to explore this further. Baseline mSASSS remains the strongest predictor of radiographic progression in AS but explains only a fraction of the variability seen.
Chapter 5

5 Conclusions and Future Directions

5.1 Conclusions

The body of work presented here is an attempt at understanding the functional and clinical relevance of the recently reported association between ERAP1 and ankylosing spondylitis. The results of the functional studies are the first report of a functional implication of \textit{ERAP1} polymorphisms in patients with AS. It is also the first attempt at examining at the effect of \textit{ERAP1} polymorphisms on radiographic severity of AS.

I have demonstrated that the \textit{rs27044} coding polymorphisms in \textit{ERAP1} which leads to a Q730E amino acid change results in significant variation of the FHC expression on the surface of monocytes. AS patients with only the minor allele of this SNP have significantly lower surface FHC compared to those with the major allele. There was no significant effect on the B cell FHC expression. The total HLA-B27 expression in the whole PBMC population was lower in patients with the minor allele, but this was not statistically significant. There was no effect of the \textit{rs30187} SNP on FHC or total HLA-B27.

The \textit{ERAP1} SNPs studied here did not show any significant effect on the serum cytokine or cytokine receptor levels for IL-1, IL-6 and TNFα. Thus the cytokine receptor shedding function of ERAP1 appears not to be a significant contributor in the pathogenesis of AS.
Suppression of ERAP1 levels by siRNA can affect both the conformation of HLA-B27 and the peptide display on HLA-B27. Furthermore, these properties appear to be differentially affected in cells expressing AS-associated HLA-B27 subtypes as opposed to those not associated with AS.

The ERAP1-B*27 interaction could be the missing link in the pathogenesis of AS. The recent evidence emerging on a genetic interaction between ERAP1 and HLA-B27 is additional evidence of a possible functional interaction and could be pathogenic in AS. The ERAP1 association with AS is seen only in HLA-B27 positive AS patients. Similarly, in psoriasis ERAP1 association is seen only in patients with the HLA-C risk allele. The association between MHC-I and ERAP1 seen in genetic studies strengthens the evidence of functional interaction that I have shown. Patients who had the Q730 variant of ERAP1, that has been shown to have higher aminopeptidase activity than the E730 variant, had lower FHC. Hence it seems that ERAP1 functioning results in properly formed MHC-I molecules with less FHC formation. It is evident from the population studies that the common variant of ERAP1 (E730) seen in the majority of normal population and AS patients has less aminopeptidase activity than the minor variant (Q730). A gain of function would thus lead to more MHC-peptide complexes being formed. Putting all the observations together, it seems the pathogenic event is a gain of function with resulting increase in MHC-peptide complexes.

The FHC hypothesis cannot be entirely excluded, although a higher FHC expression would be expected in controls compared to AS patients. FHC can form dimers that are recognized by inhibitory Killer Immunoglobulin receptors (KIR).
KIR3DL1 and KIR3DL2 [167]. Thus a loss of function could be protective by increased inhibitory signals through the KIRs. The β2m overexpression study is a great example of the complexity of AS pathogenesis. If FHC are pathogenic, by stabilizing HLA-B27, disease should be ameliorated. HLA-B27 was stabilized by β2m overexpression in HLA-B27 transgenic rats which paradoxically resulted in the arthritis worsening and colitis improving [174]. Worsening of arthritis with minimal effect on colitis was seen by another experiment that resulted in increased MHC-peptide complexes. A B27-specific influenza peptide was overexpressed in HLA-B27 transgenic rats which led to more arthritis but no effect on colitis [175]. This would suggest that the pathogenesis of arthritis is peptide mediated while that of colitis could be due to the UPR.

The discovery of the ERAP1-AS association has generated a lot of interest in the pathogenesis of AS. With high throughput studies and an explosion of molecular biology and other translational research techniques, we are closer to the answer than ever before.

5.2 Future Directions

The results obtained and elaborated in previous chapters raise several questions. ERAP1 can potentially influence FHC as well as HLA-B27-peptide complexes recognized by MARB4 antibody. It needs to be further explored as to how this interaction could be potentially pathogenic in AS.
I am planning to take this study to the next step and will attempt to answer the interesting questions raised here. I hope this attempt will help take our understanding of the pathogenesis of AS further.

5.2.1 Studies in Ankylosing Spondylitis Patients

5.2.1.1 ERAP1 variants lead to changes in surface FHC. Does this lead to altered KIR binding?

To address this question, peripheral blood mononuclear cells isolated from B27-positive AS patients will be stained with recombinant human Fc Chimeras (R&D systems) of KIR3DL1 and KIR3DL2 (both known to bind B27 FHC dimers). The secondary antibody FITC-anti human IgG (Fc gamma specific) will be used and KIR binding assessed by mean fluorescence intensity (MFI) values in flow cytometry (FACS). The antigen presenting cells monocytes and B cells will be separately gated with anti-CD14 PE and anti-CD-19 APC. Patients will be genotyped using the allelic discrimination assay for the rs30187 and rs27044 SNPs of ERAP1. Differences in KIR binding between patients with different ERAP1 variants will be assessed.

5.2.1.2 Does ERAP1 polymorphism affect NK cell function?

Monocytes will be isolated using the antiCD-14 MACS separation system (Miltenyl Biotech, Germany) from AS patients genotyped for rs27044 and rs30187 ERAP1 SNPs. The 7-AAD/CFSE assay (Cayman Chemicals, USA) will be used to study NK cell cytotoxicity. Briefly, monocytes will be co-incubated with YTS cells (NK cell line; from Dr. Deborah Burshtyn: collaboration letter appended) expressing KIR3DL1, which are known to bind B27 dimers. Different effector:target ratios will be used with the
monocytes pre-incubated with CFSE. The dye 7-AAD will be used for staining the dead
target cells and quantification will be done by FACS. It is expected that the isolated
monocytes expressing B27 dimers can affect NK cell activity of the YTS cells through
the KIR receptors. In parallel, the co-incubated cells will separately be analyzed for
cytokine production by intracellular staining for IFN-\(\gamma\), TNF-\(\alpha\) and IL-10 followed by
FACS. Thus NK cell cytotoxicity and cytokine production will be compared between
patients with different variants of ERAP1.

5.2.2 In vitro studies

The following system will be established for the in vitro studies. C1R cells stably
transfected with HLA B*2704, 05, 06 and 09 will be used to establish cell lines with
ERAP1 stably knocked down using HIVH1 shRNA clone set targeting human ERAP1
(Puromycin selective, with fluorescent mCherry reporter and all targets at 3'UTR
region, in an HIV based lentiviral vector). Once a stable knock down of ERAP1 is
established and confirmed by western blot analysis, one of the following will be
introduced into the respective C1R cells: either wild type human ERAP1 ORF clone or
one of the human ERAP1 mutant clones, rs30187: amino acid 528 K->R (AAG->AGG)
or rs27044: 730 Q->E (CAA->GAA) in pReceiver-Lv184 (HIV-based lentiviral vector
with CMV promoter, C-3xHA tag and hygromycin selection marker). As the shRNA is
against the 3'UTR, there should not be a problem in expressing ORF cDNA based
ERAP1. The antibiotic selection is based on the fact that C1R cells have Geneticin
selection incorporated already. Negative control vector for pReceiver-Lv184 and HIVH1
shRNA scrambled control clone for psiHIV-H1 will be used as controls.
5.2.2.1 Does ERAP1 polymorphism lead to altered unfolded protein response?

Once the C1R cell lines with different B27 subtypes, each stably expressing the wild type or variant ERAP1 forms are established, the UPR will be assessed by a PCR array. The Human UPR RT² Profiler™ PCR Array (SABiosciences, USA) profiles the expression of 84 key genes recognizing and responding to misfolded protein accumulation in the endoplasmic reticulum (ER). Significant differences noted in the array, between cells expressing different allelic forms of ERAP1 will be confirmed by western blot.

5.2.2.2 Does ERAP1 polymorphism lead to changes in the MHC-peptide repertoire?

Peptide repertoire variation, if any, in humans in the context of HLA B27 and different allelic forms of ERAP1 is important considering the arthritogenic-peptide hypothesis. Here, the MHC-peptide repertoire of HLA B27 expressing C1R cells with different ERAP1 variants will be studied.

Briefly, the C1R cells with different B27 subtypes, each expressing the AS-associated and wild type ERAP1 variant will be subjected to MHC-peptide isolation by the acid elution technique. Cell suspensions will be pelleted and the supernatant isolated for ultrafiltration. Analysis of the peptidome will be done by mass spectrometry using a LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific).
5.2.2.3 Does the presence of different ERAP1 variants affect NK cell cytotoxicity and activation? Does this differ depending on the B27 subtype expressed?

Similar to aforementioned experiments in question II, the C1R cells will be co-incubated with YTS cells in different effector:target ratios and NK cell cytotoxicity and activation (cytokine production) will be studied. These in vitro studies will corroborate studies in AS patients and will also help understand if the NK-cell effect varies with the B27 subtype expressed on the C1R cell.


27. Roussou E, Sultana S. Spondyloarthritis in women: differences in disease onset, clinical presentation, and Bath Ankylosing Spondylitis Disease Activity and Functional


77. Fernandez-de-Las-Penas C, Alonso-Blanco C, Alguacil-Diego IM, Miangolarra-Page JC. One-year follow-up of two exercise interventions for the management of


93. van Denderen JC, van der Paardt M, Nurmohamed MT, de Ryck YM, Dijkmans BA, van der Horst-Bruinsma IE. Double blind, randomised, placebo controlled study of


147. Hammer RE, Maika SD, Richardson JA, Tang JP, Taurog JD. Spontaneous inflammatory disease in transgenic rats expressing HLA-B27 and human beta 2m: an


152. Madden DR, Gorga JC, Strominger JL, Wiley DC. The three-dimensional structure of HLA-B27 at 2.1 A resolution suggests a general mechanism for tight peptide binding to MHC. Cell 1992 Sep 18;70(6):1035-1048.


252. Mehta AM, Jordanova ES, Corver WE, van Wezel T, Uh HW, Kenter GG, et al. Single nucleotide polymorphisms in antigen processing machinery component ERAP1


276. Chan Y, Walmsley RP. Learning and understanding the Kruskal-Wallis one-way analysis-of-variance-by-ranks test for differences among three or more independent groups. Physical therapy 1997 Dec;77(12):1755-1762.


318. Hayashi T, Faustman D. NOD mice are defective in proteasome production and activation of NF-kappaB. Molecular and cellular biology 1999 Dec;19(12):8646-8659.


