University of Alberta

Exploration of methods for sequence based HLA typing and application to patients with hair dye allergy

by

Carlos Ricardo Garcia-Batres

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Immunology

Medical Microbiology and Immunology

Fall 2011
Edmonton, Alberta

I dedicate this thesis to my parents, whose sacrifices allowed me to excel

ABSTRACT

Determining an individual's spectrum of human leukocyte antigens (i.e. HLA type) by sequencing their HLA genes is known as sequenced-based typing (SBT). In the first part of this thesis, non-polymerase chain reaction (PCR) methods of achieving SBT were explored. HLA-DQ cDNA sequences were cloned into a novel M13 vector so that antisense sequence was displayed, and attempts were made to specifically capture the DQ sequences using two strategies: 1) complementary oligonucleotides covalently bound to the surface of beads, and 2) complementary biotinylated oligonucleotides which were then captured using streptavadin-beads. Only the second protocol was successful, enabling enrichment of sequences about 100-fold, but the method was deemed unsuitable to achieve SBT. In the second part of this thesis an RT-PCR method to amplify the entire coding region of HLA-DPB1 transcripts was developed. By cloning and sequencing the amplified fragments, the HLA-DPB1 type could be assigned. Methods were also developed to extract total cellular RNA from human blood samples, and RNA samples from 16 hair dye-allergic patients were obtained. These were subjected to RT-PCR and HLA-DPB1 alleles assigned. The allelic distribution of the patient samples did not differ from a control population, suggesting that hair dye allergy is not associated with certain HLA-DPB1 alleles.

Acknowledgements

I would like to first thank Dr. John Elliott, whose guidance and support challenged me and made me an overall better scientist. I would like to thank the members of my supervisory committee, Dr. David Evans and Dr. Katharine Magor, for helping to guide my research project. Special thanks to Dr. Deborah Burshtyn, for taking part in my Master's examining committee.

I would like to thank my friend Doug Quilty for his friendship over the past several years; I knew I could always depend on him to help at work and at home with my family. I would also like to thank my friends in the MMI department, notably my friends in the Barry Lab (Logan Banadyga, Nick van Buuren, Stephanie Campbell and Kristin Burles) who welcomed me since day one and who I have developed life-long friendships with.

I would like to thank my parents and my siblings, who have always supported me and encouraged me to be my best at whatever I do. Their sacrifices have allowed me to accomplish my goals.

Finally, I would like to thank my wife Katie whose friendship and love have enabled me to become a better person in every aspect of my life. Over the course of this thesis, she has blessed our home with two lovely children (Naiya and Cahlen). She is my strength and has always been there to make me laugh and has stood by me helping me throughout my graduate career. This work would not be possible without her help.

TABLE OF CONTENTS

1. INTRODUCTION/BACKGROUND

1.1 The Immune System 1

1.2 Major Histocompatibility Complex 2

1.2.1 HLA Class I molecules 4

1.2.2 HLA Class II molecules 4

1.3 Human Leukocyte Antigen typing 6

1.3.1 Serological HLA typing 7

1.3.2 Sequence based typing 8

1.3.2.1 Polymerase Chain Reaction – Sequence Specific Primers (SSP) 8

1.3.2.2 Polymerase Chain Reaction – Sequence Specific Oligo Probes (SSOP) 8

1.3.2.3 Sequence Based Typing (SBT) 9

1.4 Human leukocyte antigens and disease 11

1.4.1 Akylosing spondylitis – HLA-B27 12

1.4.2 Insulin dependent diabetes mellitus (IDDM) 12

1.4.3 Multiple sclerosis 14

1.4.4 Berylliosis 16

1.4.5 HLA and allergic contact dermatitis 17

1.5 Sequence-based HLA typing requires enrichment of specific desired DNA molecules from complex mixtures 19

1.5.1 Cloning methods to isolate desired DNA molecules from complex mixtures 20

1.5.2 M13 bacteriophage cloning vector 26

1.6 Objectives and project overview 28

1.7 References 30

2. MATERIALS AND METHODS

1.1 Bacterial strains, media, and culture conditions 34

1.2 Preparation of transformation competent cells 35

2.2.1 Chemical competent cells 35

2.2.2 Chemical transformation 36

2.2.3 Electrocompetent cells 36

2.2.4 Electroporation 37

1.3 Protocols for M13 Bacteriophage 38

1.3.1 Preparation of large bacteriophage cultures 39
2.3.2 Extraction of single stranded phage DNA 42
2.3.3 Purification of replicative form (RF) phage DNA 44
2.3.4 Cesium chloride gradients 46
1.4 Eukaryotic cell culture 49
1.5 Purification of total cellular RNA using guanidinium thiocyanate-phenol-chloroform extraction 50
1.5.1 Extraction of total cellular RNA from small numbers of cells 51
1.5.2 Extraction of total cellular RNA from large number of cells 52
1.6 Extraction of cytoplasmic RNA from cell lines using NP40 lysis 54
1.7 Extraction of total cellular RNA from human blood 58
1.8 References 63

3. MODEL EXPERIMENTS TO SPECIFICALLY CAPTURE M13/HLA-DQ CDNA CONSTRUCTS DIRECTLY OUT OF SOLUTION
1.1 Introduction 64
1.2 Two approaches to capture specific DNA sequences directly out of solution 65
1.3 Identification of potential binding sites/sequences for capture oligonucleotides for DQA1 and DQB1 genes 70
1.3.1 Site 211B is present in nearly all MHC class II beta genes 71
1.4 Subcloning DQA1 and DQB1 cDNA fragments into an M13 vector 73
1.4.1 Subcloning using the gel slice ligation method 79
1.5 Production of capture oligonucleotide 83
1.6 Mode experiment to capture M13JFE7-DQ8B SSC-DNA onto OAS-211B beads 87
1.6.1 Model experiment to capture a soluble complementary oligonucleotide onto OAS-211B beads 90
1.7 Model experiments to capture specific SSC-DNAs from solution using biotinylated oligonucleotides and streptavidin beads 94
1.7.1 Preventing non-specific binding of SSC-DNA to streptavidin dynabeads 101

1.7.2 Preventing non-specific binding of SSC-DNA to streptavidin dynabeads 103
3.8 References 109

4. SEQUENCE BASED HLA-DPB1 TYPING RESULTS: MODEL EXPERIMENTS USING RT-PCR AND CLONING, AND APPLICATION TO HAIR DYE ALLERGIC PATIENTS
4.1 Introduction 110
4.2 cDNA synthesis 114
4.3 PCR reactions 116
4.3.1 Cloning into pCSD7s vector 118
4.4 Sample preparation for sequencing 120
4.5 TOPO TA cloning method 121
4.6 Assigning HLA allele classification 123
4.6.1 Allele assignment to homozygous B-cell lines 124
4.6.2 Allele assignment to diploid samples 125
4.7 Allergic contact dermatitis: Patient study 126
4.7.1 Patient blood samples 126
4.7.2 Sequencing patient samples 128
4.7.3 Patient sequence analysis 129
4.8 References 135

5. GENERAL CONCLUSIONS AND DISCUSSION
5.1 Summary of results 136
5.2 Solution hybridization 136
5.3 RNA extractions 137
5.4 HLA and disease 140
5.5 HLA-DPB1 and acute contact dermatitis 140
5.6 References 145

APPENDIX
Supplemental figures 146

LIST OF TABLES AND FIGURES

CHAPTER 1: Introduction/Background

Figure 1.1 The Major Histocompatibility Complex in humans 3
Figure 1.2 Sub-unit structure of the MHC class I and class II molecules 5
Figure 1.3 Screening a plasmid library of clones by filter hybridization 21
Figure 1.4 Double stranded compared to single stranded circular DNA molecules 24
Figure 1.5 Overview of GeneTrapper cDNA positive selection system 25
Figure 1.6 M13 Bacteriophage lifecycle 27

Table 1.1 Diseases and human leukocyte antigens association 13

CHAPTER 2: Materials and Methods

Figure 2.1 DNA obtained from large-scale M13mp19 bacteriophage cultures, analyzed by agarose gel electrophoresis 41
Figure 2.2 Total RNA extracted from WT49 cells and analyzed using the Experion automatic gel electrophoresis system 55
Figure 2.3 Electropherograms of total RNA compared to cytoplasmic RNA extracted from WT 49 cells 56
Figure 2.4 Electropherograms of total RNA extracted from two different human blood samples 61
Figure 2.5 Electropherograms of total RNA from two patient blood samples, showing varying quality of RNA obtained 62

CHAPTER 3: Solution based hybridization model experiments

Figure 3.1 Overview of proposed procedure to capture specific M13 target sequences directly out of solution using oligo affinity supports (OAS) beads 66
Figure 3.2 Overview of proposed procedure to capture specific M13 target sequences directly out of soltion using biotinylated oligonucleotides and ferro-magnetic streptavidin beads 67
Figure 3.3 Site 211B nucleotide sequence is highly conserved among HLA class II beta chain genes, and the corresponding

(highly conserved) peptide is located in the floor of peptide-binding cleft interest in HLA beta-chain genes 74
Figure 3.4 DNA sequence of M13JFE7s polylinker 77
Figure 3.5 Schematic diagram showing details of Sfi I 'sticky ends' used in directional cloning strategy and sub-cloning of HLA-DQ cDNA inserts from the plasmid vector pCSD7 into M13JFE7s ligations 78
Figure 3.6 Low melting point agarose gel showing the various DNA bands which were excised and used to set up the gel slice ligation reactions 81
Figure 3.7 OAS-211B beads failed to capture detectable amounts of M13JFE7-DQ8B SSC-DNA 89
Figure 3.8 Schematic diagram of experiment to test capture of a soluble complimentary oligonucleotide by OAS-211B beads 91
Figure 3.9 Results of model experiments to capture M13JFE7-DQ6A constructs out of solution using biotinylated oligonucleotides and streptavadin beads 99
Figure 3.10 Results of model experiments to capture M13JFE7-DQ8A constructs out of solution using biotinylated oligonucleotides (BTN-211B) and streptavadin beads 102

Table 3.1 Conserved sequences within HLA-DQA1 and DQB1 with potential to be target sites for capture oligonucleotides 72
Table 3.2 List of nucleotide polymorphisms and amino acid substitutions in site 211B, found by searching all known HLA-DRB1, -DQB1, and -DPB1 DNA sequences 75
Table 3.3 Plasmid and M13 vectors and derived DQ cDNA constructs 84
Table 3.4 Oligonucleotides used for DNA sequencing and model solution hybridization/capture experiments 85
Table 3.5 Radioactivity present in various fractions after incubating a ^{32}P-labelled soluble complementary oligonucleotide with two different OAS beads 95
Table 3.6 Blue/White pseudoplaque counts after one cycle of selection 105

CHAPTER 4: Sequence based typing results: model experiments

Figure 4.1 cDNA fragments generated by PCR using the DPB1 and Sfi-DPB1 primer pairs 117

Figure 4.2 DPB1 cDNA fragments generated by RT-PCR using reverse transcribed patient RNA, and representative cloned DPB1 cDNA fragments excised from pCSD7 by Sfi I digestion 119

Table 4.1 Oligonucleotides used to PCR amplify and then add Sfi I sites to HLA-DPB1 cDNA fragments 112

Table 4.2 Numbers of point mutations observed in cloned DPB1 RT-PCR fragments amplified using different thermostable DNA polymerases 113

Table 4.3 Clones with inserts and assigned alleles in patient study 122

Table 4.4 HLA transcripts encoded within pCR 2.1 TOPO vector assembled in this study 127

Table 4.5 Allele assignments for the patient samples with ACD to PPD 130

Table 4.6 Allelic distributions of patient study in this study and patients with alleles 132

Table 4.7 Allelic distribution of control population and patients in this study 134

Figure A.1 Haplotype crossover 147

LIST OF ABBREVIATIONS

β2M	Beta2-microglobulin
γ-^{32}P-ATP	^{32}P labeled – gamma phosphate on adenosine triphosphate
mg	Microgram
ml	Microliter
mM	Micromolar
ACD	Allergic contact dermatitis
AGC	Applied Genomics Center
AMV	Avian myeloblastosis virus
AS	Ankylosing spondylitis
ATP	Adenosine triphosphate
B-cells	B lymphocytes
BAC	Bacterial artificial chromosome
BLASTn	Basic local alignment search tool - nucleotide
B-oligo	Biotinylated-oligonucleotide
bp	Base pairs
BW1 (or 2) buffer	Binding and washing 1 (or 2) buffer
CBD	Chronic beryllium disease
CDC assay	Complement-dependent cytotoxicity assay
cDNA	Complementary deoxyribonucleic acid
CFU	Colony forming units
CNS	Central nervous system
CsCl/EtBr	Cesium chloride/Ethidium Bromide
DEPC	Diethylpyrocarbonate
dCTP	Deoxycytidine triphosphate
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphates
dsDNA	Double stranded deoxyribonucleic acid
eAMV-RT	Enhanced avian myeloblastosis virus reverse transcriptase
EBV	Epstein Barr virus
E. coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
EL buffer	Erythrocyte lysis buffer
ELISA	Enzyme-linked immunosorbent assay

Exo III	Exonuclease III
FW solution	Filter washing solution
g	Gram
GSC	Genome Sciences Center
HERO	Human Ethics Research Online
h	Hour
HLA	Human leukocyte antigen
IDDM	Insulin dependent diabetes mellitus
IPPD	N-isopropyl-N'-phenyl-p-phenylenediamine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	Kilobases
KIR	Killer cell immunoglobulin receptors
kV	Kilovolts
L	Liter
LB	Luria-Bertani broth
LMP	Low melting point
LS	Liquid scintillation
LT-α -β	Lymphotoxin-alpha, -beta
M	Molar
Mb	Megabases
mg	Milligram
MHC	Major histocompatibility complex
min	Minute
ml	Milliliters
mM	Millimolar
mRNA	Messenger ribonucleic acid
MS	Multiple sclerosis
NCBI	National Center for Biotechnology Information
ng	Nanogram
nm	Nanometer
OAS	Oligonucleotide affinity supports
OD	Optical density
oligo	Oligonucleotide
oligo-dT	Oligonucleotide deoxythymidine
PBMC	Peripheral blood mononuclear cells
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol

PFU	Plaque forming units
pmol	Picomole
PP	Polypropylene
PPD	Para-phenylenediamine
RCF	Relative centrifugal force
RF	Replicative form
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	Revolutions per minute
RPMI	Roswel Park Memorial Institute
RT-PCR	Reverse-transcriptase polymerase chain reaction
SBT	Sequence based typing
SDS	Sodium dodecyl sulfate
SNP	Single nucleotide polymorphism
ssDNA	Single stranded deoxyribonucleic acid
sscDNA (SSC-DNA)	Single stranded circular deoxyribonucleic acid
SSOP	Sequence specific oligo probes
SSP	Sequence specific primers
SSPE	Saline sodium phosphate EDTA
T-cells	T lymphocytes
T_H cells	T helper cells
T_{reg} cells	T regulatory cells
TBE buffer	Tris Base Boric acid EDTA buffer
TCA	Trichloroacetic acid
TCC	T-cell clone
TdT	Terminal deoxynucleotidyl transferase
TE buffer	Tris-HCl EDTA buffer
TNF-α, -β, C	Tumor necrosis factor alpha, beta, c
tRNA	Transfer ribonucleic acid
UV	Ultraviolet
X-GAL	bromo-chloro-indolyl-galactopyranoside
YT-MT medium	Yeast tryptone – magnesium chloride Tris medium

CHAPTER 1

INTRODUCTION/BACKGROUND

1.1 THE IMMUNE SYSTEM

The immune system consists of an elaborate composition of specialized cells, tissues and organs that are used by the body to defend itself against bacterial, viral, and fungal infections as well as tumor cell growth [1]. There are two key responses of the immune system: the innate immune response and the adaptive immune response. The innate immune response is the primary line of defense against foreign pathogens. The fast response time of the innate immune system restrains oncoming infections to provide time for the stronger and more specific adaptive immune response, including antibody production and cell mediated immunity, to develop [1].

The adaptive immune response is the antigen-specific response directed by pathogen specific lymphocytes [1, 2]. This response is not only stronger and more direct than the innate immune response; it also induces the development of immunological memory. Antigen presentation is how the adaptive immune response is activated. Immune cells process pathogenic proteins, and present antigenic peptides to T-cells to activate the T-cell response [2, 3, 4]. The molecules responsible for antigen presentation are encoded within the major histocompatibility complex (MHC).

1.2. MAJOR HISTOCOMPATIBILITY COMPLEX

The major histocompatibility complex (MHC) is a large genomic locus of 3.6Mb located on Chromosome 6 (6p21.3) in humans [5, 6] (Figure 1.1). The MHC locus can be subdivided into three classes: The MHC class III subgroup encodes several proteins that play major roles in innate immunity including the complement proteins, tumor necrosis factor α (TNF-α), lymphotoxin-alpha (LT-α, also known as TNF-β), lymphotoxin-beta (LT-β, also known as TNFC), and other important immune components (Figure 1.1) [7]. The MHC class I and the MHC class II regions contain the MHC molecules responsible for antigen presentation (Figure 1.1) [8].

The MHC molecules in humans are known as the human leukocyte antigens (HLA). The HLA region is the most polymorphic region in the human genome [9]. This is due to the large variability of alleles in the HLA molecules. To date, there have been a total of 6403 different class I and class II alleles reported, and these numbers increase steadily fairly regularly [9, 10]. The HLA polymorphisms are mostly located in the sites affecting antigenic peptide binding to the MHC molecule, but can also affect the contacts between a T-cell receptor and the MHC molecule or between other signaling proteins such as KIR proteins [11].

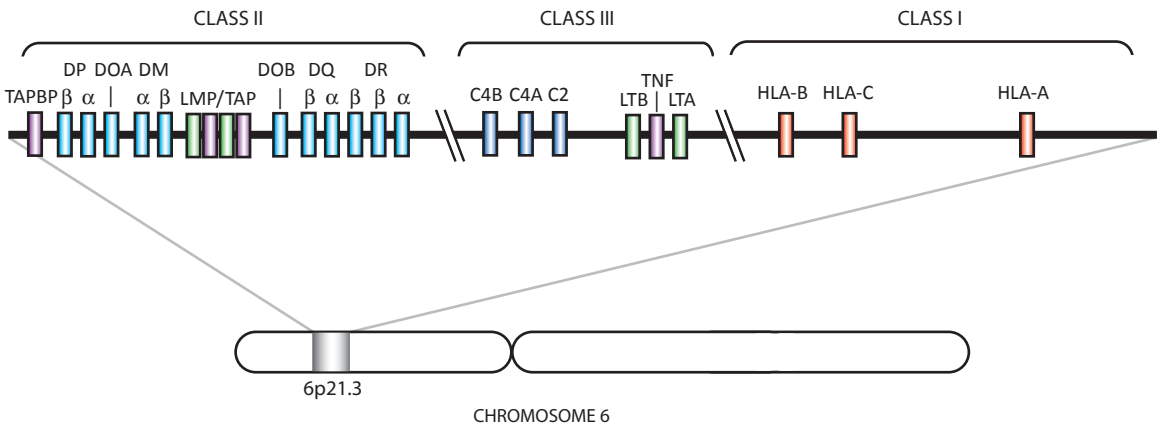


Figure 1.1 The major histocompatibility complex (MHC) in humans. The human MHC is located at 6p21.3, and can be divided into three regions: Class I, Class II and Class III. Certain molecules encoded in class I and II play a key role in antigen presentation, including HLA-A, HLA-B and HLA-C of class I, and HLA-DQα/DQβ, HLA-DRα/DRβ, and HLA-DPα /DPβ of class II. The proteins encoded in the class III region are also important for normal immune responses, but they play no direct role in antigen presentation.

1.2.1 HLA CLASS I MOLECULES

The HLA class I molecules are expressed in nearly all nucleated cells of the body [8]. These molecules present cytosol-derived peptides to CD8+ cytotoxic T-cells. The antigen presented to the T-cells can be self or foreign in origin since antigen processing through the immunoproteosome is indiscriminate [7, 13]. If the antigen presented is identified by an immature T-cell as foreign, the T-cell is activated, initiating proliferation and differentiation into an effector cytotoxic T-cell.

There are three classical HLA class I genes that form the class I receptor: HLA-A, HLA-B and HLA-C. Each MHC class I receptor is a heterodimer composed of two different proteins, the α-chain encoded within the MHC class I genes, and the beta-2-microglobulin (β2M) encoded within chromosome 15 in humans (Figure 1.2) [5]. The α-chain is polymorphic, while the β2M is non-polymorphic. Class I proteins (alpha chain) contain their highly polymorphic sites within the alpha-1 and alpha-2 domains. These two domains form the peptide-binding cleft that holds the antigen for presentation (Figure 1.2) [5].

1.2.2 HLA CLASS II MOLECULES

The HLA class II molecules present antigenic peptides generated in intracellular vesicles to CD4+ T-cells. Unlike class I molecules, MHC class II molecules are only expressed by a subset of cell types known as professional antigen presenting cells, which includes dendritic cells, macrophages and

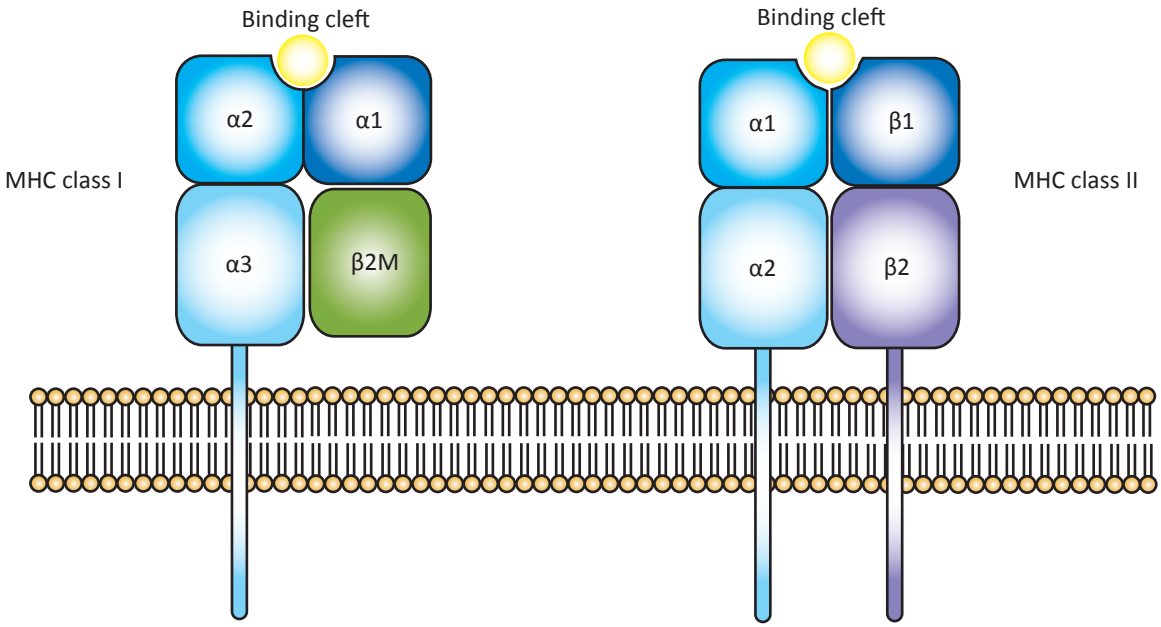


Figure 1.2 Sub-unit structure of the MHC class I and class II molecules. The MHC class I molecules (i.e. HLA-A, -B, and -C in humans) are heterodimeric proteins composed of an alpha-chain paired with beta-2-microglobulin (β2M). The alpha-chain is folded into three domains (α1, α2 and α3), and the α1 and α2 domains make up the binding cleft that holds the antigenic peptide (typically 9 amino acids long) which is presented to CD8+ T cells. The α3 domain binds non-covalently to β2M to stabilize the molecule. The α chain also contains a transmembrane segment and a cytoplasmic domain used in signal transduction. Nearly all of the allelic polymorphisms found in MHC class I molecules are found within exons II and III of the alpha chain gene, which encode the α1 and α2 domains respectively. MHC class II molecules (DQ, DR, and DP in humans) are also heterodimers, consisting of two membrane anchored proteins (alpha and beta chains). Each chain is folded into two domains, yielding α1/α2 and β1/β2. The α1 and β1 domains form the peptide-binding cleft that holds antigenic peptides (typically ≥ 12 amino acids long) which are presented to CD4+ T-cells. Virtually all of the allelic polymorphisms found in MHC class II molecules are found in the α1 and β1 domains, which are encoded primarily by exon II of the respective genes.

B lymphocytes (B-cells) [5]. There are six classical HLA class II genes: HLA-DQA1, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DPA1 and HLA-DPB1. Each HLA class II molecule is a heterodimeric protein composed of two transmembrane glycoprotein chains, designated the alpha and beta chains, which are in turn each folded into two key domains, α1/α2 and β1/β2. The α1 domain together with the β1 domain forms the peptide-binding groove that holds peptides of at least 13 amino acids in length and are presented to CD4+ T-cells (Figure 1.2).

Activation of CD4+ T-cells via antigen presentation leads to proliferation and differentiation into T-helper (T_H) cells. Cytokines released by the activated T_H cells determine their function: T-helper 1 (T_H1) cells promote cell-mediated responses (e.g. macrophage activation), whereas T-helper 2 (T_H2) cells promote B-lymphocyte mediated humoral responses. Other subsets of T_H cells include T_H17 cells that excrete interleukin 17, and T regulatory (T_{reg}) cells that function to regulate immune responses [2, 14].

1.3 HUMAN LEUKOCYTE ANTIGEN TYPING

The Human Leukocyte antigens (HLA) were first discovered in the late 1950s through the observation of antigenic variations between white blood cells obtained from different individuals [15]. Several different methods for HLA typing have been developed since the discovery of the HLA genes. These methods have played a direct role in identifying the unique characteristics of the HLA locus. Existing methods for determining HLA type can be divided into two

main groups: serological typing, which focuses on differences in the HLA proteins expressed in blood or tissue samples, and sequence based typing, which focuses on differences in the DNA sequence of the corresponding genes [16].

Both groups of techniques have their advantages and disadvantages, which is why both techniques are still in use today. However, DNA typing has become more routinely used due to advances in molecular biology in the past two decades.

1.3.1 SEROLOGICAL HLA TYPING

HLA serotyping methods are based on the detection of HLA molecules expressed on the surface of isolated T-cells (class I) and B-cells (Class II) using varying panels of antisera. The typing resolution of these methods is low, yet serological typing is extremely useful for determining whether or not a specific HLA allele is expressed on the cell surface [17].

Serotyping is a common method used in transplantation medicine due to the fast turnover of results and direct benefits on a clinical level. Serotyping can be accomplished by HLA antibody screening by complement-dependent cytotoxicity (CDC) assay, HLA antibody screening by enzyme-linked immunosorbent assay (ELISA), or HLA antibody screening by flow cytometry/X-map luminex techniques [18]. The results obtained from these assays are compared to known alleles to directly identify the alleles in question.

1.3.2 SEQUENCE BASED TYPING

Sequence-based HLA typing does not require live lymphocytes, making it a more favorable method for most, especially in instances where repeated testing of the sample is required. There are several protocols for sequence based typing, but all are based on polymerase chain reaction (PCR) amplification of HLA gene sequences. Sequence based typing has several benefits including rapid turnaround (at least for some protocols) and the ability to design PCR primers in-house and to purchase commercially synthesized oligonucleotides at a low cost. Below are some of the most prominent sequence based typing methods used:

1.3.2.1 Polymerase Chain Reaction – Sequence Specific Primers (SSP): The use of sequence specific primers to identify a particular HLA allele or allele group is fairly common [19]. The assignment of a specific allele is based on the presence of an expected PCR product as determined via gel electrophoresis analysis [20]. For example, a primer that is allele-specific will only produce a PCR product in those patient genomic samples that have the specific single nucleotide polymorphisms (SNPs) which are complimented in the primer. Low resolution typing is generally accomplished by this method, but secondary PCR reactions can be used to achieve higher resolution typing [21]. This procedure is labor intensive and is not suited for large numbers of samples.

1.3.2.2 Polymerase Chain Reaction – Sequence Specific Oligo Probes (SSOP): This method uses short (13-15bp) oligonucleotide (oligo) probes that are labeled with

γ-^{32}P-ATP or other detectable markers (e.g. biotin). The probes are hybridized with the PCR product to analyze if a particular sequence is present in the amplified sample [22]. Only those probes that have a 100% sequence match will hybridize with the PCR product. The binding (or not) of the labeled probes allows for allele assignment. Multiple oligo probes are required per locus in order to assign the allele (e.g. in a recent kit, HLA-C required 64 different probes) [23]. A recent variant of this technique uses fluorescently labeled microbeads in a X-Map tech flow cytometer (luminex) [24]. This approach is now being used increasingly for clinical HLA typing, and is ideal for processing fairly large numbers of samples. Since the luminex technology can simultaneously discriminate between a large number of different bead sizes and fluorescent intensities, placing each different DNA probe on different beads allows the system to become well suited to automation.

1.3.2.3 Sequence Based Typing (SBT): SBT is used to achieve high-resolution HLA typing results. For example, typing HLA-A, -B, or -C can be achieved by producing a large PCR amplicon that encompasses all of exon 2, intron 2 and exon 3 from genomic DNA. The PCR product is sequenced and the sequences are used to identify the alleles via a bioinformatics approach [25]. A similar strategy has been used to perform SBT for HLA Class II alleles, by PCR amplifying and then sequencing most or all of exon 2 of the corresponding genes. This method amplifies the most polymorphic regions in the HLA genes and has the potential

to identify new alleles that have unidentified polymorphisms within the areas sequenced [16].

One of the limitations of the existing SBT methods is that a limited amount of sequence is obtained (exons 2 and 3); thus any SNPs or other polymorphisms that lie outside of the genomic areas analyzed remain unknown. One way to obtain additional sequence information about polymorphisms that lie within any of the coding exons (particularly exon 1 may be of interest) is to isolate mRNA from patient samples and then perform reverse transcriptase-PCR (RT-PCR) to amplify the entire coding region of the HLA genes of interest, which can in turn be sequenced. However, this approach has not been widely used because of the technical challenge of purifying high quality mRNA from patient blood samples. This requires freshly isolated cells, specialized chemicals (e.g Trizol) and flash freezing or rapid processing of the blood samples. As well, RT-PCR requires RNAse inhibitors, priming oligonucleotides (oligo-dT, random hexamer, or sequence-specific), and the reverse transcriptase enzyme, and the process is much more complex than direct PCR amplification of one or a few exons from human genomic DNA.

Another limitation (and subtlety) of the SBT method relates to the fact that most patients will be heterozygous for most HLA genes. Thus if the PCR (or even RT-PCR) product is size selected and then directly sequenced, two overlapping sequences will be obtained; this single degenerate DNA sequence

must then be deconstructed into the two original contributing allelic sequences in order to complete the HLA typing. This adds further complexity to the SBT process, and not all heterozygous combinations can be unambiguously deconstructed. To evade this problem many investigators have used the approach of ligating the PCR products into a plasmid vector, transforming this into *E. coli*, plating and picking individual bacterial colonies, and then sequencing a dozen or more of the individual plasmid clones obtained. Typically this yields DNA sequences which are either one or the other of two clearly identifiable alleles, with only rare clones containing 'mixtures of known pure alleles' which arise as a PCR artifact (Richard Moore, BC Cancer Agency Genome Sciences Centre, personal communication). Some investigators have dubbed this the 'PCR then clone' approach to SBT.

1.4 HUMAN LEUKOCYTE ANTIGENS AND DISEASE

HLA molecules play a crucial role in the prevention of pathogenic infections by enabling T-cells to discriminate foreign antigens from self-antigens. The presence or absence of certain HLA alleles can lead to susceptibility to or protection from infection by certain specific pathogens, and/or to the development of autoimmune diseases. The latter occur when self-antigens are mistaken as foreign, and the adaptive immune system attacks healthy self-tissues leading to inappropriate inflammation and tissue destruction. In the most general case, by typing the HLA alleles expressed by a population of individuals

suffering from a given disease X, and then comparing their allelic distribution to that of a healthy control population, investigators have been able to discover associations between disease X and certain specific HLA alleles. One key benefit to identifying disease-specific susceptibility and protective alleles is that the data can be used to further our understanding of immunogenic profiling, risk management and therapeutic assessment. There are several autoimmune diseases, as well as two environmental diseases that have displayed correlations with certain specific class I or class II HLA alleles, and some of the most well known of these are listed in Table 1.1 [26, 27].

1.4.1 ANKYLOSING SPONDYLITIS - HLA-B27

Ankylosing spondylitis (AS) is a chronic, inflammatory disease that affects the joints of the axial skeleton, specifically the sacroiliac joints and spine [28]. Notably, AS was the first disorder identified to have a direct HLA association [29, 30]. HLA-B27 was identified in 90-95% of patients with AS and has also been found to correlate with other autoimmune disorders including spondylarthropathy, reactive arthritis, and inflammatory bowel disease [31].

1.4.2 INSULIN DEPENDENT DIABETES MELLITUS (IDDM)

Insulin dependent diabetes mellitus (IDDM) is an autoimmune disorder that is characterized by the killing of beta cells in the islets of Langerhans in the pancreas [32]. This leads to the inability to produce insulin and the development

Table 1.1 Diseases and human leukocyte antigens association.

Disease	Category	HLA genes associated		Pathology	Relative Risk	Sex Ratio (F:M)
		Susceptible	Protective			
Ankylosing spondytilis (AS)	Autoimmune	HLA-B27		Inflammatory disease targeting multiple articular and para-articular structures in the axial skeleton	87.4	0.3
Insulin dependent diabetes mellitus (IDDM/Type I)	Autoimmune	DR3,DQ2/DR4,DQ8	DQ6	Autoimmunity to beta cells in islet of Langerhans in the pancreas	25	~1
Systemic lupus erythematosus	Autoimmune	DR3	-	Multisystem autoimmune connective tissue disorder with various clinical presentations	5.8	10-20
Multiple Sclerosis (MS)	Autoimmune	DRβ*15/*17	DRβ*01/*10/*11/*14	Chronic inflammation associated with widespread primary demyelination and glial scarring in the CNS	4.8	10
Rheumatoid Arthritis	Autoimmune	DR4	-	Chronic, progressive inflammatory disease targeting the connective tissues of the body	4.2	3
Celiac Disease	Environmental (dietary protein)	DQ2 (DQ8 in some populations)	-	Inflammation of the lining of the small intestine which interferes with absorption, caused by eating gluten	2-31†	?
Chronic Berillium Disease (CBD)	Environmental (inhaled metallic cation)	DPB1*02:01	DPB1*04:01	Hypersensitivity disorder targeting the lungs that develops as a result of exposure to beryllium in the workplace	-	ND
Allergic Contact Dermatitis (ACD)	Environmental (small molecules which penetrate the skin)	No HLA associat. (Nickel) Unknown (Hair Dye = PPD)	-	Delayed hypersensitivity reaction in the skin, characterized by erythema, edema and vesiculation in response to specific environmental agents	-	ND

† The relative risk varies whether the individual is heterozygous or homozygous. DQ2 heterozygous individuals with a paired DQ8 gene have a higher relative risk than DQ2 heterozygous individuals with other DQ alleles. ND – not determined

There are several human diseases where risk is correlated with the presence of certain specific human leukocyte antigen (HLA) alleles. Susceptibility alleles make the individual more prone to develop the disorder, while protective alleles reduce the relative risk of developing the disorder.

Adapted from Murphy, K.P., et al., *Janeway's immunobiology*. 7th ed. 2008, New York: Garland Science. xxi, 887 p.

of hyperglycemia in the patient. IDDM has displayed a well-documented association with certain specific HLA-DQ and HLA-DR alleles [32-35].

The presence and absence of certain alleles are what determine the relative risk of developing IDDM. Certain alleles, such as HLA-DR3,DQ2 (DRB1*03:01-DQα1*05:01-DQB1*02:01; the DR-DQ genes are closely linked and these particular alleles frequently occur together), DR4,DQ8 (DRB1*04:01-DQα1*03:01-DQB1*03:02; also frequently occur together), and especially the heterozygous combination of these two alleles have been shown to give a high risk of susceptibility to development of IDDM. Some alleles, such as HLA-B*57:01, HLA-A*11:01, and especially DQ6 (DQα1*01:02-DQB1*06:02) have been shown to have protective properties [36].

More recent studies have identified that HLA-DP genes may also play a role in the relative risk of developing IDDM. Like other HLA alleles associated with IDDM, HLA-DP has certain specific combinations of alleles showing the greatest risk, such as HLA-DPB1*03:01 and HLA-DPB1*02:02, as well as protective alleles, such as HLA-DPB1*04:02 [37-39].

1.4.3 MULTIPLE SCLEROSIS

Multiple sclerosis (MS) is a neurological disorder that is characterized by demyelination in the central nervous system (CNS) [40, 41]. MS is known as a CD4 T-cell mediated autoimmune disease [42, 43], and as such is known to be associated with certain HLA-DR and HLA-DQ alleles [44]. Other HLA genes have

been shown to have a correlation with the development and severity of MS. MHC class I genes have been associated with MS, but to a much lesser extent than DR and DQ genes [45, 46].

Much like the associations seen in diabetes mellitus, the HLA association with MS is complex; it is the presence and absence of certain key alleles that appears to modulate the relative risk of MS. Within the HLA-DRB1 gene there are two alleles associated with increased relative risk of MS: DRB1*15 and DRB1*17. In contrast, there are four protective alleles with different protective mechanisms. DRB1*14 and DRB1*11 are generally protective alleles, whereas DRB1*01 and DRB1*10 are only protective when paired with DRB1*15, and show no affect on relative risk when paired with DRB1*17 [47].

In a study by Yu *et al.*, HLA-DP was found to be involved in epitope spreading during the early development of MS, whereby the optic nerve becomes involved (opticospinal MS) [48]. A second study by Wu et al., typed for HLA-DRB1 and HLA-DPB1 in a Southern Han Chinese population using SBT [49]. This study found that the allele HLA-DPB1*05:01 was present in 90% (n=60) of patients screened for conventional MS. A recent study by Field *et al.*, identified an HLA-DPB1 polymorphism (rs9277535) located in the 3' untranslated region of the gene that shows a strong correlation with MS development [50].

1.4.4 BERYLLIOSIS

Berylliosis, or chronic beryllium disease (CBD), is a hypersensitivity disorder of the lungs that is triggered by exposure to beryllium dust in the workplace [51]. CBD is maintained by the accumulation of a large number of beryllium specific T-cells in the lungs. CBD and its associations with HLA molecules is of special interest because CBD can be used as a model for how immunogenetic factors might influence susceptibility to a localized inflammatory disease triggered by exposure to a small molecule (in this case a simple metallic cation) coming from the environment. CBD differs from celiac disease, which is also a localized inflammatory disease triggered by exposure to a molecule from the environment, in that celiac disease is caused by an immune stimulus to a peptide derived from dietary gluten and covalently modified by the action of transglutaminase enzymes in the small intestine [52].

CBD has been associated with certain HLA-DPB1 class II molecules, specifically those which have the 69^{th} amino acid as glutamic acid (Glu^{69})[53, 54]. Other studies have supported these findings and have found that DPB1 alleles with Glu^{69} was present in 73-95% of CBD patients, compared to 30-48% of control patients who were also exposed to beryllium but in whom no disease developed [55, 56]. Consistent with the Glu^{69} observation, several studies have confirmed that HLA-DPB1 alleles play a role in the development of CBD, with

individuals with HLA-DPB1*02:01 being susceptible to disease, whereas those with HLA-DPB1*04:01 are protected.

The associations identified in berylliosis with certain HLA alleles are of special interest to the present study because the pathogenesis of the disease has some parallels with the development of allergic contact dermatitis (ACD) against permanent hair dyes containing the contact allergen para-phenylenediamine (PPD). In both cases the development of disease involves a reactive small molecule (a beryllium cation in CBD and a small organic molecule in PPD-ACD) that presumably interacts with self-proteins to create neoantigenic peptide epitopes. One plausible theory is that it is recognition of these neoantigenic peptide epitopes by memory T-cells which triggers the localized inflammatory response. Thus studies that have identified HLA associations in berylliosis can be used as a working model as we investigate HLA associations in patients with ACD to PPD.

1.4.5 HLA AND ALLERGIC CONTACT DERMATITIS

Allergic contact dermatitis (ACD) is a T-cell mediated type IV delayed hypersensitivity response characterized by erythema, edema and vesiculation in response to specific environmental agents such as certain metals (e.g. Nickel) or certain small reactive molecules (e.g. para-phenylenediamine (PPD), the active ingredient in common hair dyes) [57, 58]. The skin's reaction to contact allergens has two principal stages: sensitization, which can occur upon first exposure to

the allergen or after many exposures, and elicitation, which is when an individual sensitive to a given allergen is challenged with the same or a related allergen.

Hair dye allergy is a common occupational hazard for individuals working in the hair styling industry. It is caused by exposure to para-phenylenediamine (PPD) the main ingredient found in permanent hair dyes. PPD is a prohapten that requires oxidation in order to release the reactive metabolites that initiate hapten reactions [59]. Knowledge of the specific details about the nature of the immune response directed against PPD is limited, yet a study by Siebens *et al*. showed that peripheral blood mononuclear cells (PBMC) from hair dye-allergic individuals could be specifically stimulated to proliferate in the presence of PPD [60]. In these experiments, proliferation was significantly decreased (up to 95% reduction in activation) when HLA-DP molecules were blocked with specific monoclonal antibodies [60], suggesting that DP molecules may play a role in the immune response against hair dye.

A few other contact allergens have shown HLA associations with ACD development, at least in small studies. Mercury displayed a weak association with HLA-DR6 [61] and N-isopropyl-N`-phenyl-p-phenylenediamine (IPPD) showed a somewhat stronger association with HLA-Dw3 [62]. On the other hand, contact allergies against nickel have in general been shown to have no direct association with particular HLA alleles [63], but this cation has been recently

shown to signal directly to the innate immune system via toll-like receptor 4 (TLR4) [64].

1.5 SEQUENCE-BASED HLA TYPING REQUIRES ENRICHMENT OF SPECIFIC DESIRED DNA MOLECULES FROM COMPLEX MIXTURES

Sequence-based HLA typing (SBT) requires that specific genomic or cDNA fragments encoding the HLA molecules to be typed are somehow obtained in pure form so that they can be subjected to DNA sequencing. The technical challenge is that the desired DNA fragments are always first obtained as very minor constituents of complex mixtures. For example, if we wish to obtain exon II of the DPB1 gene (264 bp in length) and we are using human genomic DNA as the starting material, the desired DNA fragment is present at about 1 in 10 million (about 300 bp in 3 x 10^9 bp of genomic DNA, with one gene per haploid genome). Using cDNA as starting material represents considerable enrichment, since only a small portion of the genome is transcribed. For example, assuming the average MHC class II positive antigen presenting cell contains a total of 400,000 mRNA molecules, of which about 50 would encode DPB1 (i.e. DPB1 would fall on the lower side of a moderately abundant mRNA), and the desired cDNA fragment would be present at about 1 in 8000. In practice, the 1:8000 frequency would only be true if a pure population of antigen presenting cells was used to make the starting mRNA (e.g. an Epstein Barr virus [EBV] transformed human B cell line), but it would be about 10-fold less frequent if for example

total buffy coat cells from human blood were used, since only a subset of these cells express MHC class II proteins. With either genomic DNA or cDNA as starting material, it is clear that there is a 'purification problem' that must be solved before SBT can be performed.

In theory, there are two possible approaches to solving the purification problem: 1) start with a minute quantity of either genomic or cDNA, and then perform massive specific amplification by PCR, so that the desired DNA molecule becomes overwhelmingly abundant compared to the starting mixture of DNAs; or 2) use some kind of classical DNA cloning method to deliberately isolate the desired DNA molecule. Since PCR is now so ubiquitous and powerful, the second approach to the purification problem has largely been abandoned. However, one of challenges we set for this research project was to consider if certain novel DNA cloning technologies could be applied to the 'purification problem' that must be solved before SBT can be performed.

1.5.1 CLONING METHODS TO ISOLATE SPECIFIC DESIRED DNA MOLECULES FROM COMPLEX MIXTURES

The traditional approach to cloning genes is based on *in situ* filter hybridization, as illustrated in Figure 1.3. With this method phage or plasmid libraries are plated either randomly or in regular arrays on the surface of semi-solid agar growth medium and an imprint of the plaques/colonies is transferred to a solid membrane. These membranes are then subjected to hybridization

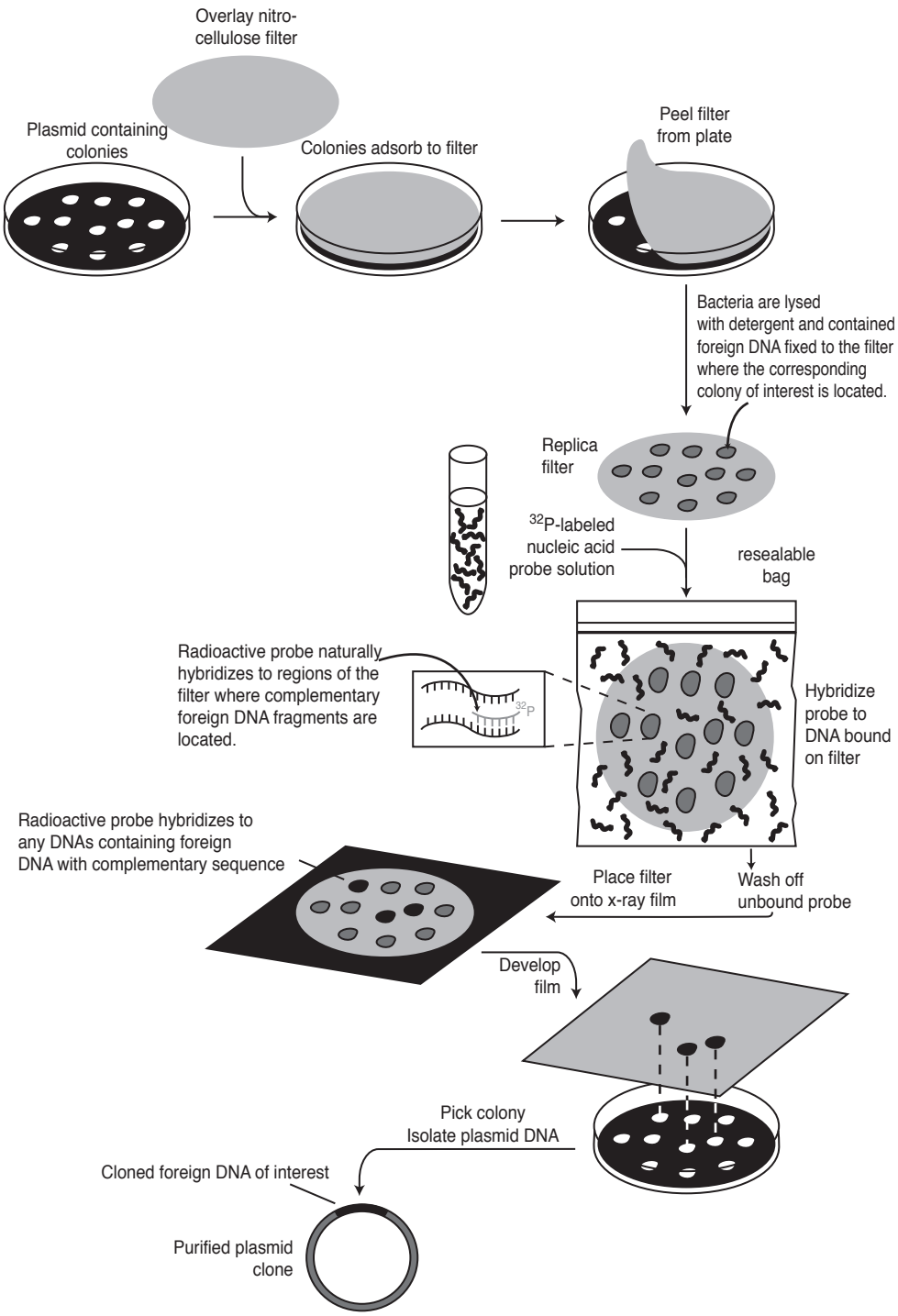


Figure 1.3 Screening a plasmid library of clones by filter hybridization. Colonies from the randomly plated library are 'tattooed' onto a nitrocellulose membrane filter, and the original petri plate stored in the cold room. The bacterial colonies are absorbed into the filter, the cells are lysed and the plasmid DNA is melted and then fixed (much of it in single stranded form) to the filter where the corresponding colony was located. A radioactively labeled DNA probe is incubated with the filter, and it naturally hybridizes to areas of the filter containing complementary DNA sequences. The filter is washed and exposed on X-ray film. By aligning the developed X-ray film with the original petri plate, positive colonies can be located. These are picked, grown up, and the corresponding plasmid DNA isolated for analysis.

with radioactively labeled probes in order to find clones that contain the specific desired DNA molecules. This method is slow and labor intensive, in part because in order to screen sufficient numbers of clones, the phage or plasmid libraries are typically plated at random at very high density (i.e. much higher than shown in Figure 1.3, which is for illustrative purposes only). Because of the high density, the isolation of pure clones requires secondary and tertiary screenings at decreasing clone densities. With random plating, the process allows screening of at most one to a few million clones per experiment. With certain valuable libraries, such as bacterial artificial chromosome [BAC] genomic libraries, clones may be plated randomly at lower densities and the resulting individual clones picked into 384-well plates using a colony-picking robot. The same or a different robot is then used to array the clones onto filters in a regular grid pattern, and the resulting membranes subjected to hybridization with radioactively labeled probes as above. This process is also very slow and labor intensive, and the numbers of clones that can be screened is even lower than for randomly plated libraries.

An alternate approach to obtaining desired DNA molecules from complex mixtures is based on solution phase hybridization of tagged synthetic oligonucleoties to complex libraries of clones in the form of pure DNA in solution. An important point about this approach is that virtually all vectors commonly used to make cDNA or genomic libraries yield DNA in a double-stranded form (e.g. plasmid, lambda phage, BAC, and most other vectors).

However, because it is stably base-paired, double stranded DNA is not available to hybridize with tagged oligonucleotides, and in fact only single-stranded DNA has this capacity (Figure 1.4). Thus in order to accomplish solution phase hybridization and selection of desired DNA molecules, the molecules or libraries of molecules must be in the form of single-stranded DNA. In this regard, of historic interest is a kit marketed by Invitrogen/Life Sciences called GeneTrapper cDNA positive cloning system.

The GeneTrapper cDNA Positive Selection System enables rapid purification of cDNA clones from complex cDNA libraries (representing 10^{12} DNA molecules). As outlined in Figure 1.5, the GeneTrapper procedure requires an oligonucleotide complementary to the target cDNA that is biotinylated at the 3´-end with biotin-14-dCTP using terminal deoxynucleotidyl transferase (TdT). A sample of the complex cDNA library (constructed in a double stranded phagemid – pSPORT1-CAT) is converted to ssDNA using Gene II (phage F1 endonuclease) and (*E. coli*) Exonuclease III (Exo III). The biotinylated oligonucleotide probe and the ssDNA library are allowed to hybridize under ideal conditions. The hybrids formed between the biotinylated oligonucleotide and ssDNA are then captured on streptavidin-coated paramagnetic beads. A magnet stand is used to pellet the paramagnetic beads from the solution, leaving any ssDNA constructs not bound to the oligonucleotide probe behind. The captured ssDNA clones are released from the paramagnetic beads using elution buffer (provided in the kit). The captured ssDNA clones are primed and converted to double stranded DNA using

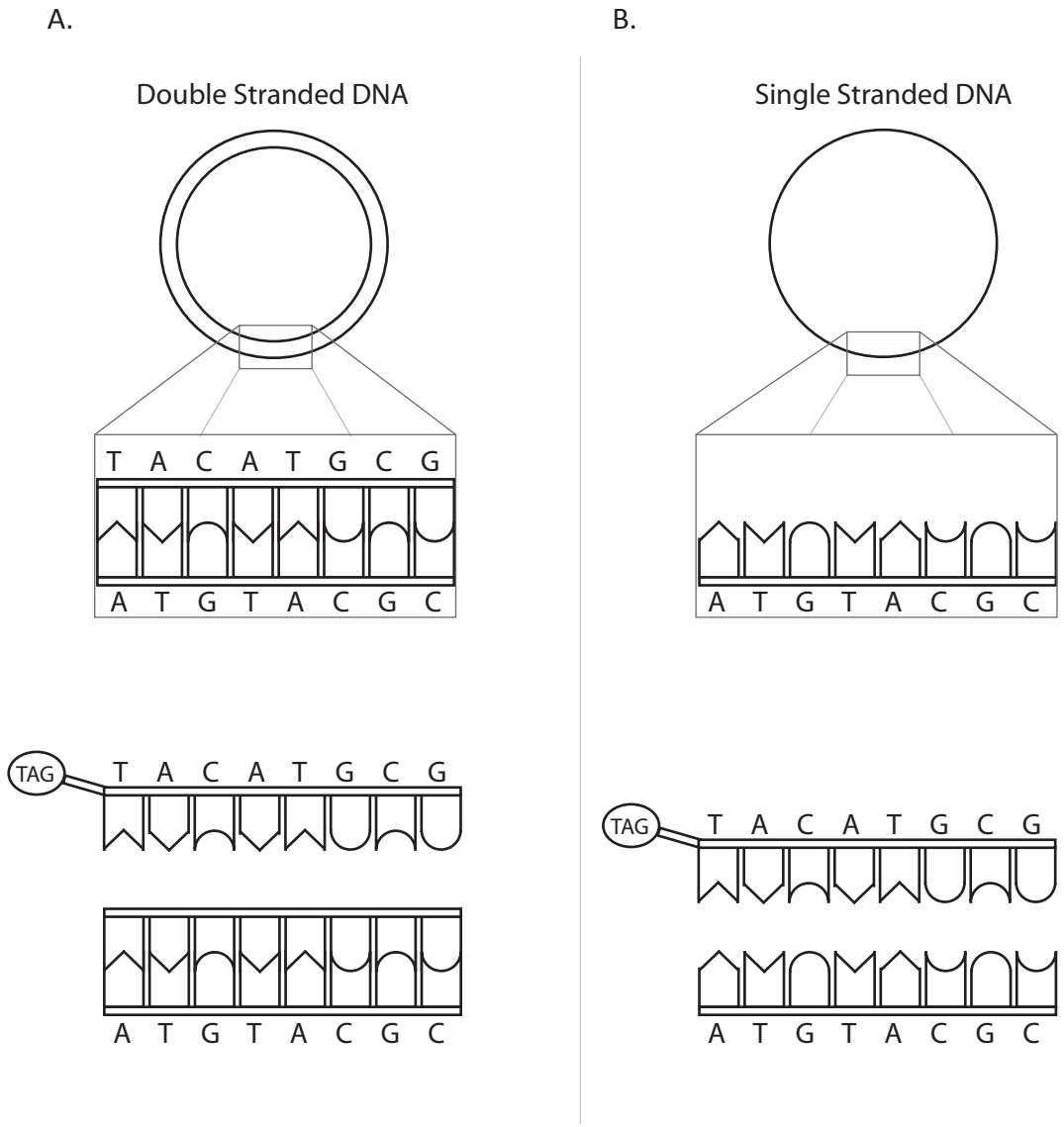


Figure 1.4 Double stranded compared to single stranded circular DNA molecules. Double-stranded DNA is not able to hybridize with tagged oligonucleotides, whereas single-stranded DNA can be selected in this way. In order to perform solution phase hybridization and selection of desired DNA molecules, the target molecules must be in single-stranded form.

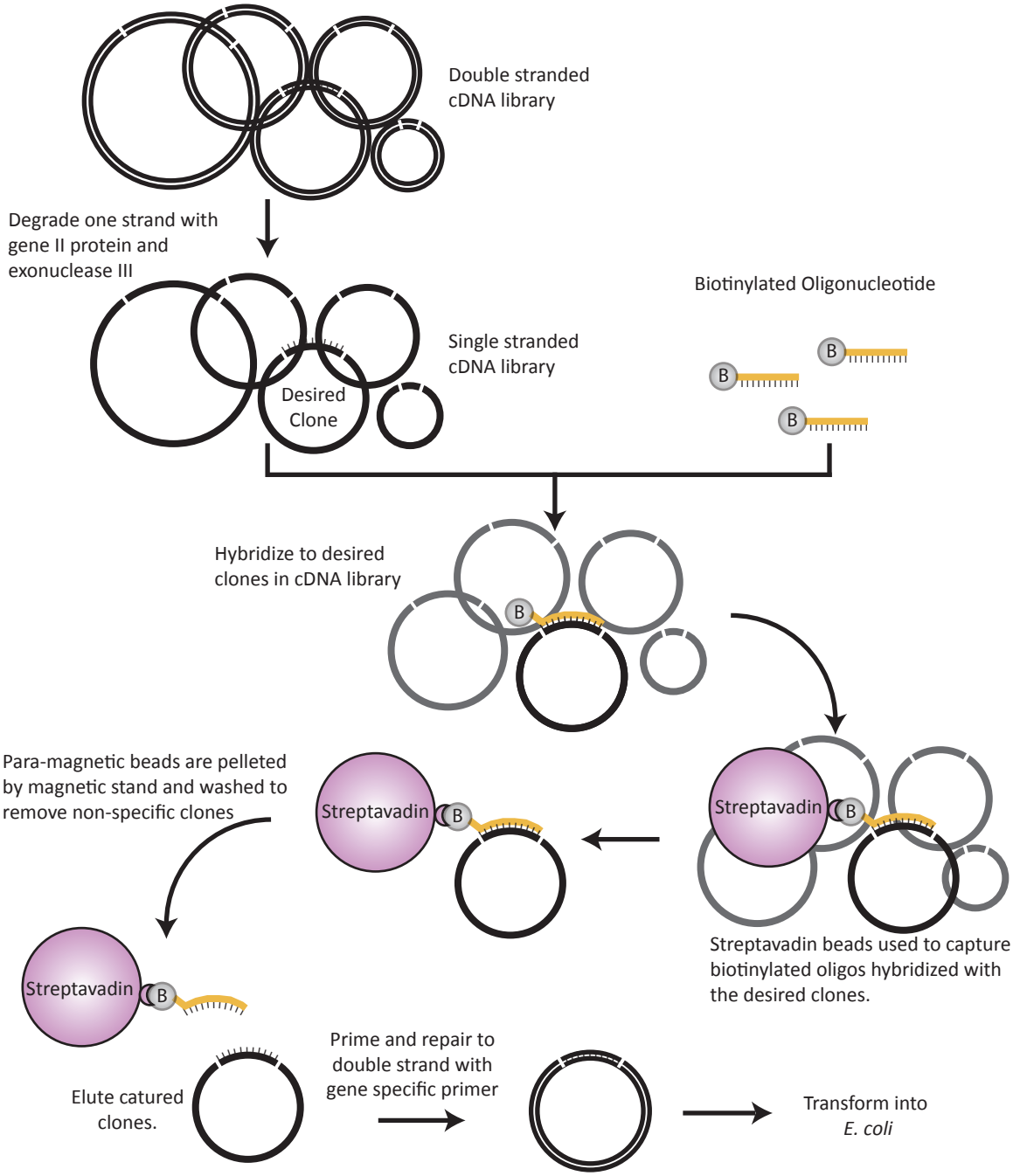


Figure 1.5 Overview of GeneTrapper cDNA positive selection system. A complex cDNA library of 10-100 x 10^6 independent clones is constructed in a double-stranded phagemid vector such as pSPORT1-CAT. Because the phagemid contains an F1 origin of replication, one of the two strands of each dsDNA molecule can be specifically cut by treatment with phage F1 endonuclease (i.e. gene II protein). The nicked double-stranded circular molecules are enzymatically converted into single stranded DNA by treating with exonuclease III, and a biotinylated oligonucleotide probe is then able to hybridize to the desired clones in the library. Biotinylated probes are captured with para-magnetic streptavadin beads, which are pelleted/washed using a magnetic stand. The captured target molecules are eluted (details of the elution buffer are unclear—possibly 33% formamide in TE at room temperature), converted to double stranded form using a target sequence-specific primer, and transformed into *E.coli.*

a ‘Repair enzyme’ – likely Klenow DNA polymerase. The dsDNA clones are transformed by electroporation into an *E. coli* cell line (UltraMAX-DH5α-FT cells) and plated. The colonies retrieved contain the cDNA clone of interest.

1.5.2 M13 BACTERIOPHAGE CLONING VECTOR

M13 is a single stranded DNA filamentous bacteriophage that targets F-pilus positive *Escherichia coli (E.coli)* bacteria. The M13 vector has been used as a cloning vector in the past [65-67], but has dropped in popularity due to the labor-intensive nature of growing and plating phage.

The lifecycle of M13 bacteriophage begins with infectious filamentous phage attaching to the F-pilus of host cells (Figure 1.6). The bacteriophage releases its single stranded DNA into the host cell where it uses host proteins to synthesize its second strand generating double stranded replicative form (RF) genome. The RF DNA generates progeny RF molecules, from which the rolling circle model synthesizes single stranded clones. The single stranded clones are packaged into viral coat proteins and are secreted from the infected cells. Infectious M13 phage are rapidly secreted from a newly infected host cell and can produce up to 1000 phage particles per infected cell within the first hour of infection. The secretion of M13 bacteriophage does not lyse the host cell, which produces pseudoplaques when M13 is plated on a bacterial lawn (M13 is not a lytic virion, but instead reduces the growth rate of infected cells forming turbid plaques we denote as pseudoplaques).

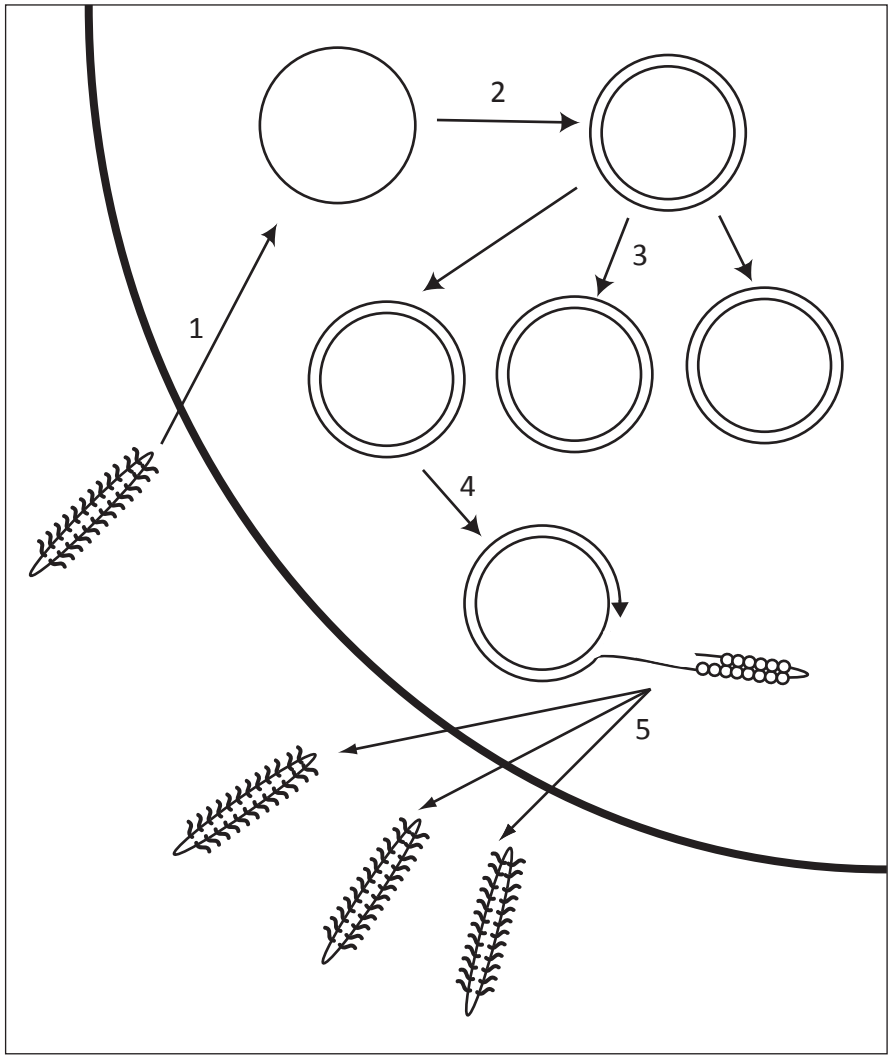


Figure 1.6 M13 Bacteriophage Lifecycle. M13 is a filamentous phage which can be used to prepare single stranded DNA. The life cycle can be divided into 5 steps: 1) The phage enters host cells (E.coli) through their F-pilus. Once in the cell, the phage loses its protein coat, revealing the single stranded circular DNA. 2) Using host proteins, the single stranded DNA is converted into double stranded replicative form (RF). 3) DNA replication produces progeny RF molecules. 4) RF molecules give rise to single stranded viral DNA by rolling circle replication. 5) The single stranded clones are packaged with viral coat proteins and released from the host cell without causing cell lysis. The phage can be harvested from the bacterial culture supernatant using polyethylene-glycol (PEG) precipitation.

The RF DNA can be extracted from infected cells and used as a cloning vector to produce M13 based constructs. The single stranded DNA can be extracted in high quantities from the phage present in the media of infected cultures. M13 has again become an area of interest as of late because of the ability to use genetically modified M13 as a biological scaffold for inorganic molecules. For example, recently M13 with modified coat protein was utilized to generate cobalt-platinum crystals that self-assembled on M13 molecules [68].

1.6 OBJECTIVES AND PROJECT OVERVIEW

The original long-term objective of this research project was to develop a novel technique to enable complete coverage HLA typing of patient samples in a cost-effective manner. In this case 'complete coverage' would mean that all coding exons of all classical HLA genes would be typed, and all possible alleles, including those not previously reported, would be identified. To accomplish this we initially proposed a SBT method involving cloning of total cellular cDNA into M13-bacteriophage vectors, generation of complex libraries of single stranded circular DNA molecules (sscDNA), and selection of desired HLA sequences/M13 clones using oligonucleotide-driven solution-based hybridization. Several model experiments were performed in order to determine if the oligonucleotide-driven solution-based hybridization method could be used to specifically enrich for M13 sscDNA molecules containing desired HLA sequences. However, the approach proved to be impractical and it was abandoned.

The second part of this M.Sc. project focused on a much more circumscribed objective: To determine if certain DPB1 alleles were more highly represented in a population of 16 hair-dye allergic patients with proven contact allergy to PPD. To accomplish the DPB1 typing we used SBT with an 'RT-PCR then clone' approach. We isolated total cellular RNA from freshly drawn patient blood samples, generated first strand cDNA using oligo-dT priming and thermostable recombinant AMV reverse transcriptase, and then amplified the entire coding region of DPB1 (ATG to TAA) by polymerase chain reaction (PCR) using a high fidelity Phusion DNA polymerase. Amplified fragments were initially cloned into an in house vector (pCSD7s) in a specific orientation using asymmetrical non-self cohesive sticky ends generated using the restriction enzyme *Sfi* I. This method of cloning the DPB1 PCR fragments was successful, but it was quickly superseded by a TOPO® cloning approach using a commercial kit, which proved to be much faster and easier. For each of the 16 patients to be DPB1 typed, 24 independent plasmid clones generated from TOPO® cloning ligation and transformation reactions were sent to a high through-put commercial sequencing service for determination of insert sequences. Analysis of the resulting sequences allowed DPB1 allele assignment for all but one patient. DPB1 allelic frequencies were compared between the 15 available hair-dye allergic patients and a normal population, with data on the latter population being taken from the literature.

1.7 References

1. Murphy, K.P., et al., Janeway's immunobiology. 7th ed. 2008, New York: Garland Science. xxi, 887 p.
2. Kennedy, M.A., A brief review of the basics of immunology: the innate and adaptive response. Vet Clin North Am Small Anim Pract. **40**(3): p. 369-79.
3. Guermonprez, P., et al., Antigen presentation and T cell stimulation by dendritic cells. Annu Rev Immunol, 2002. **20**: p. 621-67.
4. Trombetta, E.S. and I. Mellman, Cell biology of antigen processing in vitro and in vivo. Annu Rev Immunol, 2005. **23**: p. 975-1028.
5. Kumanovics, A., T. Takada, and K.F. Lindahl, Genomic organization of the mammalian MHC. Annu Rev Immunol, 2003. **21**: p. 629-57.
6. Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium. Nature, 1999. **401**(6756): p. 921-3.
7. Yung Yu, C., et al., The human and mouse MHC class III region: a parade of 21 genes at the centromeric segment. Immunol Today, 2000. **21**(7): p. 320-8.
8. Trowsdale, J., Genetic and functional relationships between MHC and NK receptor genes. Immunity, 2001. **15**(3): p. 363-74.
9. Robinson, J., et al., *IMGT/HLA database--a sequence database for the human major histocompatibility complex.* Tissue Antigens, 2000. **55**(3): p. 280-7.
10. Robinson, J., et al., *The IMGT/HLA database.* Nucleic Acids Res, 2009. **37**(Database issue): p. D1013-7.
11. Vilches, C. and P. Parham, *KIR: diverse, rapidly evolving receptors of innate and adaptive immunity.* Annu Rev Immunol, 2002. **20**: p. 217-51.
12. Dausset, J., The major histocompatibility complex in man. Science, 1981. **213**(4515): p. 1469-74.
13. Cresswell, P., et al., The nature of the MHC class I peptide loading complex. Immunol Rev, 1999. **172**: p. 21-8.
14. Kutlu, A., et al., Th1-Th2 interaction: is more complex than a see-saw? Scand J Immunol, 2007. **65**(4): p. 393-5.
15. Dausset, J., [Iso-leuko-antibodies.]. Acta Haematol, 1958. **20**(1-4): p. 156-66.
16. Howell, W.M., V. Carter, and B. Clark, The HLA system: immunobiology, HLA typing, antibody screening and crossmatching techniques. J Clin Pathol. **63**(5): p. 387-90.

17. Parham, P., Typing for class I HLA polymorphism: past, present, and future. Eur J Immunogenet, 1992. **19**(5): p. 347-59.
18. Colombo, M.B., et al., *Luminex technology for anti-HLA antibody screening: evaluation of performance and of impact on laboratory routine.* Cytometry B Clin Cytom, 2007. **72**(6): p. 465-71.
19. Rajalingam, R., Cecka, M., Reed, E.F., Molecular HLA Typing Methods Used in Clinical Laboratories, in Molecular Diagnostics. 2010, Academic Press: San Diego.
20. Bunce, M., et al., Phototyping: comprehensive DNA typing for HLA-A, B, C, DRB1, DRB3, DRB4, DRB5 & DQB1 by PCR with 144 primer mixes utilizing sequence-specific primers (PCR-SSP). Tissue Antigens, 1995. **46**(5): p. 355-67.
21. Ando, H., et al., HLA-C genotyping in the Japanese population by the PCR-SSP method. Tissue Antigens, 1996. **48**(1): p. 55-8.
22. Oh, S.H., K. Fleischhauer, and S.Y. Yang, Isoelectric focusing subtypes of HLA-A can be defined by oligonucleotide typing. Tissue Antigens, 1993. **41**(3): p. 135-42.
23. Levine, J.E. and S.Y. Yang, SSOP typing of the Tenth International Histocompatibility Workshop reference cell lines for HLA-C alleles. Tissue Antigens, 1994. **44**(3): p. 174-83.
24. Itoh, Y., et al., High-throughput DNA typing of HLA-A, -B, -C, and -DRB1 loci by a PCR-SSOP-Luminex method in the Japanese population. Immunogenetics, 2005. **57**(10): p. 717-29.
25. Abbott, W.G., et al., Low-cost, simultaneous, single-sequence genotyping of the HLA-A, HLA-B and HLA-C loci. Tissue Antigens, 2006. **68**(1): p. 28-37.
26. Rioux, J.D., et al., Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases. Proc Natl Acad Sci U S A, 2009. **106**(44): p. 18680-5.
27. Ghodke, Y., et al., HLA and disease. Eur J Epidemiol, 2005. **20**(6): p. 475-88.
28. Thorsby, E., HL-A antigens and genes. I. A study of unrelated Norwegians. Vox Sang, 1969. **17**(2): p. 81-92.
29. Schlosstein, L., et al., High association of an HL-A antigen, W27, with ankylosing spondylitis. N Engl J Med, 1973. **288**(14): p. 704-6.
30. Brewerton, D.A., et al., Ankylosing spondylitis and HL-A 27. Lancet, 1973. **1**(7809): p. 904-7.
31. Smith, J.A., E. Marker-Hermann, and R.A. Colbert, Pathogenesis of ankylosing spondylitis: current concepts. Best Pract Res Clin Rheumatol, 2006. **20**(3): p. 571-91.
32. Csorba, T.R., A.W. Lyon, and M.D. Hollenberg, Autoimmunity and the pathogenesis of type 1 diabetes. Crit Rev Clin Lab Sci. **47**(2): p. 51-71.

33. Nerup, J., et al., HLA, autoimmunity, virus and the pathogenesis of juvenile diabetes mellitus. Acta Endocrinol Suppl (Copenh), 1976. **205**: p. 167-75.
34. Nepom, G.T. and H. Erlich, MHC class-II molecules and autoimmunity. Annu Rev Immunol, 1991. **9**: p. 493-525.
35. Nepom, G.T. and W.W. Kwok, Molecular basis for HLA-DQ associations with IDDM. Diabetes, 1998. **47**(8): p. 1177-84.
36. Howson, J.M., et al., Confirmation of HLA class II independent type 1 diabetes associations in the major histocompatibility complex including HLA-B and HLA-A. Diabetes Obes Metab, 2009. **11 Suppl 1**: p. 31-45.
37. Pociot, F., et al., Genetics of type 1 diabetes: what's next? Diabetes. **59**(7): p. 1561-71.
38. Noble, J.A., et al., The HLA class II locus DPB1 can influence susceptibility to type 1 diabetes. Diabetes, 2000. **49**(1): p. 121-5.
39. Varney, M.D., et al., HLA DPA1, DPB1 alleles and haplotypes contribute to the risk associated with type 1 diabetes: analysis of the type 1 diabetes genetics consortium families. Diabetes. **59**(8): p. 2055-62.
40. Sospedra, M. and R. Martin, Immunology of multiple sclerosis. Annu Rev Immunol, 2005. **23**: p. 683-747.
41. Compston, A. and A. Coles, Multiple sclerosis. Lancet, 2008. **372**(9648): p. 1502-17.
42. Hafler, D.A., Multiple sclerosis. J Clin Invest, 2004. **113**(6): p. 788-94.
43. Martin, R., H.F. McFarland, and D.E. McFarlin, Immunological aspects of demyelinating diseases. Annu Rev Immunol, 1992. **10**: p. 153-87.
44. Hillert, J. and O. Olerup, HLA and MS. Neurology, 1993. **43**(11): p. 2426-7.
45. Fogdell-Hahn, A., et al., Multiple sclerosis: a modifying influence of HLA class I genes in an HLA class II associated autoimmune disease. Tissue Antigens, 2000. **55**(2): p. 140-8.
46. Bergamaschi, L., et al., HLA-class I markers and multiple sclerosis susceptibility in the Italian population. Genes Immun. **11**(2): p. 173-80.
47. Ramagopalan, S.V., et al., The inheritance of resistance alleles in multiple sclerosis. PLoS Genet, 2007. **3**(9): p. 1607-13.
48. Yu, M., et al., HLA-DP: a class II restriction molecule involved in epitope spreading during the development of multiple sclerosis. Hum Immunol, 1998. **59**(1): p. 15-24.
49. Wu, X.M., et al., Association of susceptibility to multiple sclerosis in Southern Han Chinese with HLA-DRB1, -DPB1 alleles and DRB1-DPB1 haplotypes: distinct from other populations. Mult Scler, 2009. **15**(12): p. 1422-30.
50. Field, J., et al., A polymorphism in the HLA-DPB1 gene is associated with susceptibility to multiple sclerosis. PLoS One. **5**(10): p. e13454.
51. Samuel, G. and L.A. Maier, Immunology of chronic beryllium disease. Curr Opin Allergy Clin Immunol, 2008. **8**(2): p. 126-34.

52. Setty, M., L. Hormaza, and S. Guandalini, *Celiac disease: risk assessment, diagnosis, and monitoring.* Mol Diagn Ther, 2008. **12**(5): p. 289-98.
53. Wang, Z., et al., Beryllium sensitivity is linked to HLA-DP genotype. Toxicology, 2001. **165**(1): p. 27-38.
54. Richeldi, L., R. Sorrentino, and C. Saltini, HLA-DPB1 glutamate 69: a genetic marker of beryllium disease. Science, 1993. **262**(5131): p. 242-4.
55. Fontenot, A.P. and L.A. Maier, Genetic susceptibility and immune-mediated destruction in beryllium-induced disease. Trends Immunol, 2005. **26**(10): p. 543-9.
56. Tarantino-Hutchison, L.M., et al., Genetic determinants of sensitivity to beryllium in mice. J Immunotoxicol, 2009. **6**(2): p. 130-5.
57. Mark, B.J. and R.G. Slavin, Allergic contact dermatitis. Med Clin North Am, 2006. **90**(1): p. 169-85.
58. Saint-Mezard, P., et al., Allergic contact dermatitis. Eur J Dermatol, 2004. **14**(5): p. 284-95.
59. White, J.M., et al., *p-Phenylenediamine allergy: the role of Bandrowski's base.* Clin Exp Allergy, 2006. **36**(10): p. 1289-93.
60. Sieben, S., et al., Delayed-type hypersensitivity reaction to paraphenylenediamine is mediated by 2 different pathways of antigen recognition by specific alphabeta human T-cell clones. J Allergy Clin Immunol, 2002. **109**(6): p. 1005-11.
61. Sato, K., et al., Investigation of the association between mercury sensitization and HLA-DR6. Contact Dermatitis, 1996. **34**(4): p. 295.
62. Hegyi, E., M. Buc, and B. Busova, Frequency of HLA antigens in persons sensitive to N-isopropyl-N'-phenyl-p-phenylenediamine (IPPD). Contact Dermatitis, 1993. **28**(3): p. 194-5.
63. Schram, S.E. and E.M. Warshaw, Genetics of nickel allergic contact dermatitis. Dermatitis, 2007. 18(3): p. 125-33.
64. Schmidt, M., et al., *Crucial role for human Toll-like receptor 4 in the development of contact allergy to nickel.* Nat Immunol. **11**(9): p. 814-9.
65. Norrander, J., T. Kempe, and J. Messing, *Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis.* Gene, 1983. **26**(1): p. 101-6.
66. Messing, J., *A multipurpose cloning system based on the sing.* Recomb. DNA Tech, 1979.
67. Yanisch-Perron, C., J. Vieira, and J. Messing, *Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors.* Gene, 1985. **33**(1): p. 103-19.
68. Lee, S.K., D.S. Yun, and A.M. Belcher, *Cobalt ion mediated self-assembly of genetically engineered bacteriophage for biomimetic Co-Pt hybrid material.* Biomacromolecules, 2006. **7**(1): p. 14-7.

CHAPTER 2

MATERIALS AND METHODS

2.1 BACTERIAL STRAINS, MEDIA, AND CULTURE CONDITIONS

Escherichia coli (*E. coli*) strain XL2-Blue (endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 e14- Δ(mcrCB-hsdSMR-mrr)171 recB recJ sbcC umuC::Tn5 uvrC F'[::Tn10 proAB lacI^q Δ(lacZ)M15 Amy Cm^R]) (Stratagene) cells were used in all experiments except those involving TOPO® cloning kits, where frozen competent cells (DH5α T1^R) were already provided. Unless otherwise stated, bacteria were grown at 37°C in 2xYT-MT medium (40g yeast extract; 64g tryptone; 20g NaCl; 40ml 1M Tris (pH 7.5); 8ml 1M $MgCl_2$ per 4 litres of medium).

XL2-Blue cells express the F-pilus which makes them susceptible to M13 bacteriophage infection [1]. Cells were initially streaked onto LB agar plates containing 12.5 µg/ml tetracycline to select for F' (since this episome contains Tn10), and glycerol stocks (2xYT-MT + 10% glycerol) were made from this plate and frozen at -70˚C. Thereafter whenever fresh stocks of cells were required, a few chips of ice were transferred from the glycerol stock and streaked onto a non-selective 2xYT-MT agar plate. Since cells grown or stored at <37°C lose expression of the F-pilus, in all experiments involving the growth or plating of M13 bacteriophage extreme care was taken to always inoculate into freshly grown, still warm cultures.

2.2 PREPARATION OF TRANSFORMATION COMPETENT CELLS

For most cloning experiments competent cells were prepared in-house. To create a start up culture, a single isolated XL2-blue colony was picked from a 2xYT-MT agarose plate which had been streaked from the frozen glycerol stock the previous afternoon. The colony was transferred to 50ml of 2xYT-MT media in a 250ml baffled culture flask. The start up culture was incubated at 37°C with vigorous aeration until the absorbance reached an optical density at 600nm (OD_{600}) of 0.5 (approximately 3h). A scale up culture was inoculated with 15ml of the start up culture in 220ml of media in a 2L flask. The scale up culture was incubated at 37°C on an I26 Air Shaker (New Brunswick Scientific) until an OD_{600} of 0.5 was reached (approximately 4h). Flasks were placed in an ice-water bath for 45 minutes, then transferred to pre-chilled 250ml centrifuge bottles and spun at 950 RCF, 4˚C for 15 minutes to gently pellet cells (e.g. 2800rpm in a Beckman Avanti J26 XP centrifuge with a JS 7.5 rotor). From this point, procedures differed depending on if chemical competent or electrocompetent cells were generated.

2.2.1 CHEMICAL COMPETENT CELLS

The supernatant was discarded and cells were resuspended by gentle swirling with 50ml of TFB1 solution (30mM KOAc, 50mM $MnCl_2$, 100mM KCl, 10mM $CaCl_2$, 15% glycerol). Bacteria were centrifuged at 400 RCF, 4˚C for 15 minutes (1800rpm in an Avanti J26 XP centrifuge with a JS-7.5 rotor) and the

supernatant was discarded making sure not to disturb the cellular pellet. The cell pellet was gently resuspended by swirling with 5ml of TFB2 solution (10mM NaMOPS, 75mM $CaCl_2$, 10mM KCl, 15% glycerol). The TFB2 cell slurry was aliquoted into 100 µl fractions in 1.5ml Eppendorf tubes cooled on dry ice to ensure flash freezing of the cells. The cells were stored at -80°C until use.

2.2.2 CHEMICAL TRANSFORMATION

To test for chemical competency, cells were thawed on ice, and 4 µl of 10ng/µl pCSD7s DNA was added to 50 µl of competent cells. The cells were incubated with the DNA for 10 minutes on ice, then heat shocked at 37°C for two minutes and immediately returned to ice for 10 minutes. 1ml of 2xYT-MT media was added dropwise to the cells still in the Eppendorf tubes, and these were simply capped and transferred to a roller wheel and incubated at 37°C for one hour to allow the cells to recover. Serial dilutions were made in 2xYT-MT media (10^{-1} to 10^{-5}) and 100 µl aliquots of the dilutions were plated on 2xYT-MT agar plates containing 100 µg/ml ampicillin. Plates were incubated overnight at 37°C and colonies counted the next day. Typical efficiencies were 1 x 10^6 to 5 x 10^6 CFU/µg of pCSD7s plasmid DNA.

2.2.3 ELECTROCOMPETENT CELLS

The supernatant was discarded and cells were gently resuspended in 50ml of 10% glycerol solution in dH_2O. Bacteria were centrifuged and the pellets drained as for the chemical competent cells. The cell pellet was gently

resuspended by swirling with 5ml of 10% glycerol solution in dH_2O. The glycerol cell suspension was aliquoted into 100 µl fractions in 1.5ml Eppendorf tubes pre-chilled on dry ice to ensure flash freezing of the cells, and tubes stored at -80°C until use.

2.2.4 ELECTROPORATION

To test the frozen electrocompetent cells, an aliquot of cells was thawed on ice and 50 µl of these were placed in a pre-chilled 0.1cm gap electroporation cuvette (BioRad) together with 1 µl of pCSD7s DNA (10ng). The cuvette was placed into a Gene Pulser electroporator (BioRad), with voltage pre-set to 1.8kV. Cells were pulsed, the cuvette was removed from the electroporator and 1ml of chilled 2xYT-MT media was immediately added to the cuvette. The cells were gently mixed by pipetting and transferred to a chilled 1.5ml Eppendorf tube. 500 µl of chilled 2xYT-MT media were used to rinse out any left over cells from the cuvette , and this was added to the same chilled Eppendorf tube. The pooled cells (still in the Eppendorf tube) were incubated at 37°C for 1h to allow for recovery. After recovery, serial dilutions were made in 2xYT-MT media (10^{-1} to 10^{-5}) and 100 µl aliquots of the dilutions plated on 2xYT-MT plates containing 100 µg/ml ampicillin. Plates were incubated overnight at 37°C and colonies were counted the next day. The electrocompetent cells were considered acceptable if counts were greater than 100 colonies in the 10^{-3} dilution plate, giving a transformation efficiency of 1.0 x 10^8CFU/µg.

2.3 PROTOCOLS FOR M13 BACTERIOPHAGE

The M13 bacteriophage is a filamentous phage that is specific for F-pilus expressing *E.coli* bacteria. This phage has been widely used as a vector for construct generation and high fidelity cloning [1] (Figure 1.6).

M13mp19 bacteriophage particles were generated by transforming XL2-Blue *E. coli* chemical competent cells with 40ng of replicative form (i.e. double stranded) bacteriophage DNA. Transformed cells were not outgrown, but rather used directly to make serial dilutions in 2xYT-MT media (10^{-1} to 10^{-5}). One hundred microlitres of each dilution was plated on a lawn of freshly growing XL2-Blue bacteria. To generate the bacterial lawn, 200 µl of a fresh overnight culture of XL2-Blue cells was added to a 5 ml glass culture tube, 100 µl of the transformation reaction (appropriately diluted) was added, and after a 5 min incubation 2.5 ml of molten 2xYT-MT top agar (0.75 % agar at 42˚C) added. This was mixed gently but quickly, and the entire contents of the tube rapidly poured onto the top of a regular 10 cm petri plate already containing solidified 2xYT-MT/1.5% agar. Plates were incubated overnight at 37°C, with M13-transformed bacteria giving rise to pseudoplaques on the lawn in 12-18 hrs. Single isolated pseudoplaques were picked using a glass Pasteur pipette and the resulting agar plugs transferred into Eppendorf tubes pre-filled with 1ml dH_2O. The plaques were stored at 4°C overnight to allow phage to elute. The titer of the phage stock

was measured by plating several dilutions (typically 10^{-3} – 10^{-5}) on a fresh lawn of XL2-Blue cells.

2.3.1 PREPARATION OF LARGE SCALE M13 BACTERIOPHAGE CULTURES

This procedure generates large volume M13 bacteriophage cultures which can be used to extract single stranded circular DNA (SSC-DNA) from the culture supernatants (which contain progeny phage), and/or replicative form double stranded DNA (RF-DNA) from the infected *E. coli* cell pellets. In cases where SSC-DNA is required for solution phase hybridization and capture experiments (such as those described in Chapter 3), only the supernatant would need to be processed. In other cases where M13 vectors are used to clone or sub-clone foreign DNA inserts (e.g. sub-cloning DQ cDNA inserts into the vector M13JFE7S as described in Chapter 3), only the bacterial cell pellet would need to be processed.

One key to generating large quantities of M13 phage is to use rapidly growing, well aerated bacterial cultures, and to keep them in log phase growth at all times. Also they should be kept at 37˚C at all times so that the F-pilus is expressed at maximal levels. A single fresh XL2-Blue colony was inoculated into 2 ml of 2xYT-MT media, and placed on a roller wheel at 37°C. After 4-6 hrs this was divided equally into four separate 15 ml polypropylene tubes, each with 2ml 2xYT-MT, and these cultures were continued overnight on the roller wheel at

37°C. The following morning they were pooled and immediately used to inoculate the larger volume cultures.

Eight one litre baffled Erlenmeyer culture flasks were filled with 2xYT-MT media (250ml/flask) and pre-warmed/aerated for 30 minutes by placing them on an I26 Air Shaker (New Brunswick Scientific) set to 250rpm. Each flask was then inoculated with 1ml of the pooled growing stock bacterial culture, and cultures continued on the air shaker for a further 30 minutes after which 1 µl of M13 phage containing ≈$1x10^7$ PFU was added to each flask. Cultures were allowed to grow overnight (16h) at 37°C on the Air Shaker with vigorous aeration.

The following morning cultures were checked for turbidity without viscosity (i.e. healthy bacterial cell growth without host cell lysis). A 1.5 ml aliquot was collected for rapid analysis by Wizard mini-prep and gel electrophoresis (Fig. 2.1a). The cultures were transferred to 250 ml centrifuge bottles (maximum volume 180 ml/bottle in fixed angle rotor) and centrifuged at 10,000 RCF for 30 minutes at 4°C (10,000 rpm on Avanti J26 XP centrifuge with a JLA-16.250 rotor by Beckman Coulter). After centrifugation, the supernatants were decanted into a second set of 250 ml centrifuge bottles and the cell pellets drained. If RF form of DNA was required, pellets were subjected to alkaline lyses and processing by cesium chloride/ethidium bromide gradients (see below for method).

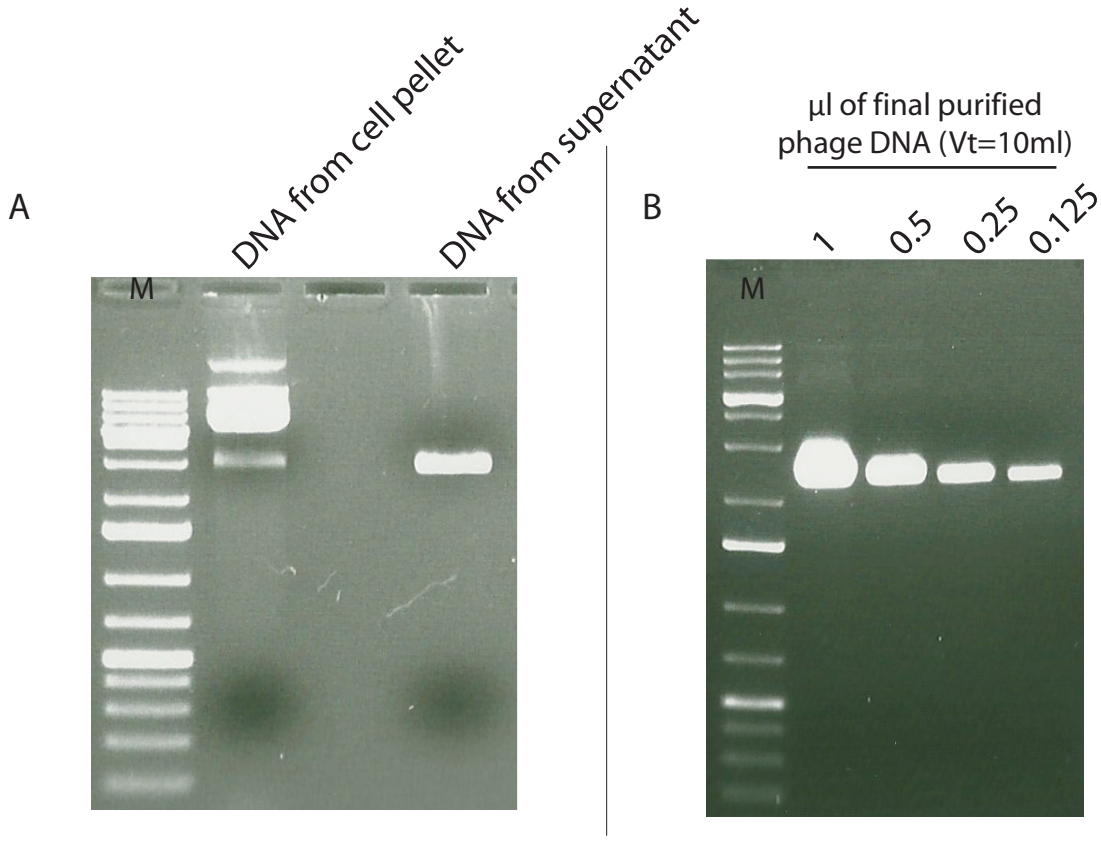


Figure 2.1 DNA obtained from large-scale M13mp19 bacteriophage cultures, analyzed by agarose gel electrophoresis. A. Results of rapid quality control test. Immediately following expansion and final growth of the ≈2 L culture (see text for details) a 1.5 ml aliquot of the culture broth was transferred to an eppendorf tube, pelleted in a microfuge, and episomal DNA present in the cell pellet extracted using a Wizard mini-prep kit (Promega). Phage DNA from the corresponding 1.5 ml of culture supernatant was precipitated by adding 0.25 ml of Polyethylene glycol/salt solution (20% PEG8000, 2.5M NaCl), incubating on ice for 5 minutes and pelleting in a microfuge. The same Wizard mini-prep protocol was used to purify the phage DNA, by simply treating the phage pellet as if it were a bacterial cell pellet. In order to rapidly visualize the DNA and determine if the bacteriophage infection had been productive, 1/5 of each Wizard mini-prep sample was run on a 1.0% agarose TBE gel at 70 volts for 30 minutes. Quality of the bands indicates that the bacteriophage infection had been successful, and that processing of the entire ≈ 2 L culture was warranted. Note that the cell pellet contains a substantial quantity of single stranded circular (SSC) DNA (lowest band in cell pellet lane, which runs at the same position as the phage DNA from the culture supernatant) B. Final purified SSC M13mp19 DNA. The doubly centrifuged culture supernatant was subjected to PEG precipitation, phenol/chloroform extraction, and ethanol precipitation. Total yield of SSC phage DNA was 3.96 mg, which was resuspended in 2 ml of TE buffer. A serial dilution of this DNA stock was run on a 1.0% TBE agarose gel at 70 volts for 1h. The gel shows highly purified phage DNA without any contaminating E.coli genomic DNA, and also without any phage deletion mutants.

2.3.2 EXTRACTION OF SINGLE STRANDED PHAGE DNA

To purify SSC-DNA, the supernatants from the first centrifugation were re-centrifuged at 19,800 RCF for 30 minutes at 4°C to remove any bacteria that may have still been present in the supernatant (14,000rpm on Avanti J26 XP centrifuge with a JLA 16.250 rotor by Beckman Coulter - again, because a fixed angle rotor was used, maximum volume per bottle was 180 ml). Supernatants from the second centrifugation run were pooled into a 4 liter Erlenmeyer flask, and a 1.5 ml aliquot of the pool was taken for rapid Wizard mini-prep analysis and gel electrophoresis. For every liter of supernatant recovered, 166 ml of PEG/NaCl solution (20% PEG 8,000, 2.5 NaCl) was added. The flask was placed in a large ice water slurry for one hour and swirled gently every 10-15 min; contents became cloudy as the phage began to precipitate.

The precipitated phage solution was aliquoted into clean 250ml centrifuge bottles (180 ml/bottle) and phage pelleted by centrifuging at 10,000 RCF, 4˚C using the JLA-16.250 rotor. The phage formed a large white pellet at the heel of the tube; supernatant was discarded and any traces of liquid removed by Pasteur pipette followed by drying the inside of the bottles with kim wipes (Kimberly-Clark). The first phage pellet was thoroughly resuspended in 10ml of TE buffer (10mM Tris-HCl, 1mM EDTA) by pipeting up and down, this solution was then transferred to the second pellet, and the process repeated sequentially until half of the pellets had been resuspended in the 10 ml volume,

which was finally transferred to a 50ml polypropylene (PP) falcon tube (Beckton-Dickinson). The first set of bottles were back-washed with an additional 10ml of TE to recover any residual phage, and this was added to the 50ml PP falcon tube. The other half of the phage pellets were resuspended in a second 10ml volume of TE buffer, this was transferred to a second 50ml PP falcon tube, and these bottles back-washed with an additional 10ml as above.

Three consecutive phenol extractions were performed on the two 20 ml aqueous phage samples: 30ml of phenol was added to each tube, these were mixed vigorously for 30 seconds, and allowed to sit at room temperature for 5 minutes before centrifuging at a low speed (400 RCF in a clinical centrifuge placed in the fume hood) and removing the aqueous phase to a clean tube. The aqueous phases were then extracted twice with 30 ml aliquots of phenol/chloroform, and finally twice with 30 ml aliquots of chloroform. These latter extractions were required to remove any residual PEG, phenol and protein contaminants. The organic phases were backwashed with 2ml of TE, and the final aqueous phases (20 ml + 2 ml) pooled into two fresh 50ml PP falcon tubes.

The total volume of the aqueous phases was reduced by mixing with an equal volume of sec-butanol, mixing vigorously, and centrifuging at low speed to separate phases (400 RCF – 1500 rpm in an Alegra 6KR Kneewell centrifuge with a GH-3.8 rotor by Beckman Coulter). Sec-butanol is an anhydrous solvent that absorbs water from the sample while leaving the DNA in the aqueous phase,

thereby concentrating the DNA. The sec-butanol treatment was repeated several times until the total aqueous volume reached 10ml (once the two aqueous volumes were small enough they were pooled and the last few sec-butanol extractions were done in a single 50ml PP falcon tube). The DNA solution was transferred into a new 50ml PP falcon tube, and 1ml of 3M NaOAc and 25ml of ice-cold 95% ethanol added. The tube was mixed thoroughly and placed at -80°C overnight to precipitate the DNA.

The following morning the sample was thawed on ice and the DNA pelleted by centrifugation at 3,000 RCF for 30 minutes at 4°C (5,000 rpm on Avanti J26 XP centrifuge with a JS-7.5 rotor, using 50 ml falcon tube adaptors by Beckman Coulter). The supernatant was removed by aspiration, the pellet washed with ice-cold 70% ethanol, dried briefly, and resuspended in 2 ml of TE buffer and stored at -80°C. The yields of the extracted single stranded DNA were analyzed by measuring the optical density of several serial dilutions; for the M13mp19 prep shown in Figure 2.1, at a dilution of 10^{-2} the OD_{260} was 0.496, giving a concentration of 1.98 μg/μl and a total yield of 3.96mg. The 260/280 ratio for this DNA solution was 1.93.

2.3.3 PURIFICATION OF REPLICATIVE FORM (RF) PHAGE DNA

Replicative form (RF) double stranded phage DNA was purified from M13 infected *E. coli* bacteria cells and the resulting vectors (e.g. M13JFE7s) used for cloning experiments as described in Chapter 3. The following protocol begins

with the bacterial cell pellets obtained in protocol 2.3.1 after the first centrifugation run, and while the pellets are still in the original 250 ml bottles.

Any traces of residual supernatant were removed from the cell pellets by aspiration. Half of the pellets were re-suspended in 25 ml of solution 1 (50mM glucose, 25mM Tris-HCl - pH 8.0, 10mM EDTA - pH 8.0) by adding the total volume to the first pellet, re-suspending by persistent pipetting up and down, moving this volume to the next pellet, etc. By a similar process, the other half of the pellets were re-suspended in a second 25 ml aliquot of solution 1, with the two cell slurries ending up in two of the original 250 ml centrifuge bottles. Once the two cell solutions were thoroughly re-suspended (i.e. homogeneous without clumps; this step is critical), 50 mls of solution 2 (0.2N NaOH, 1% SDS) was added to each bottle, and the bottles capped and mixed by gently inverting until the solution was homogeneous and translucent. Once this was achieved, 38 ml of solution 3 (5M KOAc, 11.5% Glacial acetic acid) was added; contents were gently mixed by inverting the capped tubes, and placed on ice for 10 minutes.

The samples were centrifuged at 10,000 RCF for 30 minutes at 4°C (10,000rpm on Avanti J26 XP centrifuge with a JLA-16.250 rotor) to pellet the bacterial genomic DNA, denatured cellular proteins, precipitated potassium dodecyl sulfate (SDS), etc. (these form a white precipitate). The two supernatants were poured into two plastic funnels loosely plugged at the neck with a piece of glass wool and sitting in two fresh 250 ml centrifuge bottles.

Once the supernatant had drained through the glass wool (which removes any clumps of floating material), 70ml of isopropanol was added to each bottle, and the bottles capped, mixed gently and stored at room temperature for 15 minutes to allow the DNA to precipitate. The DNA was pelleted by centrifugation at 10,000 RCF for 30 minutes at 4°C (10,000rpm on Avanti J26 XP centrifuge with a JLA-16.250 rotor), the supernatant was discarded and remnants of the supernatant were removed by aspirating and then drying the inside of the bottles with kim wipes. The pellets were resuspended in 8ml of TE buffer and prepared for ethidium bromide/cesium chloride gradient purification in order to separate double-stranded DNA from single-stranded phage DNA based on differences in density.

2.3.4 CESIUM CHLORIDE GRADIENTS

For sub-cloning experiments or cDNA library construction directly into M13 vectors, highly purified double-stranded vector DNA (i.e. RF DNA) is required, but most rapid mini-prep isolation procedures/kits result in a mixture of double-stranded and single-stranded circular DNA. This causes a problem when cloning, because the SSC-DNA does not digest with most restriction enzymes, and therefore provides a huge background of 'empty clones' in any ligation reaction. Therefore, for these types of experiments, the SSC-DNA which is present in the bacterial cell pellet must be removed from the RF-DNA, and this

is accomplished by running the material on at least two sequential ethidium bromide/cesium chloride gradients.

Cesium chloride/ethidium bromide (CsCl/EtBr) stock solution (34.48g of CsCl, 32ml of TE buffer, 3.2ml of ethidium bromide stock solution at 10mg/ml) was prepared and used to top-up the cesium chloride gradients. The DNA extracted in section 2.3.3 (8ml in TE) was adjusted for the CsCl gradients by adding 8.62 g of CsCl and 800 µl of ethidium bromide stock solution (10mg/ml in dH_2O). The solution was transferred to ultraclear QuickSeal polyallomer ultracentrifuge tubes (Beckman Coulter). The tubes were filled to the top with CsCl/EtBr stock solution and balanced to within 0.1g. The tubes were heat sealed and centrifuged at 330,000 RCF at 20°C for 16h (67,000rpm on Optima L-100 XP centrifuge with a type 70TI rotor by Beckman Coulter).

After centrifugation the tubes were removed from the centrifuge rotor so as to not disturb the bands in the gradient. The DNA bands were visualized using a 365nm UV lamp; two distinct bands formed with the lower (more dense) band representing the covalently closed circular (CCC) double stranded DNA and the upper (less dense) band representing the single stranded (or relaxed double stranded) DNA. A hole was made in the very top of the tube with a 16-gauge needle in order to allow air to enter, and the lower DNA band was removed using a 16-gauge needle and 10ml syringe. This was accomplished by piercing the sidewall of the tube, pointing the needle bevel upwards, and slowly

withdrawing the band from below while gently sweeping the needle over the lower interface of the band. The pulled band was then re-run on a new CsCl gradient by simply placing the contents of the syringe in a fresh QuickSeal tube, filling it with CsCl/EtBr stock solution, balancing, heat sealing, and centrifuging as before. The lower band was again removed from the second gradient and transferred to a 15ml screw capped polypropylene tube.

To remove the ethidium bromide from the CCC RF-DNA solution, the samples were repeatedly extracted with 10ml of H_2O-saturated n-butanol. The tubes were mixed thoroughly and centrifuged at 400 RCF to separate the organic and aqueous phases (1,500rpm on Alegra 6KR kneewell centrifuge with a GH-3.8 rotor by Beckman Coulter). The n-butanol clean-up step was repeated until the organic phase showed no more pink color. During this process, the aqueous phase was maintained at a volume of ≈5ml by addition of dH_20 as needed. In order to precipitate the DNA, the aqueous phase was transferred into fresh round-bottomed polypropylene tube and 555 µl of 3M NaOAc and 5.5 ml of isopropanol was added. The sample was mixed thoroughly and stored at -80°C for 4 h.

Samples were thawed on ice and centrifuged at 5,000 RCF for 30 minutes at 4°C (7,000 rpm on Avanti J26 XP centrifuge with a JLA-16.250 rotor). The DNA pellet was clearly visible, and the supernatant was discarded and the pellet completely drained by aspiration. One ml of ice-cold 95% ethanol was added,

the intact DNA pellet loosed from the tube by gentle flicking, and then gingerly poured en masse into a 1.5ml Eppendorf tube. This was pelleted in a microfuge, rinsed twice with 1 ml of ice-cold 70 % ethanol each time, and then dried briefly by placing the open tube in a fume hood. The final dried DNA was resuspended in 400 µl of TE buffer. The yield of RF-DNA was analyzed by measuring the optical density of several serial dilutions; for the M13JFE7s prep used in Chapter 3, at a dilution of 10^{-3} the OD_{260} was 0.176, giving a concentration of 8.8 µg/µl and a total yield of 3.52 mg.

2.4 EUKARYOTIC CELL CULTURE

APD, WT49, and WT 51 cells are all Epstein Barr virus (EBV) transformed human B-lymphocyte cell lines which are homozygous for chromosome 6 (note that the letters denoting the cell lines are not abbreviations—in some cases they may be the initials of the person from whom the cell line was made, but these names have been kept anonymous). These lines have been used previously in the Elliott lab-- in the case of APD to construct BAC libraries which were sequenced across the entire MHC region through a collaborative project with the MHC Haplotypes Consortium. All cell lines were originally obtained from the Coriell Institute for Medical Research in Camden, New Jersey. Cells were grown in RPMI 1640 (Roswel Park Memorial Institute) medium containing L-glutamine (Sigma-Aldrich), and supplemented with 10% fetal bovine serum (Fisher), and penicillin-streptomycin (100units/ml and 100ug/ml respectively) (Invitrogen).

Although all 3 cell lines were grown at one stage, the RNA purification pilot experiments were performed mainly on WT49 because it grew very well, whereas the RT-PCR sequence based typing methods were developed using RNA from APD, since the precise DNA sequence for HLA-DPB1 was know for this cell line.

2.5 PURIFICATION OF TOTAL CELLULAR RNA USING GUANIDINIUM THIOCYANATE-PHENOL-CHLOROFORM EXTRACTION

The following guanidine thiocyanate-phenol-chloroform extraction procedure is a modification of the method originally described by Chomczynski P and Sacchi N. (1987). It was obtained from the DNA microarray core laboratory of The Institute of Plant and Microbial Biology (IPMB) at the Academia Sinica, Taiwan [2].

Reagents	Conc†	Quantity added
Phenol (water saturated)	38%	380 ml
Guanidine Thiocyanate	0.8M	118.16 g
Ammonium Thiocyanate	0.4M	76.12 g
Sodium acetate (pH 5.0)	0.1M	33.4 ml of 3M stock
Glycerol	5%	50 ml
DEPC water		To final volume 1L

†Concentration in final solution

The solution was stored in an amber bottle at 4°C to minimize oxidation of phenol [2]. Guanidine-thiocyanate phenol chloroform solution is also commercially available as TRIzol (Invitrogen) and will be referred to as such in the rest of the protocols.

Ribonucleic acid (RNA) is an extremely susceptible substance, due to the large quantity of ribonuclease (RNAse) present in the environment [1]. All experiments performed to extract RNA were conducted under the most controlled conditions in order to prevent any possible RNAse contamination. Solutions used for the RNA extraction procedure were made from water treated with diethyl-pyrocarbonate (DEPC) (Ambion) at a concentration of 0.1% to inactivate any contaminating RNAses, and then autoclaved.

2.5.1 EXTRACTION OF TOTAL CELLULAR RNA FROM SMALL NUMBERS OF CELLS

This protocol was used for pilot experiments, typically involving one to a few million cells grown *in vitro*. The PBS washed cell pellet was lysed by adding an appropriate volume of TRIzol solution (usually 1ml), pipetting up and down, and incubating at room temperature for 5 minutes. Chloroform (0.2 volumes of the amount of TRIzol solution used) was added to the homogenate to cause separation of the aqueous and organic phases, the tube vortexed for 15 seconds, and allowed to sit at room temperature for 3 minutes. The sample was centrifuged at 1,200 RCF for 15 minutes at 4°C (4,000rpm on Eppendorf 5415C centrifuge with stock rotor), and the aqueous phase carefully removed making sure not to disturb the interface. The volume of aqueous phase was measured and an equal volume of 70% ethanol (made with DEPC treated H_2O) was added, and the tube mixed by inverting. The sample was loaded into an RNeasy MINI

column (Qiagen), centrifuged at 8,000 RCF for 30s (10,000rpm on Eppendorf 5415C centrifuge with stock rotor) and the flow through discarded.

To clean and then elute the RNA bound to the column, the RNeasy cleaning procedure (provided by Qiagen in the RNeasy MINI kit) was followed. RW1 buffer (700 µl) was placed in the column, it was centrifuged at 8,000 RCF for 30s, and the flow through discarded. This was followed by two consecutive washes with 500 µl RPE buffer each time. The column was transferred to a fresh 1.5 ml Eppendorf tube and 50 µl of DEPC-treated H_2O premixed with 0.5 µl Superase-IN RNAse inhibitor (Ambion) was added to the column. After a 2 minute incubation at room temperature, the column + 1.5 ml Eppendorf tube were centrifuged at 8,000 RCF for 1 minute to elute the RNA (10,000rpm on Eppendorf 5415C centrifuge with stock rotor). A (second) 0.5 µl aliquot of Superase-IN RNAse inhibitor (Ambion) was added to the eluted RNA, and the samples were stored at -80°C.

2.5.2 EXTRACTION OF TOTAL CELLULAR RNA FROM LARGE NUMBERS OF CELLS

This protocol was used for larger quantities of cells, typically 10- to 100-million , and details of a representative experiment are provided. Fifty million WT49 cells were pelleted by centrifugation at 234 RCF (1,200 rpm on Alegra 6KR Kneewell centrifuge with GH-3.8 rotor) and washed with 5ml of fresh RPMI 1640 /L-glutamine media. The cells were again pelleted, supernatant discarded, and cells were resuspended in 2ml of phosphate buffered saline (PBS). To also have

material for genomic DNA, 600 µl of the cell slurry in PBS was transferred to a 1.5ml Eppendorf tube, and centrifuged at 4,000 RCF for 5 minutes at 4°C (7,000 rpm on Eppendorf 5415C centrifuge with stock rotor). The supernatant was aspirated and the cell pellet was stored at -80°C for future genomic DNA analysis.

The remaining cells (1.4 ml) in PBS were transferred to a labeled 15 ml conical screw capped polypropylene tube, and 10ml of TRIzol solution was rapidly added. Samples were capped, inverted four times to rapidly mix and incubated at room temperature for 5 minutes. Two ml of chloroform (i.e. 0.2 volumes) was added to induce separation of the aqueous and organic phases, and the mixture which was vortexed for 30 seconds to mix contents thoroughly. Samples were centrifuged at 2,100 RCF for 5 minutes at 4°C (3,600 rpm on Alegra 6KR Kneewell centrifuge with GH-3.8 rotor) and the aqueous phase recovered and re-extracted with 6 ml chloroform. The final aqueous phase was transferred to a fresh 15 ml conical screw capped polypropylene tube, and an equal volume of 70% ethanol (prepared with DEPC treated H_2O) was added. The solution was mixed thoroughly and then loaded onto an RNeasy MIDI kit column (QIAGEN). Because these columns hold a maximum of 4 ml, loading was accomplished by adding the first 4 ml aliquot of the TRIzol + ethanol solution, spinning, discarding the flow through, adding the next 4 ml aliquot, etc. until the entire volume had been passed through the column. The RNA was washed and eluted according to the instructions provided with the RNeasy MIDI kit (4ml wash with RW1 buffer, followed by two 2ml washes with RPE buffer). Finally the RNA was eluted with

250 µl of DEPC treated dH_2O which had been premixed with 2.0 µl of Superase-IN RNAse inhibitor (Ambion).

Total RNA was analyzed using an EXPERION automated electrophoresis system (BioRad), and representative electropherograms are shown in Figures 2.2 and 2.3 A.

2.6 EXTRACTION OF CYTOPLASMIC RNA FROM CELL LINES USING NP40 LYSIS

Purification of cytoplasmic RNA from eukaryotic cells is an effective way to obtain RNA without any contaminating partially transcribed or partially spliced mRNAs, and without any contaminating DNA—all of which are found in the nucleus [3]. For our project it also represented a potentially useful method to obtain all of the DNA plus all of the (useful) RNA from a relatively precious patient sample. Although we ended up not using the method on patient samples, a representative experiment using a model cell line is described in what follows.

Fifty million WT49 cells were harvested by centrifugation for 10 min at 234 RCF (1,200 rpm on Alegra 6KR Kneewell centrifuge with GH-3.8 rotor) in a 15 ml falcon tube. The cell pellet was resuspended in 5ml of fresh RPMI 1640/L-glutamine media, re-pelleted, washed in 2.5ml of fresh RPMI 1640/L-glutamine media, and the final pellet resuspended in 120 µl of ice-cold low-salt Tris buffer (20mM Tris-HCl, 10mM NaCl, 3mM MgOAc) pre-mixed with 8 µl of 200mM vanadyl sulfate:nucleoside complexes (made in-house) [4]. Once the cell pellet

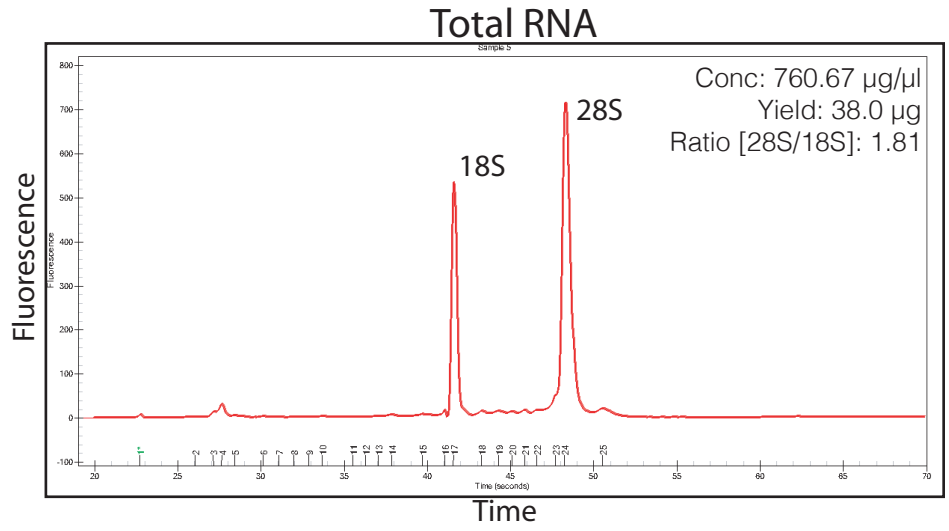


Figure 2.2 - Total RNA extracted from WT49 cells and analyzed using the Experion capillary gel electrophoresis system (Bio-Rad). Ten million WT49 cells were resuspended in 2ml of PBS and lysed by adding 10ml of TRIzol solution. Aqueous and organic phases were induced to separated by addition of 2 ml of chloroform, and the recovered aqueous phase was re-extracted with 6 ml of chloroform. The final aqueous phase was diluted with an equal volume of 70% ethanol, loaded onto a RNeasy MIDI kit column (QIAGEN), and the column washed according to the manufacturers instructions. RNA was finally eluted in 250 µl of DEPC treated water, and 2µl of this eluate was analyzed by capillary gel electrophoresis. The electropherogram shows two distinct peaks that are specific to the 18S and 28S ribosomal RNA species. The ratio of the peaks gives an indication of the quality of the RNA extracted. The electropherogram indicates that this RNA sample is of very high quality and has no background contamination.

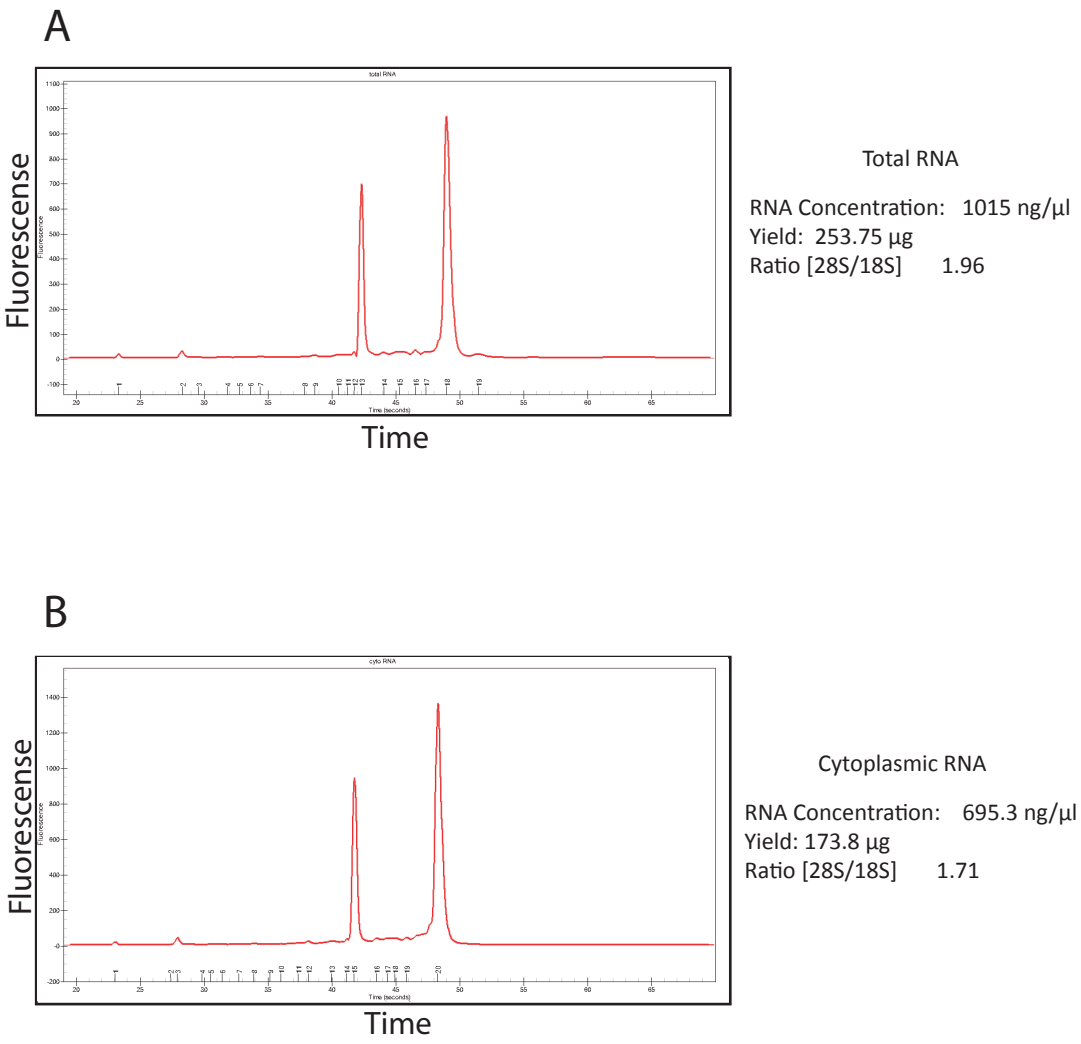


Figure 2.3 Electropherograms of total RNA compared to cytoplasmic RNA extracted from WT49 cells. Fifty million WT49 cells were used in each experiment. Total RNA was extracted using the same methods and reagent volumes as in Figure 2.2. In contrast, cytoplasmic RNA was extracted using NP-40 lysis in the presence of vanadyl-adenine complexes to preserve the RNA, and the nuclei were pelleted prior to further processing (see text for details). The lower 28S/18S ratio of the cytoplasmic RNA suggests it has undergone slight degradation.

was thoroughly resuspended, 40 µl of ice cold lysis buffer I (low-salt Tris buffer containing 5% sucrose and 1.2% NP-40 detergent) was added, and the contents mixed thoroughly by finger vortexing. The NP-40 mediated lysis step breaks open the cytoplasmic membrane while maintaining the nuclear membrane intact. The nuclei were recovered by centrifugation at 1,200 RCF for 5 minutes at 4°C (3,800 rpm in an Eppendorf 5415C centrifuge with stock rotor) and the supernatant (≈168 µl) removed and placed in a clean labeled 2.0 ml Eppendorf tube.

To extract the RNA from the supernatant, 600 µl of ACE buffer (50mM NaOAc, 10mM EDTA, pH 5.1), 40 µl of 10% Sodium dodecyl sulfate (SDS), and 800 µl of ACE saturated Phenol were added, and the tube were heated to 50°C for 5 minutes with vigorous mixing. The aqueous phase was recovered by centrifugation at 4,000 RCF for 10 minutes at 4°C (7,000 rpm on Eppendorf 5415C centrifuge with stock rotor), re-extracted with 800 µl of ACE saturated Phenol/chloroform (the chloroform containing 6% isoamyl alcohol), and finally with 800 µl of chloroform-isoamyl alcohol. To the final aqueous phase, 0.05 volumes of 2M NaOAC and then 2.5 volumes of 95% ethanol were added. The solution was mixed by inverting and placed at -80°C overnight. RNA was recovered by centrifugation at 10,000 RCF for 30 minutes at 4°C (10,000rpm on Eppendorf 5415C centrifuge with stock rotor), the RNA pellet rinsed with ice-cold 70% ethanol, dried briefly and resuspended in 100 µl of DEPC-treated H_2O premixed with 1.0 µl of Superase-IN RNAse inhibitor (Ambion). The cytoplasmic

RNA was analyzed by EXPERION automatic gel electrophoresis, and compared to total RNA obtained using the TRIzol extraction method (Figure 2.3).

2.7 EXTRACTION OF TOTAL CELLULAR RNA FROM HUMAN BLOOD

This method was first established using blood samples from healthy volunteers, and subsequently applied to patient samples. Four 10 ml EDTA vacutainers were drawn from each patient (or volunteer), and the blood from these was evenly distributed into three 15ml conical falcon tubes which were centrifuged at room temperature for 15 minutes at 1,500 RCF (3,000rpm on Alegra 6KR Kneewell centrifuge with GH-3.8 rotor). This 'hard spin' caused the erythrocytes to move to the lower half of the tube, the plasma to move to the upper half, and the white blood cells (lymphocytes, monocytes, neutrophils, etc.) to be deposited as a gray layer lying just on top of the erythrocyte layer, at the interface (so called 'buffy coat layer'). The entire buffy coat layer was gently drawn up into a Pasteur pipette, while attempting to minimize the number of erythrocytes taken up at the same time. However, since some erythrocytes were always inadvertently drawn up, an erythrocyte lysis step is added to obtain the pure buffy coat cells.

The three buffy coat interfaces from each patient (~500 µl/interface; these can all be drawn up using the same Pasteur pipet) were pooled to a prechilled 50 ml PP conical falcon tube on ice, after which 50 ml of pre-chilled Erythrocyte Lysis (EL) buffer (155mM NH_4Cl, 10mM $KHCO_3$, 0.1mM Na_2EDTA – pH

7.4; [5]) was added. Note that the pH of the EL buffer must be 7.4 to cause optimal erythrocyte lysis, and that the pH of the solution will change over time and must be checked before each use. Tubes were placed on ice for 30 minutes to lyse nearly all red blood cells, with mixing every 10 minutes by gently inverting. The tubes were centrifuged at 234 RCF for 10 minutes at 4°C (1,200rpm on Alegra 6KR Kneewell centrifuge with GH-3.8 rotor) to pellet the buffy coat cells, and the hemoglobin-containing supernatants removed by decanting and then aspiration. Cell pellets were gently resuspended in 10ml of ice-cold EL buffer (to lyse any residual erythrocytes), transferred to fresh 15 ml conical screw capped polypropylene tubes, and re-centrifuged at 234 RCF for 10 minutes at 4°C. The final supernatant was discarded and the buffy coat cells resuspended in 2ml of ice cold PBS.

For future genomic analysis, 600 µl of each cell slurry in PBS was transferred to a 1.5ml Eppendorf tube, which was centrifuged at 2,000 RCF, 4°C for 5 minutes (5,000 rpm on Alegra 6KR Kneewell centrifuge with GH-3.8 rotor). The supernatant was aspirated and the cell pellet stored at -80°C for future use.

The remaining cells (1.4ml) were transferred to a labeled 15ml conical PP tube, and TRIzol lysis and RNeasy MIDI column (Qiagen) purification performed as described in section 2.5.2 above. The RNA was eluted from the column with 200 µl of DEPC-treated H_2O premixed with 1 µl of Superase-In RNAse inhibitor

(1U /μl), and analyzed on an EXPERION automated electrophoresis system (Figures 2.4 and 2.5).

The quality and quantity of the RNA obtained depended on several factors. When patient samples were analyzed, we found that different patients gave different quality RNA preparations. One key concept is that the reagents used must always be fresh in order to maintain the RNA quality. For example, Figure 2.5 shows two different patient samples each extracted from 40ml of blood using the process outlined above. Patient 42 total RNA is of ideal quality, whereas patient 46 has a poor RNA profile, indicating marked degradation. Upon further analysis, it was discovered that the erythrocyte lysis buffer which was used to process Patient 46's blood sample had an elevated pH (this had shifted during storage), making it slow and ineffective at inducing lysis of the erytrocytes, many of which made it through to the TRIzol lysis tube. The quality of the RNA preparation obviously has downstream consequences. Of the 16 different patient blood samples that we processed for total cellular RNA, only the Patient 46 sample failed to yield an assignable HLA sequence after RT-PCR amplification and cloning (see Chapter 4).

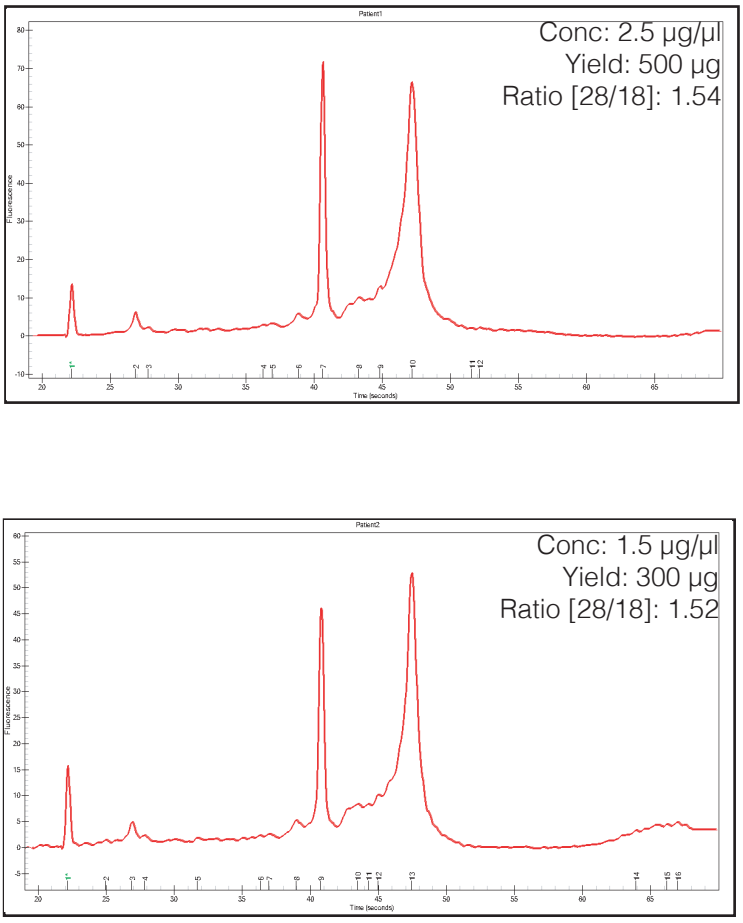


Figure 2.4 Electropherograms of total RNA extracted from two different human blood samples. Samples were given by volunteers, and therefore could be processed under ideal conditions (i.e. immediately after collection). Four 10 ml EDTA vacutainers were collected from each donor. The blood was centrifuged to obtain buffy coat cells, and these were pooled for each patient and contaminating erythrocytes lysed by addition of NH4Cl-containing buffer. The final erythrocyte-free buffy coat cells were resuspended in 2 ml PBS and 0.6 ml of the cell suspension transferred to a fresh tube, centrifuged and the drained pellet stored at -80˚C for future genomic DNA extraction. The remaining 1.4 ml of the cell suspension was lysed by adding 10ml of TRIzol solution, and total RNA was extracted as described in Figure 2.2 and the text. The reagents used for erythrocyte lysis and RNA extraction were freshly made and the total RNA obtained was of relatively high quality based on the appearance of the 18S and 28S ribosomal RNA peaks. This procedure appears to be very reproducible when samples are processed immediately after collection.

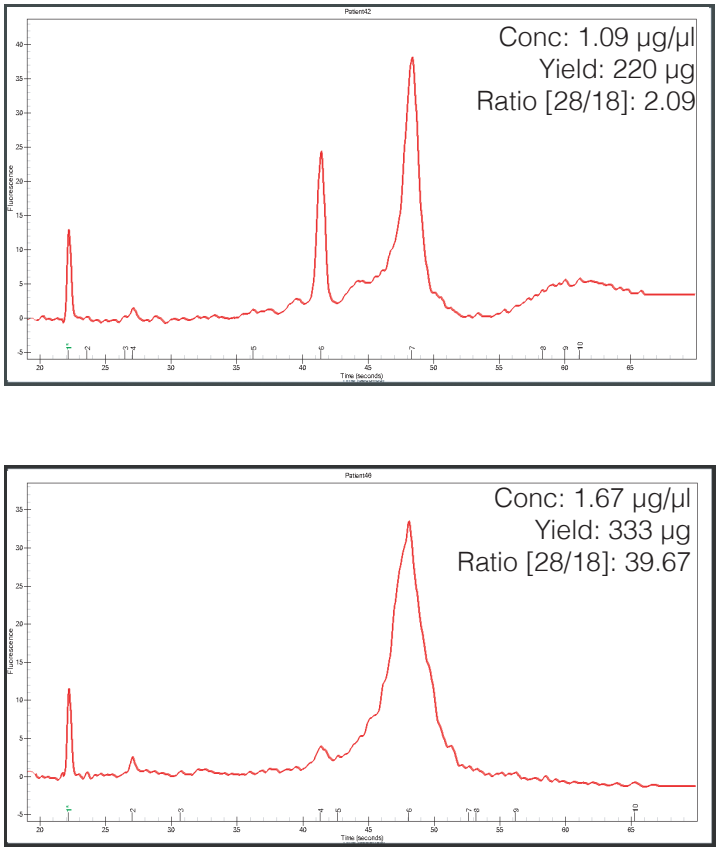


Figure 2.5 Electropherograms of total RNA extracted from two different patient blood samples, showing varying quality of RNA obtained. Patient blood samples were collected over several months as patients with allergic contact dermatitis to para-phenylenediamine were identified in the university dermatology clinic. Because of this different patient samples ended up being treated somewhat differently, for example room temperature storage times prior to processing varied between 15 minutes and 3 hours. These and other factors likely influenced the quality of the total RNA obtained from particular patients. The upper electropherogram is representative of the very best patient RNA samples obtained (patient 42; with a nearly ideal profile), whereas the lower electropherogram shows the worst patient RNA sample obtained (Patient 46; both 18S and 28S bands are highly degraded). In retrospect we discovered that the poor quality of the Patient 46 sample was due to elevated pH of the erythrocyte lysis buffer (when freshly made the pH is 7.4, but this had drifted up to 8.0 with simple storage at room temperature). The elevated pH resulted in minimal lysis of the erythrocytes and possibly also caused direct degradation of the RNA within the buffy coat cells. In fact Patient 46 is the only RNA sample where RT-PCR failed to yield an assignable HLA sequence (see Chapter 4).

2.8 References

1. Sambrook, J. and D.W. Russell, *Molecular cloning : a laboratory manual*. 3rd ed. 2001, Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.
2. Chomczynski, P. and N. Sacchi, *Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction.* Anal Biochem, 1987. **162**(1): p. 156-9.
3. Carninci, P., et al., *Cytoplasmic RNA extraction from fresh and frozen mammalian tissues.* Biotechniques, 2002. **33**(2): p. 306-9.
4. Gray, J.C., *The inhibition of ribonuclease activity and the isolation of polysomes from leaves of the french bean, Phaseolus vulgaris.* Arch Biochem Biophys, 1974. **163**(1): p. 343-8.
5. Verheyen, G., et al., *Comparison of four mechanical methods to retrieve spermatozoa from testicular tissue.* Hum Reprod, 1995. **10**(11): p. 2956-9.
6. Robinson, J., et al., *IMGT/HLA database--a sequence database for the human major histocompatibility complex.* Tissue Antigens, 2000. **55**(3): p. 280-7.
7. Robinson, J., et al., *The IMGT/HLA database.* Nucleic Acids Res, 2009. **37**(Database issue): p. D1013-7.
8. Sahal, D. and Y. Fujita-Yamaguchi, *Protein kinase assay by paper-trichloroacetic acid method: high performance using phosphocellulose paper and washing an ensemble of samples on flat sheets.* Anal Biochem, 1987. **167**(1): p. 23-30.
9. Holmberg, A., et al., *The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures.* Electrophoresis, 2005. **26**(3): p. 501-10.

CHAPTER 3

MODEL EXPERIMENTS TO SPECIFICALLY CAPTURE M13/HLA-DQ cDNA CONSTRUCTS DIRECTLY OUT OF SOLUTION

3.1 INTRODUCTION

As discussed in chapter 1, sequence-based HLA typing (SBT) requires that specific genomic or cDNA fragments encoding the HLA molecules to be typed are somehow obtained in pure form so that they can be subjected to DNA sequencing. One of challenges we set for this research project was to consider if certain novel non-PCR based DNA cloning technologies could be applied to the 'purification problem' that must be solved before SBT can be performed. Specifically we set out to explore if oligonucleotide-driven solution-based hybridization and capture methods, applied to HLA-cDNA sequences cloned into M13-bacteriophage targets might be applicable to achieve SBT. As explained in chapter 1, this approach is similar to that used by the 'GeneTrapper cDNA Positive Selection Kit' formerly marketed by Invitrogen/Life Sciences, but for reasons already described we chose to use M13 vectors rather than phagemid vectors. This chapter describes a number of model experiments that were performed to explore this approach to SBT.

For our model experiments we chose to focus specifically on the HLA-DQ genes, namely DQA1 and DQB1. The work was done in a series of steps, and the chapter has been organized on this basis. The steps included: 1. Exploring what

possible ‘capture from solution’ strategies might be utilized; 2. Identifying oligonucleotide sequences capable of specifically hybridizing to all DQA1 or all DQB1 alleles; 3. Subcloning several different DQA1 and DQB1 cDNA fragments into an M13 vector in order to generate single-stranded circular DNA (SSC-DNA) targets for capture; 4. Purchasing (or in some cases synthesizing special forms of) the capture oligonucleotides designed in 2.; 5. Performing the model solution phase hybridization and capture experiments using the oligonucleotides and M13 SSC-DNAs made for this purpose. Although the model experiments did not lead to a new method for SBT, results obtained illustrate the strengths and weaknesses of the approaches taken, and the technologies explored may never-the-less have utility to solve other biological problems where use of PCR might not be appropriate.

3.2 TWO APPROACHES TO CAPTURE SPECIFIC DNA SEQUENCES DIRECTLY OUT OF SOLUTION

Two possible approaches to capturing specific M13 target sequences directly out of solution are shown in Figures. 3.1 and 3.2. The first protocol makes use of oligonucleotides covalently bound to the surface of macroscopic beads such as the oligonucleotide affinity supports (OAS) marketed by Glen Research Inc. This product is a polystyrene/PEG co-polymer bead upon which an oligonucleotide is directly synthesized. The second approach uses biotinylated oligonucleotides in solution and secondary capture with streptavidin beads.

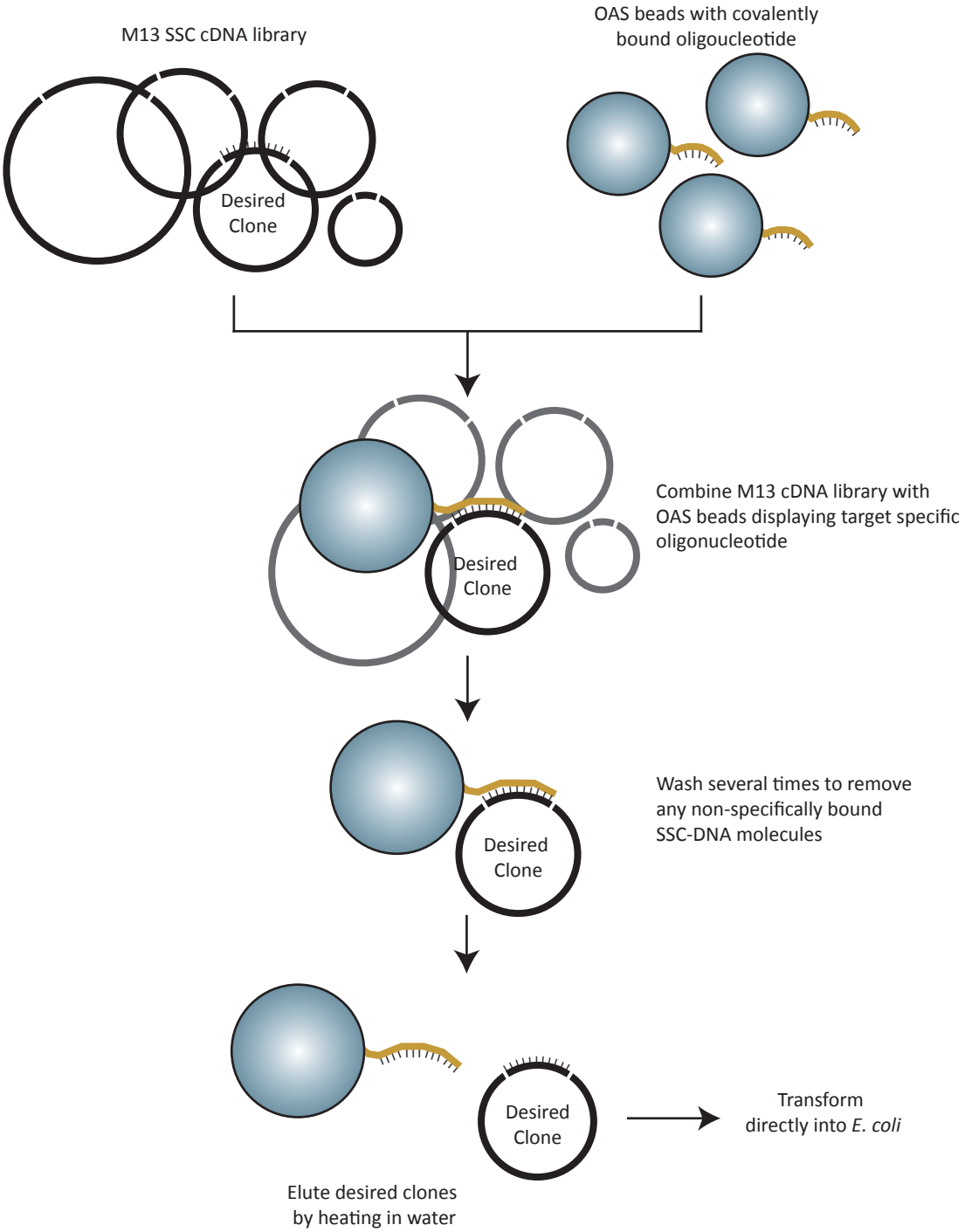


Figure 3.1 Overview of proposed procedure to capture specific M13 target sequences directly out of solution using oligo affinity supports (OAS) beads. The capture oligonucleotide is designed to specifically hybridize to the desired sequence element (i.e. desired clone) in the M13 library. Solid phase DNA synthesis of this oligonucleotide is done directly on the surface of the OAS beads, and it remains permanently attached to the surface of the beads following deprotection.

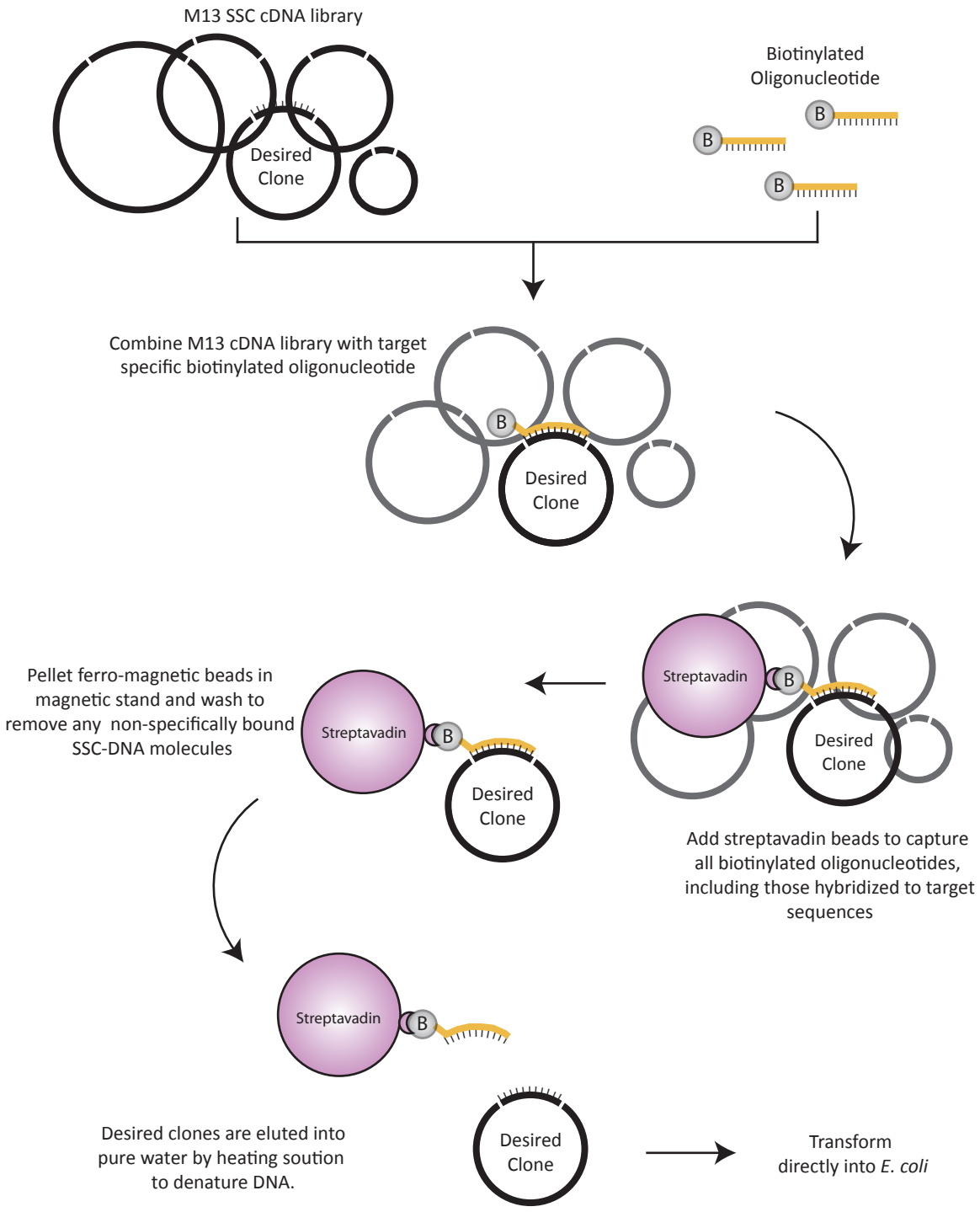


Figure 3.2 Overview of proposed procedure to capture specific M13 target sequences directly out of solution using biotinylated oligonucleotides and ferro-magnetic streptavadin beads. This involves a two-step process: 1) biotinylated oligonucleotides hybridize to complementary sequences in the M13 library, and 2) all biotinylated oligonucleotides (hybridized or free-floating) are captured onto ferro-magnetic streptavadin beads, and these are pelleted and washed using a magnetic stand. The selected SSC-DNA molecules are eluted by heating beads in water and transformed directly into *E.coli*.

With OAS approach (Figure 3.1), a complex cDNA library is cloned directly into an M13 vector and purified SSC-DNA generated from that library. This purified material is incubated in solution in the presence of the OAS beads, which have on their surface a specific oligonucleotide with sequence complimentary to the desired target clone. After hybridization of the OAS-oligo beads to the M13 DNA, the beads are separated from the solution by either filtration or centrifugation. They are washed to remove any nonspecifically bound M13 SSC-DNA molecules, and the final washed beads are heated in pure water to elute the desired clone.

The OAS approach suffers from the fact that the solid phase oligonucleotides do not have a high effective concentration in solution, and also because they may be located close to the surface of the beads and therefore not be especially accessible for hybridization. On the other hand, the approach potentially has an advantage in simplicity, and also at least the theoretical possibility of simply reusing the beads over and over to capture desired clones from a series of different libraries. Since the accessible surface area of the beads is quite large and the beads should be quite mobile in solution (especially if tubes are placed on an agitator or roller wheel) we felt this approach was certainly worth trying. The OAS solid phase beads have been used primarily in the past to capture sequence-specific DNA binding proteins, but they have also been reported to be useful in hybridization experiments (Glen Research).

The second approach to capturing specific M13 target sequences out of solution makes use of a biotinylated oligonucleotide capable of hybridizing to the specific desired clones (Fig. 3.2). These biotinylated oligonucleotides are mixed with purified SSC-DNA made from a complex M13 cDNA library, and allowed to hybridize in solution. Following hybridization, streptavidin ferro-magnetic beads are added, and these capture all of the biotinylated oligonucleotides, including those hybridized to the target M13 SSC molecules. The ferro-magnetic beads are pelleted using a magnetic stand, after which they can be repeatedly washed to remove any non-specifically bound M13 molecules. The desired M13 target molecules are finally eluted by heating the streptavidin beads in water.

With either of the methods described above, the final eluted M13 SSC-DNA molecules can be transformed directly into *E.coli* and large quantities of the desired clone(s) can be generated. This final step differs from the GeneTrapper kit (see Chapter 1) in that the M13 SSC-DNA does not need to be converted to double-stranded form in order to achieve high transformation efficiencies. This is likely because SSC is the form of the phage DNA that enters the *E.coli* cell during infection, and host cell machinery rapidly converts it to double-stranded form in any case (see Fig. 1.6). Thus, transformation efficiencies – whether single-stranded or double-stranded M13 DNA is used, are relatively comparable. This differs from phagemid SSC-DNA such as used in the GeneTrapper kit, where converting from single stranded to double stranded form increases

transformation efficiencies 20- to 100-fold (GeneTrapper manual and our unpublished observations).

One other potential advantage of using M13 SSC-DNA over phagemid double stranded DNA constructs in these types of solution based hybridization and capture strategies is that with M13 vectors, the DNA selection process can be repeated in an iterative fashion more easily and rapidly. For example one cycle of selection, transformation, amplification in *E. coli*, and purification of 1X-selected/amplified SSC-DNA could likely be accomplished in about 24 hours; if necessary 2X-selected/amplified SSC-DNA could be generated in the subsequent 24 hour period, and so on.

3.3 IDENTIFICATION OF POTENTIAL BINDING SITES/SEQUENCES FOR CAPTURE OLIGONUCLEOTIDES FOR DQA1 AND DQB1 GENES

As previously mentioned, nearly all of the HLA genes (and hence their transcribed mRNAs) show marked polymorphism across the human population, and especially so at their 5' ends, which makes identification of exon I or exon II sequence elements conserved within all known alleles of a given HLA gene quite challenging. For these model experiments, our efforts were focused on identifying conserved sequence elements within the HLA-DQA1 and HLA-DQB1 cDNAs. The process was made easier by studying the allelic alignments for these genes available on the Anthony Nolan Research Institute public database (http://hla.alleles.org/alleles/text_index.html).

The allelic sequence alignments for DQA1 and DQB1 were analyzed to identify candidate hybridization sites for capture oligonucleotides. The candidate sites were identified based on the following criteria: 1. Location within a non-polymorphic region as near as possible to the 5' end of the mRNA; 2. Length of 18-25 bases; 3. Melting temperature of 50-62 ˚C; 4. Absence of predicted secondary structure.

Candidate oligonucleotide sequences were analyzed with the NCBI BLASTn tool (http://www.ncbi.nlm.nih.gov/blast/) to identify any possible matches to unrelated sequences within the human transcriptome. With one notable exception (Site 211B), the sequences that had no matches outside of their specific HLA gene were considered most ideal. Of all the sites identified, six sequences were considered to be superior candidates (Table 3.1): for the DQA1 alleles, Sites 001A, 103A, and 149A; and for the DQB1 alleles, Sites 110B, 211B, and 327B. The numbers given to each specific oligonucleotide identify the nucleotide position (i.e. number of nucleotides from the ATG) with respect to the reference sequences DQA1*01:01:01 and DQB1*05:01:01:01.

3.3.1 SITE 211B IS PRESENT IN NEARLY ALL MHC CLASS II BETA GENES

Our BLASTn searches indicated that the Site 211B sequence had a somewhat unusual property. This nucleotide sequence was found not only in almost all DQB1 alleles, but also in nearly all other classical MHC class II beta chain genes (i.e. HLA-DRB1 and HLA-DPB1). For this reason, we chose to

Table 3.1 Conserved sequences within HLA-DQA1 and DQB1 with potential to be target sites for capture oligonucleotides

Name	Target Gene	Sequence	Length (bases)	Calc Tm (50mM NaCl)	Blast hits
Site 001A	HLA-DQA1	5'-ATG.ATC.CTA.AAC.AAA.GCT.CTG-3'	21	58°C	HLA-DQA1, DQA2
Site 103A	HLA-DQA1	5'-GGT.GTA.AAC.TTG.TAC.CAG.T-3'	19	51.5°C	HLA-DQA1
Site 149A	HLA-DQA1	5'-C.CAT.GAA.TTT.GAT.GGA.GA-3'	18	50.0°C	HLA-DQA1, DQA2, MYCBP2
Site 110B†	HLA-DQB1	5'-AG.GAT.TTC.GTG.(T/C)(A/T)C.CAG.TT-3'	19	54°C-56°C	DQB1, DRB5
Site 211B‡	HLA-DQB1	5'-CGC.TTC.GAC.AGC.GAC.GTG.G-3'	19	61.5°C	DRB1, DRB3, DRB5, DQB1, DPB1
Site 328B†	HLA-DQB1	5'-GT(G/A).TGC.AGA.CAC.AAC.TAC.(G/C)AG-3'	21	54.4°C	DQB1, DQB2

† Mixed bases are present in the sites 110B and 328B.

‡ Site 211B contains a conserved sequence that is found in HLA-DQB1, HLA-DRB1 and HLA-DPB1.

Note: Oligonucleotides generated with sequences specific to a target gene will only hybridize with constructs containing the target gene cDNA (eg. An oligonucleotide with the sequence to site 103A (HLA-DQA1 target) will hybridize with an M13JFE7-DQ6A [HLA-DQA1] construct and will not hybridize with an M13JFE7-DQ8B [HLA-DQB1] construct).

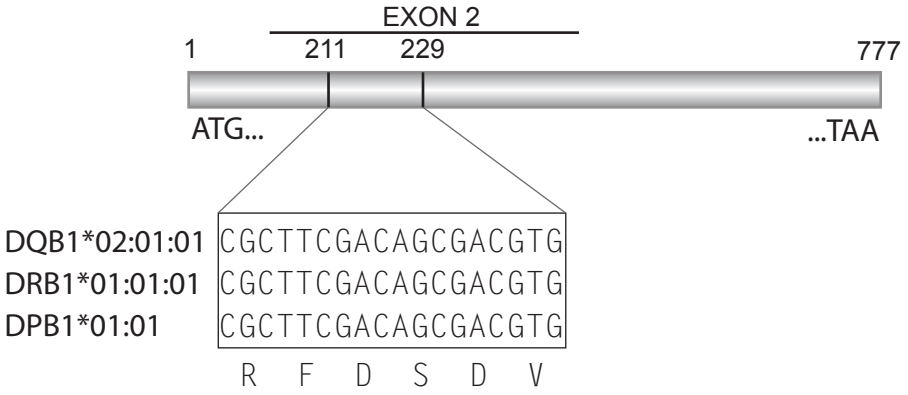
investigate the DNA and encoded peptide sequence further, and these results are shown in Figure 3.3 and Table 3.2.

Of the 1230 allelic (nucleotide) sequences known for DRB1, DQB1, and DPB1, only 29 non-matching alleles were identified, and all of these differed by only a single nucleotide (Figure 3.3 B). The sequence of Site 211B encodes the peptide "RFDSDV", which is obviously also highly conserved. In the protein structure, the peptide sequence occurs in a beta strand and a partial turn in the floor of the antigen-binding cleft (region indicated by the asterisk in the ribbon diagram shown in Figure 3.3 C). We can speculate that this part of the MHC class II beta chain might play some role which is required of all class II molecules, such as binding to the invariant chain. Many of the nucleotide polymorphisms in this region lead to no change (i.e. silent mutations) or to conservative changes in the peptide sequence (Table 3.2). It is also possible that the conserved nucleotide sequence may play some role in gene regulation and/or RNA splicing, but this is also speculative.

3.4 SUBCLONING DQA1 AND DQB1 cDNA FRAGMENTS INTO AN M13 VECTOR

In order to generate single stranded circular DNA molecules we required an M13 vector. For this purpose we used M13JFE7s, which was constructed from M13mp19 by deleting all non-M13 sequences and replacing them with a 570 bp polylinker + stuffer element, as detailed in Figure 3.4 and shown schematically in Figure 3.5 (J. Elliott and A. Atrazhev, unpublished). When

A.



B.

	DQB1		DPB1		DRB1	
	DNA	Protein	DNA	Protein	DNA	Protein
perfect match	153	104	149	128	902	691
one mismatch	2	2	1	1	26	13
two mismatches	-	-	-	-	-	-

C.

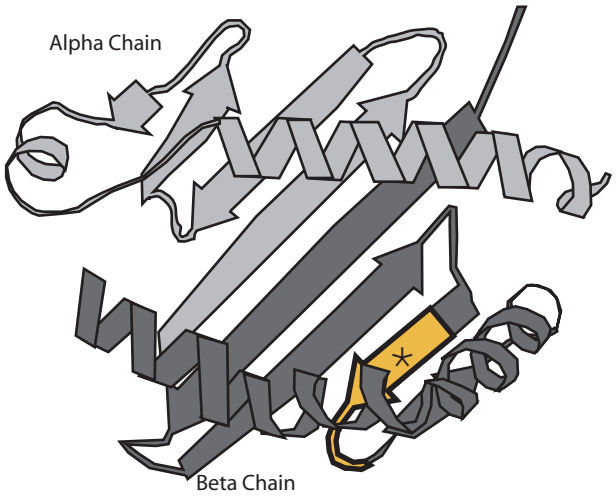


Figure 3.3 Site 211B nucleotide sequence is highly conserved among HLA class II beta chain genes, and the corresponding (highly conserved) peptide is located in the floor of peptide-binding cleft. A. Schematic showing location of 211B sequence in DQB1 cDNA, and encoded peptide in representative HLA class II beta chain genes. B. Numbers of perfect matches of nucleotide and protein sequence among all HLA class II beta chain genes (includes all DRB1's, but not DRB5's or DRB9's, etc). Note that the number of protein sequences is less than the number of DNA allelic sequences, since certain DNA sequences are assigned to different alleles, yet they encode the same protein. C. Ribbon diagram of structure of HLA-DQB1 indicating the location of the 211B peptide (adapted from Dai, S. et al). The peptide contributes a beta sheet and turn to the floor of the peptide-binding cleft. Of the 29 nucleotide sequences that are not a perfect match with 211B, 13 alleles contain silent mutations generating no protein change, 7 alleles produce neutral amino acid substitutions and 9 alleles generate non-conservative amino acid substitutions (see Table 3.2).

Table 3.2 List of nucleotide polymorphisms and amino acid substitutions in site 211B, found by searching all known HLA-DRB1, -DQB1, and -DPB1 DNA sequences

Allele	DNA sequence	Encoded peptide	Amino acid substitution
Oligo211	CGCTTCGACAGCGACGTG	RFDSDV	
DR*01:01:07	----------A	------	None
DR*01:01:09	----------T---	------	None
DR*01:01:12	----------C	------	None
DR*01:01:17	----------T	------	None
DR*01:02:05	----------A	------	None
DR*04:05:02	--G----------	------	None
DR*04:05:10	----------A	------	None
DR*11:01:13	----------T------	------	None
DR*12:02:05	----------A	------	None
DR*13:03:05	----------A	------	None
DR*15:01:10	----------T---	------	None
DR*15:01:15	-----T----------	------	None
DR*15:02:07	----------T---	------	None
DQ*05:06	-----G----------	-L----	F → L (conserv.)
DQ*05:07	-A----------	H-----	R → H (conserv.)
DR*04:33	-A----------	H-----	R → H (conserv.)
DR*10:01:01	----A----------	-Y----	F → Y (conserv.)
DR*10:01:02	----A----------	-Y----	F → Y (conserv.)
DR*10:01:03	----A----------	-Y----	F → Y (conserv.)
DR*10:02	----A----------	-Y----	F → Y (conserv.)
DP*120:01N[†]	------T----------	--Y---	D → Y (non-conserv.)
DR*04:96	------A----------	--N---	D → N (non-conserv.)
DR*09:11	----------G-	-----G	V → G (non-conserv.)
DR*11:61	------A----------	--N---	D → N (non-conserv.)
DR*11:99	----C----------	-S----	F → S (non-conserv.)
DR*13:91	T----------	C-----	R → C (non-conserv.)
DR*13:95	----------T-----	----Y-	D → Y (non-conserv.)
DR*15:56	------C----------	--H---	D → H (non-conserv.)
DR*15:55	----------G-	-----G	V → G (non-conserv.)

The sequences were identified by alignment of all known transcripts of HLA-DPB1, HLA-DQB1 and HLA-DRB1 genes (sequences from http://hla.alleles.org). The sequences with single nucleotide polymorphisms in the site of interest (site 211B) were further analyzed by identifying their translated peptides. The amino acid mutations identified were then considered conservative or non-conservative mutations depending on the properties of exchanged amino acids.

M13JFE7s RF DNA is digested with the restriction enzyme *Sfi* I and the stuffer fragment removed, it yields a vector with two asymmetric, non-complimentary, non-self cohesive ends (-GAG and –AGA; figure 3.5). This allows *Sfi* I inserts excised from other vectors with the same polylinker to be inserted into M13JFE7 in a specific orientation, and with high efficiency since circle closure of the empty M13JFE7 vector is prevented (i.e. only the 'stuffer-fragment' or another fragment with the proper 3'-overhanging ends [-CTC and –TCT] are capable of closing the circle; see Figure 3.5). For the case of cDNA fragments being inserted into *Sfi* I cut M13JFE7s, if the ATG start codon is nearer the –CTC end, then the SSC phage DNA will yield a strand which is anti-sense to the original mRNA, whereas if the ATG start codon is nearer the –TCT end, then it will yield SSC DNA making the sense strand (see also legend to Figure 3.4).

Fortunately, several HLA-DQA1 and HLA-DQB1 cDNA constructs were previously made in the Elliott lab in a plasmid vector, pCSD7s (Table 3.3, upper half). By design pCSD7s contains the same polylinker and *Sfi* I cloning sites as were originally made in M13JFE7s, therefore it was straightforward to subclone the HLA-DQ constructs from pCSD7 into M13JFE7 (Figure 3.5). This subcloning was accomplished using the gel slice ligation method, which is rapid, simple to execute, and very inexpensive. Because of these advantages, we felt it was worthwhile to provide a clear and detailed explanation of the method, in this case by describing a representative experiment.

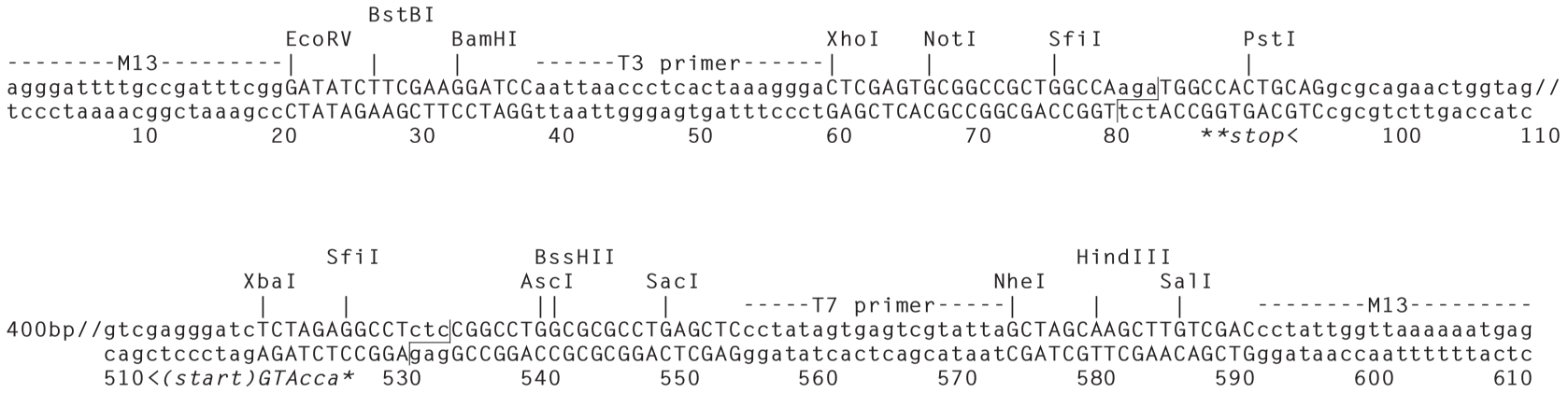


Figure 3.4 DNA sequence of M13JFE7s polylinker. The 570 bp segment inserted into the intergenic region of M13 contains a number of unique restriction sites as well as T3 and T7 priming sites which can be used for DNA sequencing or PCR amplification of cloned inserts. When this vector is digested with *Sfi* I it releases an approximately 450 bp insert (='stuffer fragment') and leaves two asymmetric, non-annealing, non-self cohesive 3' overhanging 'sticky ends' : --AGA and --GAG. These allow for directional cloning of any DNA fragments which have the complementary 3' overhanging ends : --TCT and --CTC. All of the DQ cDNA segments subcloned into this vector using the *Sfi* I strategy have their ATG (i.e. start codon; indicated *) near the *Sfi* I site with the GAG overhang, and their stop codon (**) near the *Sfi* I site with the AGA overhang. (Note that the Start codon* should be read on the lower strand in the 5' -> 3' direction, and that in the subcloned DQ cDNA's part of a Koziak consensus sequence is added just upstream of the ATG, the rest of the Koziak consensus sequence being contributed by the final CC of the adjacent *Sfi* I site, i.e. to yield ccaccATG.) As drawn, it is the upper stand of the polylinker which is found in the M13 virions, thus the DQ cDNA segments subcloned always yield single stranded DNA which is anti-sense to the original mRNA. Hence oligonucleotides derived from the sense strand will be able to hybridize with the DQ-derived single stranded circular (SSC) DNA molecules described in this chapter. Also, when sequencing SSC-DNA, it is the T7 primer which is required (anneals to the upper strand), and the sequence obtained starts with ATG and yields sense sequence. The Xho I site indicated is not unique in M13JFE7s, since it also occurs at positions 124 and 504 within the stuffer fragment, but it may be unique once the stuffer fragment is replaced with another DNA element.

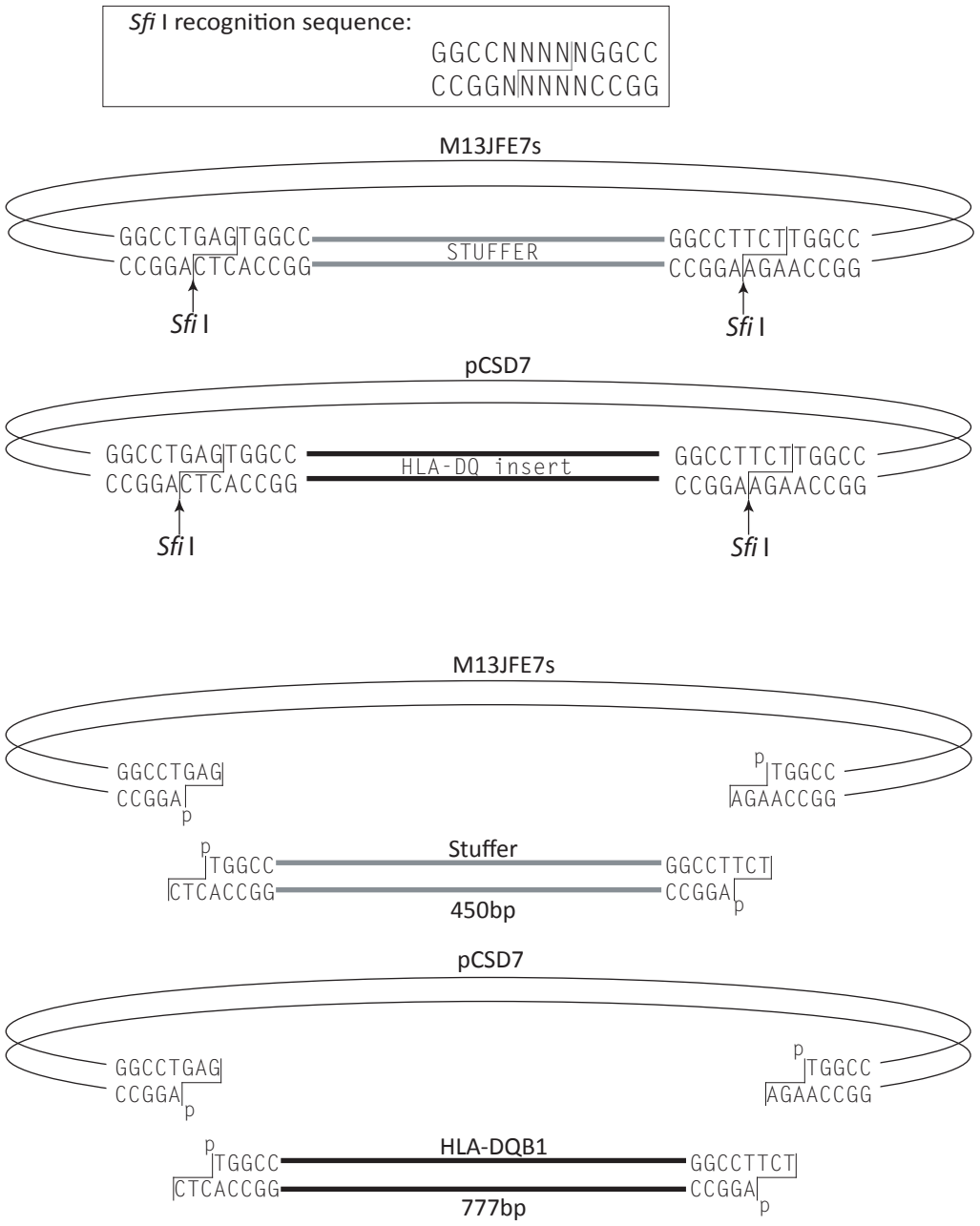


Figure 3.5 Schematic diagram showing details of *Sfi* I 'sticky ends' used in directional cloning strategy and sub-cloning of HLA-DQ cDNA inserts from the plasmid vector pCSD7 into M13JFE7s. Because the *Sfi* I sticky ends generated on the vector after digestion are non-annealing, the vector cannot close on itself, and because these ends are asymmetric, the DQ cDNA inserts are cloned in a specific known orientation. This cloning strategy leads to low or zero background of empty clones, and is highly efficient in generating clones with the desired insert in a defined orientation.

3.4.1 SUBCLONING USING THE GEL SLICE LIGATION METHOD

The double-stranded RF form of M13JFE7s was prepared as described in chapter 2, being careful to purify this away from contaminating single stranded M13JFE7s DNA (i.e. phage DNA still in the bacterial cell awaiting packaging) by running the DNA on two sequential cesium chloride/ethidium bromide gradients. If the single stranded M13JFE7s is not removed, the cloning will not work because there will be an overwhelming background of M13JFE7s pseudoplaques (i.e. 'empty clones') arising from the single stranded DNA which fails to cut with *Sfi* I, but still transforms with high efficiency.

One microgram each of the M13JFE7s vector, as well as pCSD7 constructs containing cDNAs for HLA-DQB1*02:02 (DQ2) and DQB*03:02 (DQ8) were digested to completion with the restriction enzyme *Sfi* I (New England Biolabs) following the manufacturers directions. As a control for the size of the pCSD7 vector and stuffer fragments, 1 µg of pCSD7s was also digested at the same time. Following digestion one-tenth volume of load dye containing 25 mM EDTA pH 8.0 was added to each reaction, and the tubes were heated to 70˚C for 10 minutes to inactivate any contaminating exonucleases. The digested fragments were separated on a 1.0% low melt point (LMP) agarose gel made with TBE buffer and cast containing 0.2 µg/ml ethidium bromide. Electrophoresis was continued for 90min at 70 volts, after which the DNA bands were immediately visualized by placing the gel on a 365nm UV transilluminator

(Figure 3.6; note that shorter wavelengths can mutate DNA, or damage the DNA and render it unclonable). The appropriate fragments were excised with a scalpel, any excess gel trimmed away, and the gel pieces placed in separate labeled 1.5ml Eppendorf tubes.

The Eppendorf tubes were centrifuged briefly, the volume of each gel slice was estimated and an equal volume of sterile dH_2O was added. Note that this dilution step makes it much easier to melt, mix, and pipet the agarose containing DNA solutions. The gel slices were melted by incubating tubes in a 70°C heating block for 10min with occasional mixing, and the ligation reactions set up according to the table below, using enzyme and 10X buffer supplied by New England Biolabs.

	Ligation Reaction			
Reagents added	Vector only (Neg. control)	Vector + stuffer (Pos. control)	Vector + DQ2B	Vector + DQ8B
Sterile dH_20	30.5µl	20.5µl	20.5µl	20.5µl
10X Lig. Buffer	4µl	4µl	4µl	4µl
Sfi I cut M13JFE7*	5µl	5µl	5µl	5µl
Stuffer*	-	10µl	-	-
DQ2B insert*	-	-	10µl	-
DQ8B insert*	-	-	-	10µl
T4 DNA Ligase	0.5µl	0.5µl	0.5µl	0.5µl
Total Volume	40µl	40µl	40µl	40µl

* indicates volume of diluted/melted gel slice added directly to the ligation reaction

The ligation reactions were simply left on the bench overnight (≈16h), after which 10 µl of each reaction was transferred into 100 µl of XL2-Blue chemically competent *E.coli* cells, and the tube contents gently but thoroughly mixed by flicking with the index finger. The tubes were incubated on ice for 10

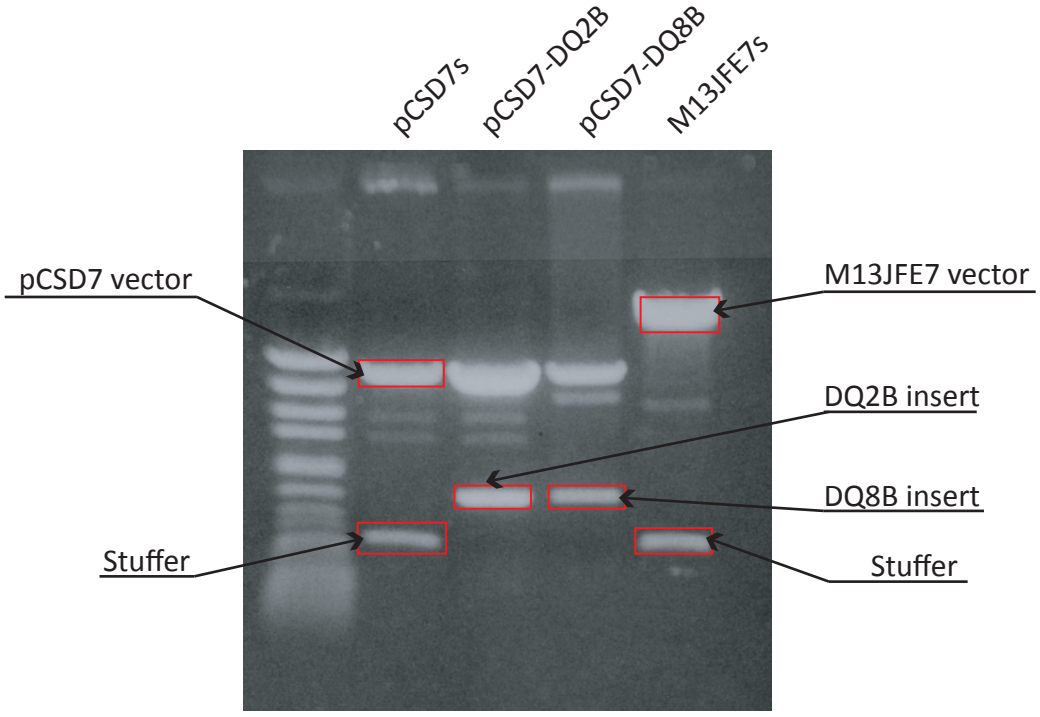


Figure 3.6 Low melting point agarose gel showing the various DNA bands which were excised and used to set up the gel slice ligation reactions. The left-most lane contains DNA size markers, and all other lanes contain the indicated constructs digested with Sfi I. The gel was made with 0.5X TBE buffer, and is cast containing 0.2 µg/ml of ethidium bromide. Photograph was taken on a 365 nm UV trans-illuminator, since this wavelength causes minimal damage to the ethidium-stained DNA. The ethidium does not need to be removed from the gel slices prior to using them in the ligation reactions.

minutes, heat shocked at 37°C for two minutes, returned to ice for 10 minutes, and then 1 and 3 µl of each transformation reaction was plated on a lawn of *E.coli* XL2-Blue (see chapter 2 for further details; note that unlike plasmid transformation, M13 transformation requires not outgrowth step). After overnight incubation at 37°C, pseudoplaques were counted and the numbers obtained are summarized below.

Volume of transformation plated	Vector only (Neg. control)	Vector + stuffer (Pos. control)	Vector + DQ2B	Vector + DQ8B
1µl	0	58	39	52
3µl	0	119	139	148

These results indicate that the cloning was highly successful since *Sfi* I cut vector (M13JFE7) ligated by itself generated no pseudoplaques. Note that some investigators have suggested that inclusion of the negative and positive control ligations all the way through to the transformation, plating, and pseudoplaque counting steps represents an unnecessary expenditure of effort. However, we find that in practice this approach requires little additional time and effort, but it is very useful since the numbers of transformants obtained in each of the ligation reactions immediately tells if the subcloning has been successful, as well as allowing instant trouble-shooting if the experiment has been unsuccessful (e.g. no transformants in the positive control indicates failure of the ligase, equal numbers of transformants in the positive and negative controls indicates contamination with of the gel box or buffer with exonuclease, etc.).

Individual well-isolated pseudoplaques were picked using sterile 200 µl pipette tips and placed in 1ml of sterile dH_2O to make phage stocks. To verify that the pseudoplaques picked contained the desired clones, 1 µl of phage stock was used as an inoculum into a fresh 2 ml culture of XL2-Blue bacteria. The cultures were grown at 37°C for 5 hrs. on a roller wheel to ensure optimum phage growth. RF DNA was extracted from the cell pellets using a Wizard miniprep alkaline lysis kit, and constructs analyzed by mapping with *Sfi* I. The constructs were also sequenced using T7 and T3 primers on RF DNA. Both mapping and DNA sequencing showed the construct to be correct. In the particular experiment described above, M13JFE7 clones were produced for the alleles HLA-DQB1*02:02 (M13JFE7-DQ2) and HLA-DQB*03:02 (M13JFE7-DQ8), but other DQ cDNAs were also inserted into the M13JFE7s vector as summarized on Table 3.3. Note that all of these constructs generated SSC phage DNA which is anti-sense to the original mRNA, and therefore all of the capture oligonucleotides for our model experiments would need to be of the sense strand.

3.5 PRODUCTION OF CAPTURE OLIGONUCLEOTIDES

Although six different potential target sites for capture of DQ sequences had originally been identified (Table 3.1), oligonucleotides corresponding to only a subset of these sites were actually made, and these are summarized in Table 3.4. Purchase of the T3 and T7 sequencing oligonucleotides, as well as the

Table 3.3 Plasmid and M13 vectors and derived DQ cDNA constructs

Plasmids	Description	Source
pCSD7s	pUC18 backbone, Amp^r, polylinker from M13JFE7s	D. Denney / J. Elliott
pCSD7-DQ2α	pUC18 backbone, Amp^r, DQα*05:01	C. Compson / J. Elliott
pCSD7-DQ2β	pUC18 backbone, Amp^r, DQβ*02:01	C. Compson / J. Elliott
pCSD7-DQ6α	pUC18 backbone, Amp^r, DQα*01:02:01	C. Compson / J. Elliott
pCSD7-DQ8α	pUC18 backbone, Amp^r, DQα*03:01	C. Compson / J. Elliott
pCSD7-DQ8β	pUC18 backbone, Amp^r, DQβ*03:02	C. Compson / J. Elliott
M13JFE7s	M13 backbone, polylinker shown in Figure 3.4	A. Atrazhev / J. Elliott
M13JFE7-DQ2α	M13 backbone, DQα*05:01	C. Garcia-Batres / J. Elliott
M13JFE7-DQ2β	M13 backbone, DQβ*02:01	C. Garcia-Batres / J. Elliott
M13JFE7-DQ6α	M13 backbone, DQα*01:02:01	C. Garcia-Batres / J. Elliott
M13JFE7-DQ8α	M13 backbone, DQα*03:01	C. Garcia-Batres / J. Elliott
M13JFE7-DQ8β	M13 backbone, DQβ*03:02	C. Garcia-Batres / J. Elliott

Table 3.4 Oligonucleotides used for DNA sequencing and model solution hybridization/capture experiments

Name	Modification	Sequence	Length (bases)	Tm (50mM NaCl)
T7 primer*	unmodified	5′ - TAATACGACTCACTATAGGG - 3′	20	47.9°C
T3 primer*	unmodified	5′ – AATTAACCCTCACTAAAGGGA - 3′	21	50.8°C
C′-211B-S-dT**	unmodified	5′ - CCACGTCGCTGTCGAAGCGGCCAAGAAGGCC $(T)_{25}$ - 3′	56	Unknown
OAS-103A**	Covalently attached OAS	5′-GGT.GTA.AAC.TTG.TAC.CAG.T - **OAS**	19	51.5°C
OAS-211B**	Covalently attached OAS	5′ - CGC.TTC.GAC.AGC.GAC.GTG.G - **OAS**	19	61.5°C
OAS-328B**	Covalently attached OAS	5′ - GT(G/A).TGC.AGA.CAC.AAC.TAC.(G/C)AG - **OAS**	21	54.4°C
BTN-103A*	Biotinylated	**BTN** - GGT.GTA.AAC.TTG.TAC.CAG.T - 3′	19	51.5°C
BTN-211B*	Biotinylated	**BTN** - CGC.TTC.GAC.AGC.GAC.GTG.G - 3′	19	61.5°C

*Purchased from Integrated DNA Technologies Inc.

**Synthesized in house

biotinylated oligonucleotides was straightforward. However, we were unable to find a company that was willing to synthesize oligonucleotides on OAS beads, and therefore this synthesis had do be done in house. Once we had the DNA synthesizer up and running, we also chose to synthesize the longer unmodified C'-211B-S-dT oligonucleotide in house as well.

Oligonucleotide affinity supports (OAS; polystyrene/PEG copolymer beads upon which oligonucleotides can be synthesized but where the 3' end remains permanently attached) as well as phosphoramidites and solvents for our Applied Biosystems PCR Mate EP model 391 DNA synthesizer were all purchased from Glen Research. Protocols were based on recommendations from Glen Research and followed the PCR mate instruction manual. Trityl groups were collected automatically (in a fraction collector) for each nucleotide synthesis step, and the absorbance of these fractions indicated DNA synthesis had proceeded efficiently. Unfortunately we could not remove the oligonucleotides from the OAS beads, so the trityl fractions were the only indication we had of the quality of the synthesis. Following synthesis the OAS beads were dried, transferred from the synthesis cartridge to a glass vial, and resuspended in 3ml of fresh ammonium hydroxide (NH_4OH). The vial was tightly capped, sealed with parafilm, and incubated in a 55°C oven overnight to cleave the protecting groups. The ammonium hydroxide was decanted and any last traces removed by drying the beads in a vacuum centrifuge. Once completely dry, the beads were re-suspended in 500 µl of TE buffer, transferred to a 1.5ml Eppendorf tube,

pelleted in a microfuge, drained and resuspended in 1 ml of fresh TE. This wash step was repeated 8 times and the beads finally resuspended in 1ml of TE and stored at -80°C ready for use.

Note that all OAS oligonucleotides were synthesized using the equivalent of 0.5 µMoles of beads, but the DNA synthesizer was programmed to run a 1.0 µM scale synthesis. In this way the phosphoramidites and all other reagents would be in a 2-fold higher molar excess than usual, which we anticipated would drive synthesis to be more complete at each step. Thus our final beads ended up with a total of 0.5 µMoles (or 500 nMoles) of oligonucleotide on their surface.

3.6 MODEL EXPERIMENT TO CAPTURE M13JFE7-DQ8B SSC-DNA ONTO OAS-211B BEADS

Having generated the necessary reagents, we were somewhat uncertain how to proceed with our initial model experiments using the OAS oligonucleotides. In order to make the readout fast and straightforward, we decided in the first instance to simply use agarose gel electrophoresis to detect the capture of SSC-DNA molecules out of solution. The first experiment simply looked to see if any capture would occur, using OAS-211B beads and M13JFE7-DQ8B single stranded circular (SSC) DNA.

The OAS-211B bead stock was mixed thoroughly and 10 µl of the slurry (displaying ≈5 nmoles of oligonucleotide) was transferred to a labeled 1.5 ml Eppendorf tube. This was centrifuged at 2,000 RCF (5,000 rpm in an Eppendorf

5415C centrifuge with stock rotor) for 2 min, the supernatant removed and the beads thoroughly resuspended in 100 µl of hybridization solution (1X Denhardt's solution, 0.1% SDS and 5X saline sodium phosphate EDTA [SSPE]). Two µl of a solution of M13JFE7-DQ8B (containing 236ng or ≈1 x 10^{-4} nmoles of SSC-DNA) was added and the bead-DNA mixture placed on a rocker in a 50°C oven for 16h (calculated T_m for 211-B is 61.5˚C; hybridization temperature was chosen to be just below T_m to ensure an ideal hybridization). The tube was removed from the oven, allowed to cool to room temperature for 30min, and the beads pelleted by centrifugation as above. The supernatant was removed and saved for agarose gel analysis, 100 µl of 1X SSPE solution was added to the drained beads, and they were resuspended and incubated at 50°C for 10min. The centrifugation, draining, resuspension, and 50°C incubation steps were repeated a total of 3 times, and the 300 µl of wash material pooled. Finally, the beads were resuspended in 100 µl of TE and placed at 65°C for 10 min. The beads were pelleted and the supernatant containing the (presumably) eluted SSC-DNA was collected. The beads were placed in 50 µl of TE buffer for possible further analysis.

The supernatant, pooled washes, and eluted fraction were all analyzed by agarose gel electrophoresis (Figure 3.7). If the capture procedure was successful, at least part of the SSC-DNA should have been transferred from the supernatant to the eluted fraction. However, whereas there was detectable SSC-DNA remaining in the supernatant, we saw no visible band in the eluted fraction, suggesting that no significant capture by the beads had occurred. To be sure

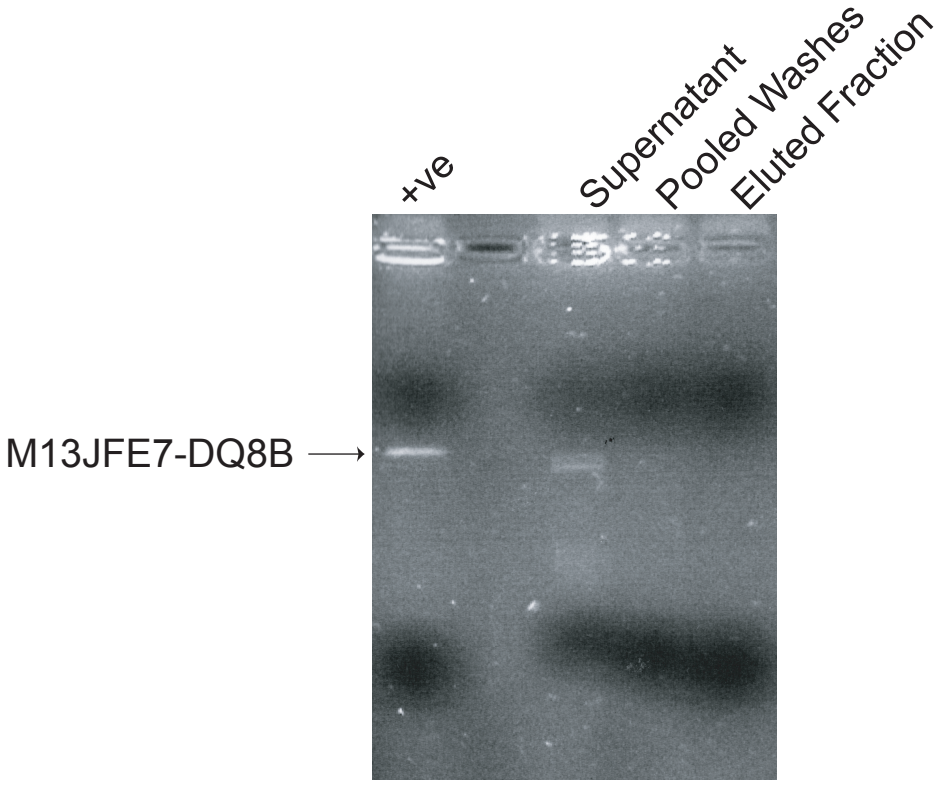


Figure 3.7 OAS-211B beads failed to capture detectable amounts of M13JFE7-DQ8B SSC-DNA. The first lane shows the amount of M13-SSC DNA (236ng) used in the original capture reaction. The supernatant and eluded fractions both had a volume of 100 µl, and 35 µl (≈1/3 of each fraction) was run on the gel. The pooled washes were in a total volume of 300 µl, and again 35 µl of this was run on the gel. The 1% regular agarose gel was made using 0.5X TBE buffer and was cast containing ethidium bromide. As expected, the high salt concentration in the hybridization solution caused the band in the supernatant fraction to run with a slightly higher mobility than the band seen in the positive control lane.

that there was no remaining SSC-DNA on the beads, they were pelleted, reuspended in dH_2O and heated to 90°C for 10min, but again no detectable DNA was released (data not shown). Variations of this experiment were repeated twice more using longer incubation times (3 and 5 days) but yielded similar results.

Failure of the OAS-211B beads to capture any SSC-DNA lead us to speculate that perhaps the 211B oligonucleotides were so close to the bead surface (and so relatively immobile) so that the very large (and slowly diffusing) SSC-DNA molecules were simply unable to hybridize to their targets to any extent over the incubation times that we were using. Also we considered the possibility that the oligonucleotides on the surface of the OAS-211B beads might be of poor quality. To explore these ideas further we designed a second experiment to test if the OAS-211B beads could capture a complementary oligonucleotide out of solution.

3.6.1 MODEL EXPERIMENT TO CAPTURE A SOLUBLE COMPLEMENTARY OLIGONUCLEOTIDE ONTO OAS-211B BEADS

The essential idea of this experiment is illustrated in Figure 3.8. A 56-base oligonucleotide (C'-211B-S-dT) was designed to contain a sequence element complementary to the 211B oligonucleotide, but also with two other sequence elements that would contribute additional length while also making the oligo

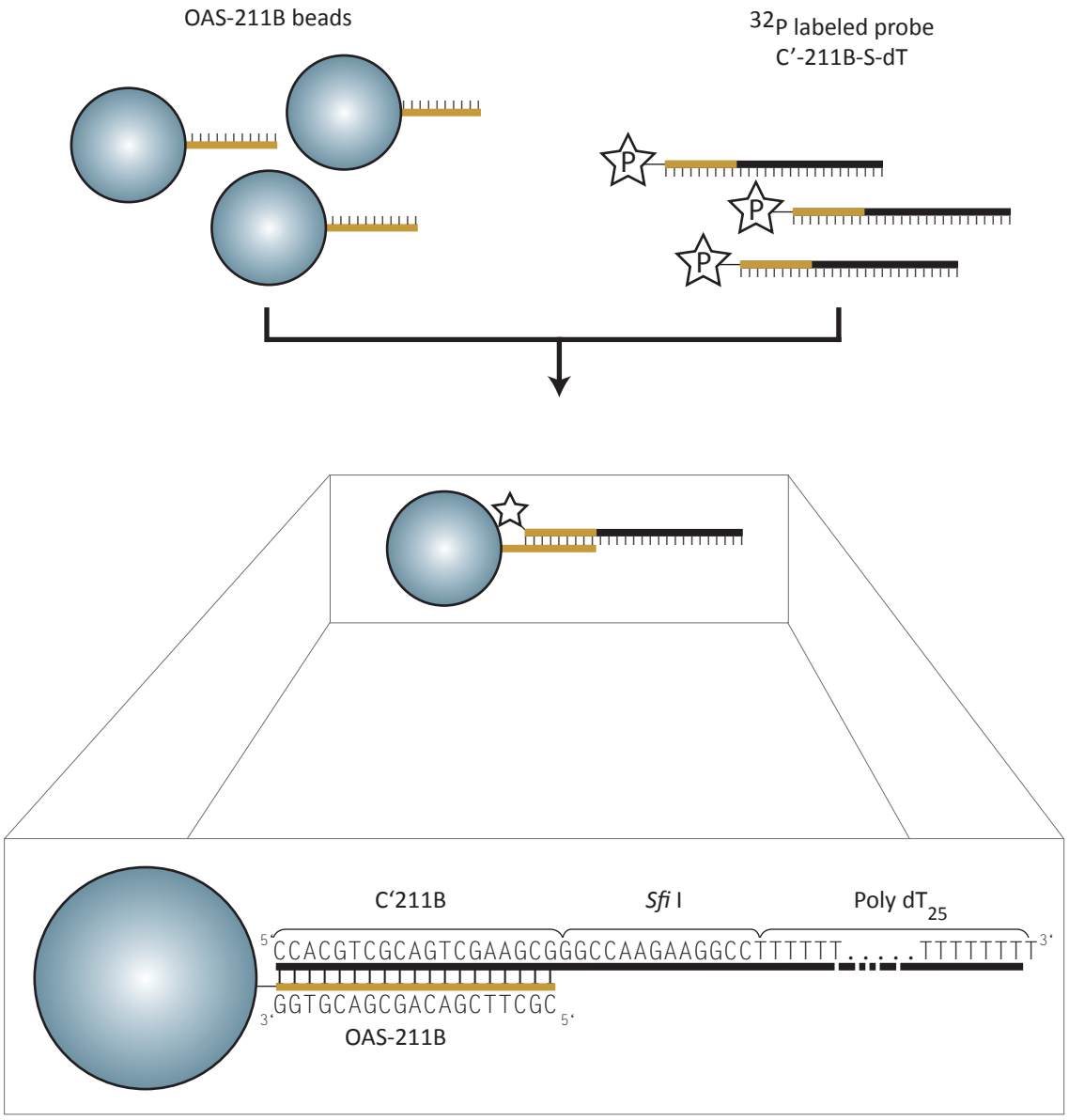


Figure 3.8 Schematic diagram of experiment to test capture of a soluble complementary oligonucleotide by OAS-211B beads. The soluble oligonucleotide is radioactively labeled at the 5'-end using T4 polynucleotide kinase and [γ-32P] ATP, and then purified by gel exclusion chromatography. An aliquot of this material can then be incubated with the OAS-211B beads, and the quantity of radioactivity which is transferred from the liquid phase to the beads measured at different times. Rather than using a minimal sized 19-nucleotide complimentary oligonucleotide, we chose to use a longer construct, since we felt this would somewhat more accurately reflect the behavior of the SSC-DNA molecules, and also because the oligonucleotide generated could potentially be used for a second purpose (e.g. priming first strand cDNA synthesis as a first step to making directional cDNA libraries using an *Sfi* I cloning strategy).

potentially useful for a second purpose (e.g. priming first strand cDNA synthesis as a first step to making directional cDNA libraries in vectors like M13JFE7s).

The oligonucleotide was radioactively labeled with ^{32}P at the 5' end using γ-^{32}P-adenosine triphosphate (ATP) and T4 polynucleotide kinase. The phosphorylation reaction contained 20 pmoles of C'-211B-S-dT oligonucleotide, 1 pmoles of $\gamma^{32}P$-ATP and 19 pmoles of ATP in a total volume of 20 µl, and was set up as recommended by the enzyme supplier (New England Biolabs). After incubating for 1h at 37°C, the kinase enzyme was deactivated by adding 1 µl of 500 mM EDTA (pH 8.0) and extracting with an equal volume of chloroform.

To separate the labeled oligonucleotide from any unincorporated $\gamma^{32}P$-ATP, the reaction was run over a Sephadex G50-fine 'drip column' run in TE buffer and assembled in a short Pasteur pipette plugged at the lower constriction with a small piece of glass wool. The pipet was filled with pre-swollen column matrix so as to leave the top ≈5 mm of the pipet empty, and it was run in an intermittent fashion by waiting until the column stops dripping at the bottom, moving the pipet tip to the next collection tube, adding 100 µl of TE to the top of column, waiting until the column stops dripping again, moving the pipet tip to the next collection tube, adding 100 µl of TE to the top of column, etc.

To determine at which fractions the labeled oligonucleotide would elute, an identical trial column was first run, loading 20 pmoles of C'-211B-S-dT oligonucleotide from an identical phosphorylation reaction but lacking the γ-^{32}P-

ATP. A total of sixteen 100 μl fractions were collected in Eppedorf tubes, and OD_{260} readings obtained using a nanodrop spectrophotometer showed a clear peak (data not shown). A virtually identical peak was obtained when the γ-^{32}P-ATP-containing phosphorylation reaction was run over a parallel drip column, but this time analyzing the column fractions by Cherenkov counting. The three fractions with the highest counts-- (which corresponded to the three fractions from the trial column which gave the highest OD_{260} readings) were pooled and used for further experiments. Trichloroacetic acid (TCA) precipitation analysis using a filter paper method [1] indicated that 87% of the ^{32}P within the pooled fractions was TCA precipitable (i.e. incorporated into the oligonucleotide).

To test the capture of the ^{32}P-labeled C'-211B-S-dT oligonucleotide by OAS-211B beads, we used the same solutions and hybridization/wash/elution volumes as described above for testing the capture of SSC-DNA. However in this case 4 μl of the OAS-211B (or OAS-103A) bead stock slurry (displaying ≈2 nmoles of oligonucleotide) were each mixed with ≈2 pmoles of the ^{32}P-labeled C'-211B-S-dT oligonucleotide (i.e. 1/10th of the pooled column fractions). Two different hybridization times were tested (1 and 3 days), after which the beads were centrifuged and washed exactly as for the experiment using SSC-DNA.

Supernatant, washes, eluted fraction and beads after elution were all counted using a LS6500 liquid scintillation counter, and results are shown in Table 3.5. The OAS-103A (negative control) beads appeared to bind very little

oligonucleotide, with eluted fraction and final beads showing essentially background counts. In contrast there appeared to be some specific binding of the ^{32}P-labeled oligonucleotide to the OAS-211B beads, and the amount of bound oligonucleotide increased with increasing incubation times. Also since about 50% of the counts remained on the beads even after elution, it appears that our elution protocol (TE buffer, 65˚C x 10 minutes) is not optimal. However, overall these results suggest that the OAS-211B beads are relatively inefficient at capturing out of solution even a small highly diffusible target (only about 10% of the counts were specifically captured after 3 days of incubation). Considering that the C'-211B-S-dT oligonucleotide is roughly 130 times smaller than the M13JFE7-DQ8B SSC-DNA, we concluded that the OAS-oligonucleotide approach would not be viable to capture the much larger SSC M13 DNA constructs. This led us to shift our focus away from OAS beads and towards the biotin/streptavidin protocol.

3.7 MODEL EXPERIMENTS TO CAPTURE SPECIFIC SSC-DNAs FROM SOLUTION USING BIOTINYLATED OLIGONUCLEOTIDES AND STREPTAVIDIN BEADS

This approach is summarized schematically in Figure 3.2, and involves a two step process whereby biotinylated oligonucleotides are first incubated with SSC-DNA molecules, and then captured using streptavidin beads. Any SSC-DNA molecules which have specifically hybridized to the oligonucleotides are also

Table 3.5 Radioactivity present in various fractions after incubating a ^{32}P-labelled soluble complementary oligonucleotide with two different OAS beads

	OAS-211B		OAS-103A	
Capture	1 day	3 day	1 day	3 day
Supernatant	11665 cpm	7986 cpm	10188 cpm	12469 cpm
Wash1	90 cpm	709 cpm	272 cpm	69 cpm
Wash2	33 cpm	103 cpm	46 cpm	15 cpm
Wash3	19 cpm	23 cpm	24 cpm	25 cpm
Elution	224 cpm	575 cpm	32 cpm	31 cpm
Beads	194 cpm	502 cpm	25 cpm	47 cpm

OAS-211B is complementary to the labelled oligonucleotide, whereas OAS-103A is not. Amount of radioactivity associated with the various fractions was measured by adding Ready-Safe liquid scintillation cocktail and counting in a Beckman LS6500 liquid scintillation counter.

captured at the same time. The biotin modification is known to not interfere with hybridization of the oligonucleotides to their target constructs [2]. In theory a variety of different streptavidin beads might be used, but for the current experiments we used ferro-magnetic streptavidin beads (Dynabeads M280 from Invitrogen). Although expensive, somewhat fragile, and having a tendency to melt at higher temperatures, Dynabeads have the advantage that they can be easily and rapidly pelleted and washed in a magnetic stand.

As with our OAS-oligonucleotide experiments, we initially used agarose gel electrophoresis to detect the capture of SSC-DNA molecules out of solution. Because we could achieve much higher effective concentrations of the capture oligonucleotides (the biotinylated oligonucleotides being considerably smaller and having a markedly higher diffusion coefficient than the OAS-oligonucleotides), we hoped to obtain more quantitative results which would be obvious on agarose gels. Once positive results were achieved using agarose gels, we went on to use transformation and analysis of relative numbers of blue and white pseudoplaques as a second method to detect capture of specific SSC-DNA molecules. For these experiments we used a mixture of two different SSC-DNAs, one giving rise to blue pseudoplaques and the other to white. Although time constraints prevented more complete development of these latter experiments, they did provide a more stringent test of the biotinylated oligonucleotide-streptavidin bead capture approach.

The Dynabeads are stored at 4°C and have a tendency to aggregate and settle out over time. Before use they must be resuspended by gentle swirling while avoiding the creation of bubbles, and then the necessary aliquot removed and washed in "binding and washing buffer 1" (BW1; consisting of 10mM Tris-HCl, 1mM EDTA, 1M NaCl). In a typical experiment (e.g. where capture from three separate hybridization tubes was required; see below) 60 μl (600 μg) of bead suspension was transferred to a 1.5ml Eppendorf tube, and the tube placed on a magnetic stand for 2 minutes to pellet the beads. The supernatant was discarded, and the beads resuspended in 150 μl of BW1 buffer by gentle pipetting. The slurry was incubated at room temperature for 2 minutes, and the beads re-pelleted on the magnetic stand for 2 minutes. The supernatant was discarded, and the wash process with 150 μl of BW1 buffer repeated two more times to remove any trace of bead storage buffer. The final washed pellet was resuspended in 30 μl of "binding and washing buffer 2" (BW2; 10mM Tris-HCl, 1mM EDTA, 2M NaCl) and stored at room temperature until use [3].

The initial experiments involved setting up 3 parallel aliquots of M13JFE7-DQ6A SSC-DNA using 0.2 ml thin-walled PCR microtubes. The tubes contained either: 1) no oligonucleotide, 2) a specific complementary oligonucleotide (BTN-103A), or 3) a non-hybridizing negative control oligonucleotide (BTN-211B). Eight microlitres of 62.5 ng/μl purified M13JFE7-DQ6A SSC-DNA (dissolved in TE buffer; total DNA added 500 ng or ≈2.1 x 10^{-4} nmoles) and 1 μl of 100mM $MgCl_2$ were added to each tube. Finally 1 μl of the appropriate oligonucleotide

(representing 250 ng or ≈4 x 10^{-2} nmoles; dissolved in TE buffer), or alternately 1 μl TE buffer (for the 'no-oligo' tube) was added, and tube contents mixed thoroughly. To promote denaturation followed by annealing, the tubes were placed in a thermocycler (Biometra T-Gradient) and heated to 95°C for 2 minutes, followed by 3 minutes at 50°C.

Following the 3 minute incubation at 50°C, the hybridization reaction tubes were removed from the thermocycler, allowed to cool to room temperature for 5 min, and then 10 μl of the washed Dynabead slurry was added to each tube. These were placed on a roller wheel in a 37°C warm room and rotated overnight (≈16 hr). The following morning the tubes were placed in a magnetic stand for 2 minutes and the supernatant removed and saved for analyses (See Figure 3.9). Beads were washed twice with 100 μl of BW1 buffer each time, and the resulting bead pellets resuspended in 100μl of BW1 buffer. This slurry was transferred to a fresh labeled 0.2 ml Eppendorf tube (this step was to prevent carryover of SSC-DNA which might have bound to the wall of the original tubes), the beads pelleted, final BW1 wash removed, and the resulting pellet resuspended in 20 μl of dH_2O.

The tube containing the beads in dH_2O was placed in a thermocycler (Biometra T-Gradient) and incubated at 70°C for 2 minutes, then 50°C for 3 minutes. (We reasoned that the higher temperature would allow the captured

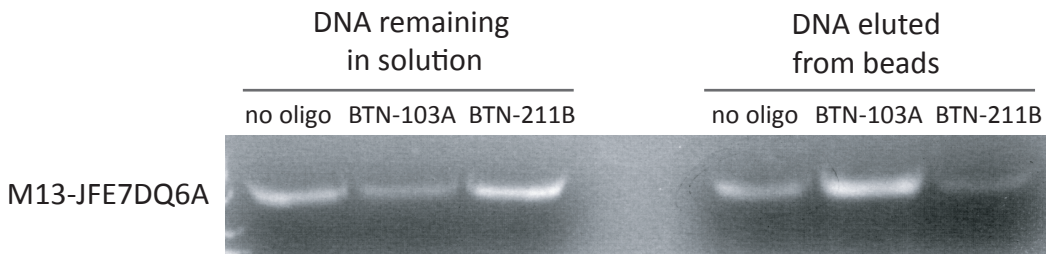


Figure 3.9 Result of model experiment to capture M13JFE7-DQ6A constructs out of solution using biotinylated oligonucleotides and streptavadin beads. To test the specificity of the solution hybridization procedure, M13JFE7-DQ6A constructs were hybridized with specific target oligonucleotide (BTN-103A), non-specific oligonucleotide (BTN-211B) or no oligonucleotide. The parameters of the protocol are outlined in the text. The DNA remaining in solution corresponds to the DNA that was not captured by the protocol. The DNA eluted from the beads corresponds to the DNA that was captured by the beads. The increased capturing capacity of the specifc oligonucleotide (BTN-103A) indicates the captures are working with some specificity, but the minimal captures in the non-specific oligonucleotide (BTN-211B) and when no oligonucleotide was used indicate that a small fraction of constructs are non-specifically binding to the streptavidin beads. These experiments were done without pre-blocking the streptavidin ferro-magnetic beads.

SSC-DNA to dissociate from the bead-bound oligonucleotides, and the lower temperature incubation would possibly allow the Dynabeads to stabilize/become more solid after the stress of the 70°C incubation.) The tubes were transferred to room temperature, beads pelleted 2 min in the magnetic stand, and the supernatant collected and analyzed by agarose gel electrophoresis (See Figure 3.9; samples labeled 'DNA eluted from beads').

The results of this initial experiment (Figure 3.9) indicate that with no oligonucleotide added, about 2/3 of the SSC-DNA molecules remain in solution, whereas about 1/3 bind non-specifically to the Dynabeads. In the presence of a specific capture oligonucleotide (BTN-103A) more of the SSC-DNA (possibly 80%) is captured by the Dynabeads. Whereas in the presence of the non-hybridizing oligonucleotide (BTN-211B) this trend was reversed, with much more of the SSC-DNA remaining in solution. Overall these results suggest: 1) the Dynabeads have a high capacity to bind SSC-DNA non-specifically, and 2) presence of a specific oligonucleotide increases the quantity of SSC-DNA captured, whereas presence of an irrelevant oligonucleotide does not. The capacity of these ferro-magnetic streptavidin beads to non-specifically bind SSC-DNA is consistent with their porous, irregularly shaped surface (manufacturer's product information) and with the fact that single stranded DNA is notoriously 'sticky'. This motivated us to find a method to prevent non-specific binding to the Dynabeads.

3.7.1 PREVENTING NON-SPECIFIC BINDING OF SSC-DNA TO STREPTAVIDIN DYNABEADS

To try and minimize non-specific binding of SSC-DNA to the streptavidin beads, several approaches were tested, including varying the binding temperature, decreasing the $MgCl_2$ concentration, and pre-blocking the beads with either denatured salmon sperm DNA or with yeast tRNA. Of all the methods tested, pre-blocking the beads with yeast tRNA appeared to be the most effective. Details of this method are described in the following paragraph, and results of a representative experiment are shown in Figure 3.10.

To accomplish pre-blocking, immediately following washing and final resuspension in BW2 buffer (for details see above), an additional 1/5 volume of 10 µg/µl yeast tRNA stock (made in BW2 buffer) was simply added to the slurry of washed beads. For example, 2 µl of the yeast tRNA solution would be added to 10 µl of washed beads in BW2 buffer, and the resulting 12 µl of bead slurry would be used to perform capture in one hybridization tube. In order to maintain the concentration of NaCl at 1 M during capture, the hybridization tube containing the biotinylated oligonucleotide + SSC-DNA was also brought up to a final volume of 12 µl by adding 2 µl of TE buffer immediately prior to adding the 12 µl of washed/blocked beads.

For the experiment shown in Figure 3.9 B, four 500 ng aliquots of M13JFE7-DQ8B were added to four 0.2 ml thin-walled PCR microtubes. Two

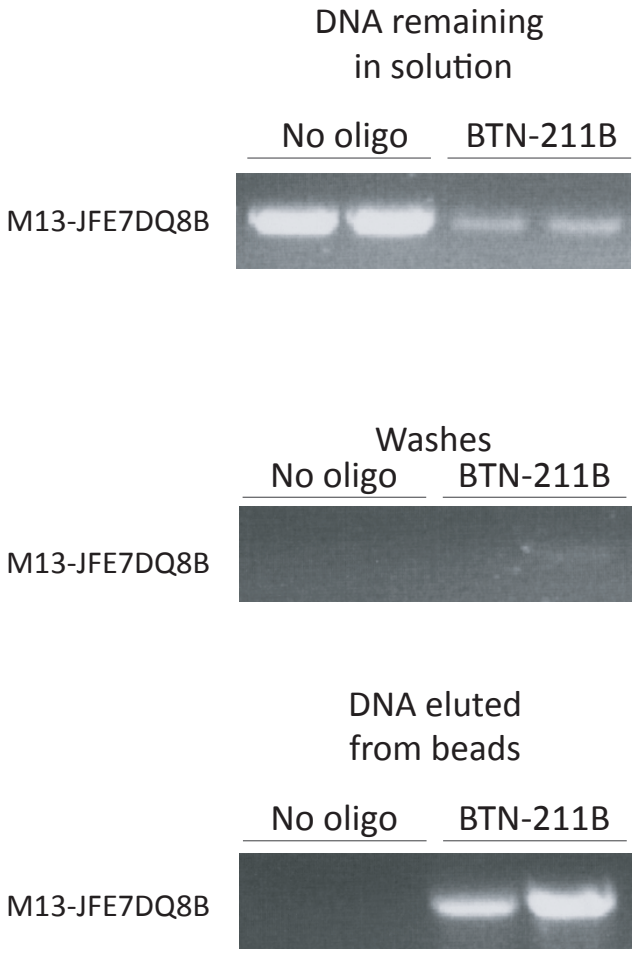


Figure 3.10 Results of model experiment to capture M13JFE7-DQ8B constructs out of solution using biotinylated oligonucleotide (BTN-211B) and streptavidin beads. The experiment was performed with pre-blocked streptavidin ferro-magnetic beads using yeast tRNA as the blocking agent. Non-specific binding of the SSC-DNA to the beads is completely prevented by this strategy. Experiment was done in duplicate, and fractions analyzed on a 1% agarose gel.

duplicate tubes had no oligonucleotide added (the missing volume being made up by TE), and two tubes had 250 ng of the specific complementary oligonucleotide BTN-211B added. The buffers, volumes, incubation times, and temperatures used were exactly the same as for the M13JFE7-DQ6A experiment shown in Figure 3.9, except that the streptavidin beads were pre-blocked with yeast tRNA and volumes for the capture reaction were adjusted slightly as described above.

Figure 3.10 demonstrates that pre-blocking with yeast tRNA completely eliminated any non-specific binding of the SSC-DNA to the streptavidin beads (at least to within the limits of detection of the agarose gel), whereas when a specific complementary biotinylated oligonucleotide was added, ≈70-90% of the SSC-DNA was captured by the beads, with only a minor fraction of the DNA remaining in solution. (It is possible that an even greater fraction of the SSC-DNA might be captured if the incubation/oligonucleotide annealing time at 50°C was increased from the 3 minutes used in this experiment.) Only a small amount of the SSC-DNA is seen in the pooled washes, indicating that the beads bind strongly to the captured SSC-DNA until the elution step.

3.7.2 TRANSFORMATION AND ANALYSIS OF RELATIVE NUMBERS OF BLUE AND WHITE PSEUDOPLAQUES TO DETECT ENRICHMENT OF SPECIFIC SSC-DNA'S

M13mp19 phage produces blue pseudoplaques in the presence of X-GAL and IPTG, whereas M13JFE7s (or DQ sub-clones) produce 'white' (really colorless) pseudoplaques. The ability to distinguish relative numbers of the two different M13 SSC-DNA molecules based on transformation, plating, and scoring of pseudoplaque color suggested another more sensitive method to test for specific enrichment of HLA-DQ sequences. Results for an experiment of this type are show in Table 3.6.

In this experiment increasing amounts of highly purified M13mp19 SSC-DNA (100 ng, 1 µg, 10 µg, and 100 µg as summarized in Table 3.6; note this purified DNA is shown in Figure 2.1 B) were mixed with 1 ng of purified M13JFE7-DQ8B SSC-DNA in 0.2 ml thin-walled PCR microtubes. The M13mp19 SSC-DNAs were all added in a volume of 7 µl, and the M13JFE7-DQ8B SSC-DNA in a volume of 1 µl. To concentrate 10 µg and 100 µg of the M13mp19 SSC-DNA into a volume of 7 µl, the appropriate volumes of stock solution were precipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, and the pellets resuspended in TE buffer—note that these two solutions ended up being very viscous. Once the four different mixtures of M13mp19 + M13JFE7-DQ8B were generated, 1 µl (250 ng) of the specific complementary oligonucleotide BTN-211B was added to each tube, as well as 1 µl of 100 mM $MgCl_2$, giving a final volume 10 µl. The tubes were placed in a thermocycler (Biometra T-Gradient) and heated to 95°C for 2 minutes, followed by 3 minutes at 50°C. The hybridization reaction tubes were removed from the thermocycler, allowed to

Table 3.6 Blue/White pseudoplaque counts after one cycle of selection

Amount of M13mp19 to M13JFE7-DQ8B SSC-DNA in starting mixture	100ng/1ng	1μg/1ng	10μg/1ng	100μg/1ng
Predicted ratio of blue to white in starting mixture*	10^2/1	10^3/1	10^4/1	10^5/1
Numbers of blue to white on 10^{-1} dilution plate after selection	41/21^	41/6^	8/9†	114/2†
Fold Enrichment#	50^	146^	1125†	1754†

* Estimated based on relative amounts of DNA; not measured by transformation/plating/counting
^ White pseudoplaques on these plates were shown by mapping to all contain the DQ8B cDNA insert, therefore fold enrichment in these columns is accurate
† White pseudoplaques on these plates were shown by mapping NOT to contain the DQ8B cDNA insert (i.e. they represent false positives); therefore fold enrichment is not meaningful for these columns
# Fold enrichment calculated based on starting ratio (predicted not measured) and numbers seen in 10^{-1} dilution plate after one cycle of selection

cool to room temperature for 5 min, and 2 µl of TE added to each tube, followed by 12 µl of washed/tRNA blocked Dynabead slurry. The mixtures were agitated on a roller wheel at 37°C overnight (≈16 hr), and beads captured and washed in a magnetic stand as for the previous experiments (all wash volumes 100 µl). Again, the beads were transferred to a fresh 0.2 ml thin-walled PCR microtube before the final pelleting, and the final drained pellet was resuspended in 20 µl of dH_2O.

The tubes containing the beads in dH_2O were placed in a thermocycler (Biometra T-Gradient) and incubated at 70°C for 2 minutes, then 50°C for 3 minutes, and finally transferred to room temperature where beads were pelleted 2 min in the magnetic stand, and the supernatants collected. The entire volume of each supernatant was used to transform 100 µl aliquots of chemically competent XL2-Blue *E.coli* cells, and the transformations were plated on LB plates containing X-GAL and IPTG, using top agar that was also supplemented with X-GAL and IPTG. Plates were incubated at 37˚C overnight to allow the bacterial lawn to become well established, and then placed at 4˚C for 8-24 hours to allow more complete color development of the pseudoplaques. Numbers of blue and white pseudoplaques were counted, and results are summarized in table 3.6.

Using the DQ8B-complementary oligonucleotide BTN-211B, relative enrichment of the M13JFE7-DQ8B SSC-DNA was clearly achieved for the 1:100

and 1:1000 mixtures selected through one cycle of hybridization and bead capture (i.e. since all of the white psuedoplaques on these plates contained the DQB8 insert). However, for the $1:10^4$ and $1:10^5$ mixtures, the white pseudoplaques that were seen were all likely due to mutations or deletions in the LacZα peptide gene of M13mp19. Appearance of these M13mp19 mutants from the mixtures with the higher amounts of input M13mp19 SSC-DNA (10 and 100 μg) may be somehow related to the amount of input DNA and/or to the high viscosity of these hybridization mixtures, but it is unclear why the white mutants would be enriched by the selection procedure. In any case these results suggest that future experiments of this type should probably limit the amount of input M13mp19 SSC-DNA to about 1 μg, and use decreasing amounts of input M13JFE7-DQ8B SSC-DNA to define the limits of the enrichment procedure. In this case selected SSC-DNA would need to be transformed into *E.coli* cells with competencies higher than we had available for our chemical competent cells.

The experiment summarized in table 3.6 represents a first attempt, and it would need to be repeated and extended in order to develop any firm conclusions. However, the results we do have suggest that enrichment of specific desired DNA sequences can indeed be achieved using biotinylated oligonucleotides and streptavidin beads, but that this enrichment is on the order of 100 to 200-fold. This is probably not sufficient to make the technique practical for selecting HLA-encoding cDNA elements from complex cDNA libraries, where enrichments on the order of at least 10,000-fold would be

required. This limitation, together with time constraints, led us to abandon this approach, and instead investigate RT-PCR based methods for achieving sequence based HLA typing.

3.8 References

1. Sahal, D. and Y. Fujita-Yamaguchi, *Protein kinase assay by paper-trichloroacetic acid method: high performance using phosphocellulose paper and washing an ensemble of samples on flat sheets.* Anal Biochem, 1987. **167**(1): p. 23-30.
2. Aldener-Cannava, A. and O. Olerup, HLA-DPB1 typing by polymerase chain reaction amplification with sequence-specific primers. Tissue Antigens, 2001. **57**(4): p. 287-99.
3. Holmberg, A., et al., *The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures.* Electrophoresis, 2005. **26**(3): p. 501-10.

CHAPTER 4

SEQUENCE BASED HLA-DPB1 TYPING: MODEL EXPERIMENTS USING RT-PCR AND CLONING, AND APPLICATION TO HAIR DYE ALLERGIC PATIENTS

4.1 INTRODUCTION

The experiments described in this chapter were motivated by the specific goal of developing a PCR-based method to achieve sequenced-based typing for HLA-DPB1. Due to the limited data available for HLA-DPB1 in the HLA database (17 complete transcripts out of 145 alleles), we were encouraged to develop a typing method that would acquire more complete HLA-DPB1 transcripts that could contribute to the quality of sequences available in the HLA database. As explained in chapter 1, there was some evidence in the literature to suggest that HLA-DP molecules might be involved in signaling to T cells involved in hair dye allergy, and therefore we chose to concentrate on typing this HLA gene in a collection of 16 hair dye allergic patients obtained from the University of Alberta Patch Test Clinic. However, before we proceeded to the patient samples we performed a number of model experiments using total RNA that we had purified from a human homozygous B lymphoblastoid typing cell line.

The model experiments made use of RNA purified from the APD cell lines, since the DPB1 haplotype of this cell line was known and therefore we knew the expected DNA sequence. These models experiments using RNA from the cell line involved a series of progressive steps: 1) establishment of the reverse

transcriptase (RT) reaction, 2) specific amplification of the DPB1 cDNA sequences using primers that extended from the ATG start code into the TAA stop codon (Table 4.1), and then re-PCR with a second set of primers which added *Sfi* I restriction sites to the ends, 3) digestion of the PCR fragments with *Sfi* I and cloning into *Sfi* I cut PCSD7, 4) picking individual resulting clones and sequencing in both directions using a local sequencing facility. For these model experiments, we tried several different DNA polymerases, including *Taq* DNA polymerase, *Pfu* DNA polymerase, a mixture of *Taq* and *Pfu* polymerases, and finally, a higher fidelity Phusion polymerase (Table 4.2). These model experiments demonstrated clearly that it was critical to use a high fidelity DNA polymerase. Also, they demonstrated that it was unnecessary to use the nested set of primers and that the specific primers with *Sfi* I sites on the ends could be used directly on the RT reaction template to generate the cloneable fragments. This meant that half as many PCR cycles were required, and therefore less mutations would be introduced into the PCR products. Once this preliminary work with RNA from cell lines had demonstrated a successful method, we applied it to the total RNA from the patient blood samples.

Once work with the patient blood samples began, it became clear that the *Sfi* I cloning strategy was too labour intensive and inefficient to generate the numbers of clones required. We were using a RT-PCR and then clone approach so that individual clones were sequenced to determine the anticipated two allelic sequences from each patient sample. This necessitated generating a fairly

Table 4.1 Oligonucleotides used to PCR amplify and then add *Sfi* I sites to HLA-DPB1 cDNA fragments

Name	Target Gene	Sequence	Length	Tm (50mM NaCl)	GC Content
5’ DPB1 primer	HLA-DPB1	5’-ATG.ATG.GTT.CTG.CAG.GTT.TC-3’	20bp	53.0°C	45%
3’ DPB1 primer	HLA-DPB1	5’-TTA.TGC.ACA.TCC.TCG.TTG.AAC-3’	21bp	52.6°C	42.8%
5'Sfi-DPB1	HLA-DPB1	5’-CTCTCTGGCCTGAGAGGCC.ATG.ATG.GTT.CTG.CAG.GTT.TC-3’	39bp	68.7°C	56.4%
3’Sfi-DPB1 primer	HLA-DPB1	5’-CTCTCTGGCCAAGATGGCC.TTA.TGC.AGA.TCC.TCG.TTG.AAC-3’	40bp	67.3°C	52.5%

The upper primer pair was designed to amplify the cDNA from start codon (ATG) to stop codon (TAA), generating a PCR product of 777 bp. The lower primer pair was designed to amplify the 777 bp product and add Sfi I restriction sites at each end, one with a -CTC overhang, and the other with a -TCT overhang. When this 815 bp fragment is digested with *Sfi* I it can be efficiently cloned into *Sfi* I cut pCSD7s in a specific orientation (See also Figure 3.5).

Table 4.2 Numbers of point mutations observed in cloned DPB1 RT-PCR fragments amplified using different thermostable DNA polymerases

Polymerase	# of clones sequenced	# of clones with no mutations	# of point mutations in non-perfect clones	Total number of point mutations
Taq	6	1	1, 2, 3, 4, 4	14
Taq + *Pfu*†	13	3	1, 1, 1, 1, 1, 1, 1, 1, 2, 2	13
Phusion	8	7	3	3

† *Taq* and *Pfu* were used in combination: 5 Units *Taq* with 1.25 Units of *Pfu*
Two sequential PCR reactions were done, each with a total volume of 100μl. A total of 50 PCR cycles were performed and the resulting PCR product was ligated into pCSD7s vector for sequencing. *Pfu* alone was used for RT-PCR reactions, but no product was produced.

large number clones for each patient – in our case we chose to do 24 clones from each of the 16 patients. Therefore, we abandoned the *Sfi* I directional cloning approach and moved to use a commercial TOPO cloning kit, which led to much higher throughput and production of the necessary plasmid clones. As well as the volumes of samples requiring DNA sequencing increased significantly in number, we established a relationship with a high throughput sequencing facility which could do the DNA sequencing much more economically, and on a much larger scale than the sequencing facility available at the University of Alberta.

These adaptations allowed us to determine the cDNA sequence and assign DPB1 alleles to 15 of our 16 patients samples. This data was then analyzed and the allele frequencies compared to those for a comparable general population. The conclusion of this work was that there did not appear to be any over-represented DPB1 alleles in the hair dye allergic population, suggesting that hair dye allergy is not associated with any particular DPB1 allele.

4.2 cDNA SYNTHESIS

Reverse transcriptase polymerase chain reactions were performed using enhanced avian myeloblastosis virus reverse transcriptase (eAMV-RT; Sigma). A total of 5.0 µg of total RNA was used as the template to generate cDNA, and an anchored poly-d(T)22 oligonucleotide (containing a 3' dV - either dG dA or dC) (Integrated DNA Technologies) was utilized as the primer for first strand cDNA synthesis. Superase-IN RNase inhibitor (Ambion) was used to protect the

RNA from degradation because it did not interfere with the production of cDNA. The following reagents were added to a 200 µl PCR tube: 5 µg of RNA template (total RNA), 1 µl deoxynucleotide mix (10mM dNTP), 1 µl of anchored poly-d(T)22 (3.5 µM) and DEPC treated dH_2O to a total volume of 10 µl. The sample was heated to 70°C to denature RNA secondary structures and allow an efficient reverse transcription reaction, and the sample was then immediately placed on ice. The contents were centrifuged at 200 RCF for 1 min at room temperature to collect all of the contents at the heel of the tube and the rest of the reagents were added: 6 µl of DEPC treated water, 1 µl of Superase-In RNAse inhibitor (20 units) (Ambion), 2 µl of 10X reaction buffer (provided with eAMV-RT) and 1 µl of eAMV-RT (20 units). The reaction was placed at 42°C for 50 minutes to allow complete first strand cDNA synthesis.

The sequence data for the HLA-DPB1 gene is available in the HLA nomenclature public database at the Anthony Nolan Research Institute [1, 2]. The database is regularly updated adding several new allelic sequences every few months. There are 145 different HLA-DPB1 allele variants in the database, with most of the alleles containing limited sequence data in the 5' and 3' ends and only 17 alleles with complete transcripts from the ATG start codon to the TAA stop codon (777bp). The complete sequences were used to generate a set of PCR primers to clone the HLA-DPB1 gene's complete transcript. A set of primers was also made with restriction enzyme recognition sequences for *Sfi* I (GGCCNNNN^NGGCC) flanking the PCR fragment produced. These primers were

constructed with varied *Sfi* I recognition motifs to generate two distinct sticky ends for directional cloning of the PCR product into the pCSD7 vector (Table 4.1)(Figure 4.1).

4.3 PCR REACTIONS

Two sequential PCR reactions were used to produce a PCR product with flanking *Sfi* I restriction sites. The first PCR reaction amplified the target sequence from the total cDNA population; the second PCR reaction amplified the desired HLA-DPB1 allele with the flanking *Sfi* I restriction enzyme sites for directional cloning.

Phusion high-fidelity DNA polymerase (New England Biolabs) was utilized to minimize possible point mutations during PCR amplification. Phusion (high-fidelity) has an error rate of 4.4×10^{-7} (New England Biolabs) producing the least number of possible errors per PCR product, making it the ideal polymerase to use. All PCR reactions were performed in a Biometra T-Gradient thermocycler with a heated lid. The PCR mixture included 1 µl of Phusion DNA polymerase (New England Biolabs), 20 µl of the Phusion HF buffer (New England Biolabs), 200 µM of all four dNTPs (dATP, dCTP, dGTP and dTTP), 1 µM of primer mix (forward [fwd] and reverse [rev] primers), and 10 µl of cDNA template in a total volume of 100 µl. The PCR reaction was set at 98°C for 30 seconds to enable denaturation of secondary structures in the template. The cycle settings were: 1.

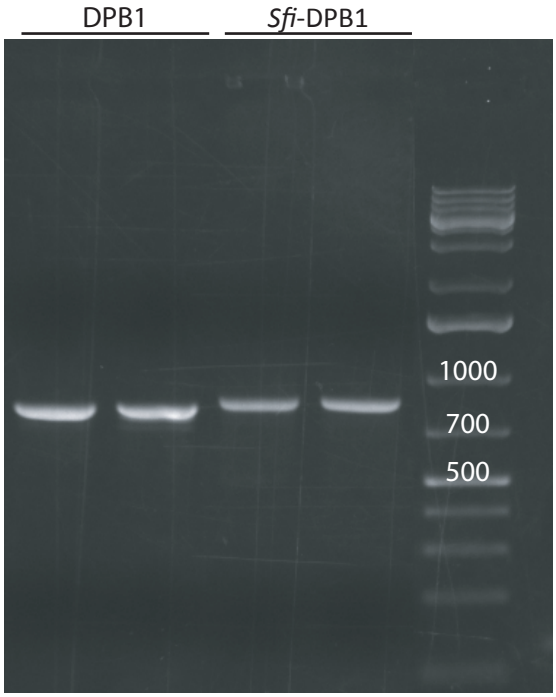


Figure 4.1 - cDNA fragments generated by PCR using the DPB1 and Sfi-DPB1 primer pairs. Template for the first PCR reactions (leftmost 2 lanes; reactions were done in duplicate) was cDNA reverse transcribed from total cellular RNA made from the APD cell line. The DPB1 oligonucleotide pair was used to prime synthesis, and 777 bp fragments were generated. The first PCR reactions were purified using QIAquick purification kits (Qiagen), and 1/200 of each was used as template for a second set of PCR reactions (next 2 lanes; reactions again done in duplicate). The second PCR reactions were primed using the Sfi-DPB1 oligonucleotide pair, and they produced 815 bp products in both cases. The PCR fragments shown in this 0.8% agarose gel were generated using Phusion DNA polymerase, but the bands are identical to those obtained when Taq DNA polymerase or a mixture of Taq and Pfu DNA polymerase were used to catalyze both reactions. Note that we were unable to obtain any bands when Pfu DNA polymerase alone was used with first strand cDNA as template. All PCR reactions were done in 100 µl total volumes (no matter which enzyme was used), using conditions recommended by the enzyme manufacturer, and 10 µl of the reaction was run on an agarose gel. The sizes of the PCR products were confirmed by direct comparison to 5ul of Fermentas 1KB Plus ladder (right lane).

98°C for 10 seconds, 2. 62°C for 30 seconds, 3. 72°C for 40 seconds. PCR reactions were cycled twenty-five times and the product stored at 4°C. The PCR product was purified using a QIAquick PCR purification kit (Qiagen), and eluted with 50 µl of TE buffer (Figure 4.2).

4.3.1 CLONING INTO PCSD7s VECTOR

The PCR products and the pCSD7-s vector were digested with FastDigest *Sfi* I (Fermentas) at 50°C for 30 minutes to ensure a complete digest. The digested samples were run on a 1.0% low melting point (LMP) agarose gel made with 60ml of 1x TBE buffer (89mM Tris Base, 89mM Boric acid, 2mM EDTA) and 1ul of 10mg/ml Ethidium bromide. The gel was run at 70 volts for 90 minutes, and the DNA fragments were visualized using a 365nm ultraviolet transilluminator before being excised from the agarose gel with a sterile razor blade. Gel slices were purified using a QIAquick gel extraction kit (Qiagen) and the DNA was eluted with 30 µl of dH_2O.

Three ligation reactions were completed (two controls and a DPB1 PCR product): 1. negative control. 2. positive control. 3. DPB1 insert.

	-ve control	+ve control	DPB1 insert
dH_2O	30.5µl	20.5µl	20.5µl
T4L 10x Buffer	4µl	4µl	4µl
Vector DNA	5µl	5µl	5µl
Stuffer DNA	-	10µl	-
DPB1 DNA	-	-	10µl
T4 ligase	0.5µl	0.5µl	0.5µl
Total volume	40µl	40µl	40µl

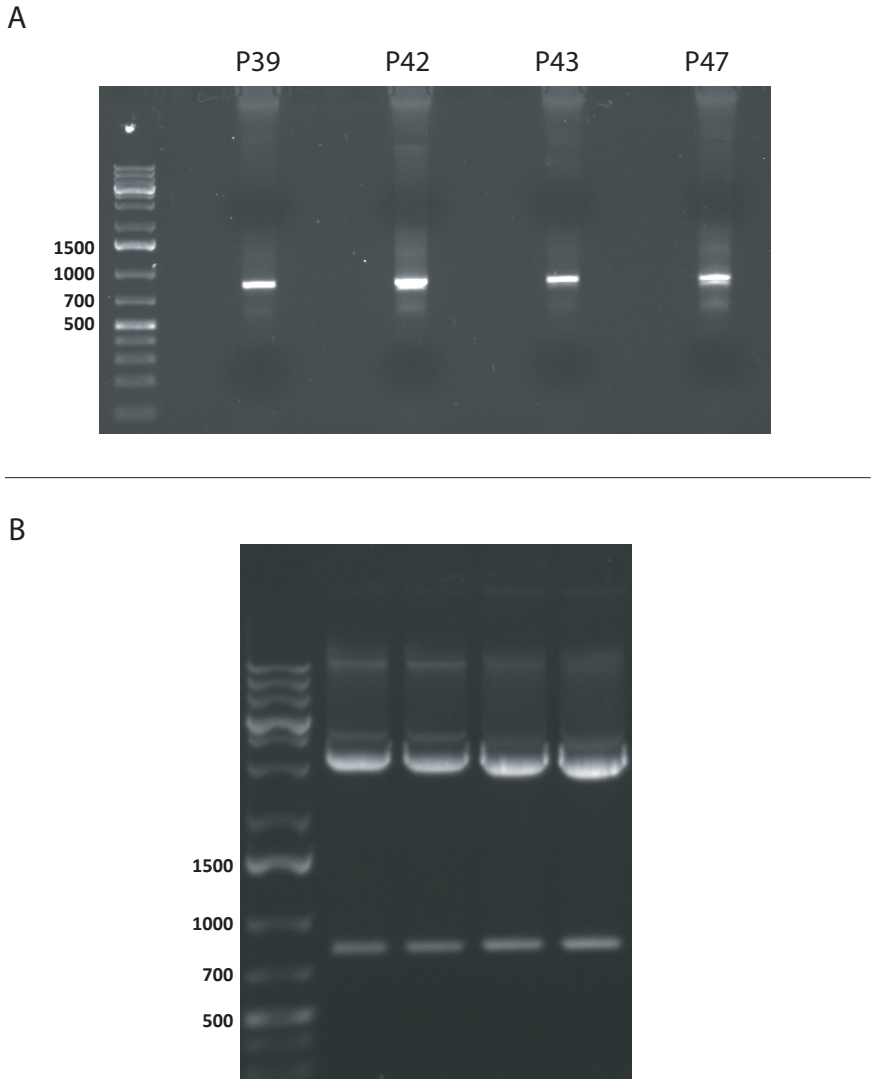


Figure 4.2 - DPB1 cDNA fragments generated by RT-PCR using reverse transcribed patient RNA, and representative cloned DPB1 cDNA fragments excised from pCSD7 by Sfi I digestion. A. RT-PCR reactions for four different patients are shown (patients 39, 42, 43 and 47; numbers were randomly assigned). Patient blood was collected in EDTA vacutainers, total RNA extracted from buffy coat cells using a modified TRIzol extraction procedure, and cDNA generated using enhanced avian myeloblastosis virus reverse transcriptase (eAMV-RT). The cDNA generated was used as template for the PCR reactions, which were primed using the Sfi-DPB1 oligonucleotide pair, and catalyzed using Phusion DNA polymerase. Ten microlitres of the 100 µl PCR reaction was run on the 1.0% agarose gel shown here, and PCR products were of the expected size. The remaining 90 µl of each PCR reaction was purified using a QIAquick purification kit (Qiagen), and half of the resulting material digested with Sfi I and cloned into Sfi I cut pCSD7s using the gel slice ligation method. B. Four random colonies were picked from the transformations resulting from the gel slice ligation reactions (one colony from each patient), and plasmid minipreps prepared, digested with Sfi I, and run on a 1.0% agarose gel. The bands show that all four constructs have inserts of the appropritate size (790bp).

Ligations were incubated at 16°C for 16h in a T-gradient thermocycler (Biometra), and the ligase was inactivated by heating the reaction to 65°C for 10 minutes. Electrocompetent XL2-Blue *E.coli* cells were thawed on ice and transformed with 5 µl from each ligation reaction as per section 2.2. The cells were resuspended in 1.0ml of 2xYT-MT media and placed in a roller wheel at 37°C for 1h. Cells were plated by streaking 100 µl from the liquid culture on 2xYT-MT agar plates containing 100 µg/ml ampicillin (Amp+ 2xYT-MT plates). The plates were allowed to dry and were placed at 37°C overnight to allow colonies to grow.

Colonies were picked, placed into 2ml of 2xYT-MT media in a 14ml falcon (PP) tube and allowed to grow in a roller wheel at 37°C for 8h. The cells were pelleted by centrifugation at 650 RCF for 5 minutes (2000rpm in an Allegra 6KR kneewell centrifuge with a GH3.8 rotor), and the plasmid DNA was extracted using a Wizard miniprep kit (Promega). Plasmid DNA was analyzed by a NanoDrop 1000 spectrophotometer (Thermo Scientific) giving an average yield of 8 µg (268.4ng/µl).

4.4 SAMPLE PREPARATION FOR SEQUENCING

The sequencing of samples was done in two separate facilities. The Applied Genomics Center (AGC) in the department of Medical Genetics at the University of Alberta was used for the trial samples from volunteers in the lab

and produced the sequence results of a small number of samples. The Genome Sciences Center (GSC) at the BC Cancer Agency was used for the sequencing of the patient samples.

Samples were prepared to a concentration of 200ng/10 µl for sequencing at the AGC. Sequencing of pCSD7 constructs was performed using the T7 universal primers for amplification of the DNA.

Samples were sent to the GSC as frozen cultures. The colonies picked from the construct generation were placed in 200 µl of freezing 2xYT-MT media (2xYT-MT media plus 15% glycerol) with 100 µg/ml Ampicillin, grown overnight at 37°C then placed at -80°C for storage until shipment. For each of the 16 patients studied, 24 different construct samples were sent for analysis producing a total of 384 samples sent for sequencing analysis (Table 4.3).

4.5 TOPO TA CLONING METHOD

The pCR 2.1 TOPO Cloning Kit (Invitrogen) allows for Taq polymerized products to be easily ligated into the pCR 2.1 TOPO vector (Invitrogen) by use of the 3' adenine overhang [11]. When using polymerases that have a proof reading domain (for example Pfu DNA polymerase [New England Biolabs] or Phusion DNA polymerase [New England Biolabs]), the PCR products generated are blunt, lacking a 3' adenine overhang [1]. In order to clone the PCR products using Phusion polymerase and the TOPO 2.1 vector the following procedure was used.

Table 4.3 Clones with inserts and assigned alleles in patient study.

Patient #	# of Colonies picked	Sequencing Plate #	# of clones with DPβ1 inserts*	# of assigned DPβ1 sequences**
37	79	1	22/24	14
38	96	1	19/24	17
39	96	1	14/24	13
40	43	1	19/24	19
41	96	2	15/24	15
42	64	2	21/24	20
43	96	2	20/24	20
44	56	2	23/24	23
45	96	3	21/24	18
46	**96**	**3**	**15/24**	**3†**
47	62	3	21/24	16
48	83	3	21/24	20
49	96	4	19/24	17
50	36	4	21/24	18
51	96	4	20/24	18
55	96	4	16/24	16

* Based on sequence data

** Where DPβ1 haplotypes can be clearly assigned

† 15 clones with inserts showed many mixed haplotype crossovers (see Appendix A.1), so that only three could be clearly assigned to a known haplotype

The number of colony picks per patient was dependent on the efficiency of construct generation and transformation of the constructs. Of the 24 colonies sent for analysis, there were only some constructs with HLA-DPB1 gene inserts. Of the constructs with gene inserts only some contained a sequence clear enough for HLA-DPB1 allele assignment.

To clone the PCR products of Phusion polymerase into the TOPO 2.1 vector, an adenine was added to the 3' end of the PCR products with Taq polymerase (produced by K. Suzuki in the Elliott Lab). The PCR products were cleaned by a QIAquick PCR purification kit (Qiagen) before being treated with 2 units of Taq polymerase, 10mM of dATP, 50mM $MgCl_2$, and 5 µl of 10X Taq Buffer (Invitrogen) in a total volume of 50 µl. Reactions were incubated at 72°C for 10 minutes, and immediately placed on ice. In order to clone only the specific 777bp PCR product desired, the reaction was run on a 1.0% LMP agarose gel at 70V for 90 minutes. The gel was analyzed in a 365nm ultraviolet transilluminator and the band specific for 777bp was excised by razor blade and purified with a QIAquick gel purification kit (Qiagen). The PCR product was eluted from the wash column in 50 µl of dH2O, which produced the product ready to be cloned into the pCR 2.1 TOPO vector.

4.6 ASSIGNING HLA ALLELE CLASSIFICATION

The sequencing files received from the GSC contained detailed chromatographs for each patient sample. The sequence chromatographs were thoroughly scanned for any clearly identifiable anomalies, including poor sequence read-out due to multiple constructs present per sample, no insert present in the sample, or general sequencing problems.

Once the sequences were confirmed to be genuine and of satisfactory quality, they were aligned using the ClustalW2 sequence alignment program [3]. The consensus sequence files obtained were further analyzed in one of two ways: 1. homozygous allele assignment or 2. diploid allele assignment.

4.6.1 ALLELE ASSIGNMENT TO HOMOZYGOUS B-CELL LINES

When typing the homozygous B-cells (APD cells), the alignment files produced with ClustalW2 were used to interrogate the sequences for point mutations that may have arisen during the PCR cloning steps. The alignment files were used to generate a consensus sequence of the sample in question. The consensus sequence was then aligned with all known HLA-DPB1 sequences collected from the HLA sequence database (version 3.2.0) found at http://hla.alleles.org [1, 2]. This alignment was done using the specialized basic local alignment search tool for nucleotide matches (BLASTn) called bl2seq accessible through the National Center for Biotechnology Information (NCBI) website [4]. The alignment was performed using the default settings for matching sequences.

The sequence matches with the highest maximal percent ID identified the sequence(s) with the highest scoring pairs. The total alignment score gave a higher score to sequences with longer pairing sequences (per basepair); given that the allele database contains sequences of different lengths, it was crucial to

look at the exact matches rather than the highest percent coverage or total score.

4.6.2 ALLELE ASSIGNMENT TO DIPLOID SAMPLES

For diploid samples, the approach to identifying the alleles present was more thorough in order to distinguish the two possible alleles in the sample. The alignment files and guide tree produced by ClustalW2 alignment were used to identify two distinct sequences based on their alignment scores. Sequences that had 100% alignment scores were grouped together as a specific allele group (i.e. Allele A – Sequence1, 2, 4 and 6 / Allele B – Sequence 3 and 5). A consensus sequence was generated from each allele group to remove any point mutations generated during the cloning steps. The consensus sequence from each allele group was then aligned with all known HLA-DPB1 sequences collected from the HLA sequence database (version 3.2.0) found at http://hla.alleles.org [1, 2]. This alignment was also completed through bl2seq from NCBI [4]. Any outlier sequences that did not classify into a specific allele group were further analyzed by directly comparing the forward and reverse sequences to identify any sequence anomalies. The chromatographs of the outlier sequences were also closely examined to determine if the single nucleotide polymorphisms (SNPs) identified in these sequences were indeed new alleles.

4.7 ALLERGIC CONTACT DERMATITIS: PATIENT STUDY

Allergic contact dermatitis triggered by para-phenylenediamine (PPD) [5, 6] has been loosely associated with the HLA protein HLA-DPB1 [7], yet a direct association between an HLA-DPB1 allele/allele group and susceptibility to ACD to PPD has yet to be identified.

A pilot study was established to identify any evident HLA-DPB1 allelic distributions present in a small patient group with ACD to PPD (approved by the Human Ethics Research Online [HERO] board at the University of Alberta). This study was set up through the Edmonton patch test clinic with the help of Dr. John Elliott. Patients with moderate to severe reaction to PPD on their patch test were asked to join the study by donating blood samples for HLA analysis. When patient samples were received the blood was processed immediately to assure the highest quality of nucleic acid products (RNA/DNA).

4.7.1 PATIENT BLOOD SAMPLES

Human blood samples were collected and total RNA was extracted by the method outlined in section 2.8. Aliquots of the cells were collected and stored at -80°C for future genomic studies. Constructs were generated using the TOPO 2.1 TA cloning kit outlined in section 3.1.4 (Table 4.4).

Table 4.4 HLA transcripts encoded within pCR 2.1 TOPO vector assembled in this study

Plasmids	Description
PCR 2.1 TOPO-DP*01:01:01	pUC ori, Amp^r, Kan^r, DPβ1*01:01:01
PCR 2.1 TOPO-DP*02:01:02	pUC ori, Amp^r, Kan^r, DQβ1*02:01:02
PCR 2.1 TOPO-DP*03:01:01	pUC ori, Ampr, Kanr, DQβ1*03:01:01
PCR 2.1 TOPO-DP*04:01:01	pUC ori, Ampr, Kanr, DQβ1*04:01:01
PCR 2.1 TOPO-DP*04:02	pUC ori, Ampr, Kanr, DQβ1*04:02
PCR 2.1 TOPO-DP*10:01	pUC ori, Ampr, Kanr, DQβ1*10:01
PCR 2.1 TOPO-DP*17:01	pUC ori, Ampr, Kanr, DQβ1*17:01
PCR 2.1 TOPO-DP*20:01:01	pUC ori, Ampr, Kanr, DQβ1*20:01:01
PCR 2.1 TOPO-DP*104:01	pUC ori, Ampr, Kanr, DQβ1*104:01
PCR 2.1 TOPO-DP*109:01	pUC ori, Ampr, Kanr, DQα1*109:01
PCR 2.1 TOPO-DP*116:01	pUC ori, Ampr, Kanr, DQα1*116:01
PCR 2.1 TOPO-DP*123:01	pUC ori, Ampr, Kanr, DQα1*123:01

The patient constructs were confirmed by sequencing of several generated constructs. The sequences were analyzed by aligning them with known sequences from the HLA database; the alleles were assigned to the constructs whose sequences had the highest identity scores by BLASTn alignment to the HLA database.

4.7.2 SEQUENCING PATIENT SAMPLES

The patient constructs were produced using a TOPO 2.1 TA cloning kit, which included a pCR 2.1 TOPO vector with an ampicillin resistance gene. The ampicillin resistance is functional only if the vector is recircularized and transformed into XL2-Blue *E.coli* hosts. Recircularization of the vector was accomplished by inserting the patient HLA-DPB1 PCR product as the insert. The HLA-DPB1 gene was cloned from patient cDNA samples and was prepared for TA cloning through treatment with Taq polymerase. The completed vector+insert constructs were transformed into XL2-Blue *E.coli* bacteria and grown in amp+ 2xYT-MT plates.

The colonies were counted and a maximum of 96 colonies were picked from the transformation plates using sterile toothpicks. Each pick was placed into 200 µl of freezing media (2xYT-MT media plus 15% glycerol) in a single well of a 96-well flat bottom plate with a total of 96 colony picks selected for each patient whenever possible (some transformation efficiencies yielded <96 colonies to pick from) (Table 4.3). The plates were incubated at 37°C for 24h to ensure culture growth. A sample set for sequencing was made by duplicating 24 cultures from each patient plate into a new 96-well flat bottom plate containing 200 µl of freezing media. The duplicate plates contained 4 patients' sample set, totaling 96 samples per plate (24 samples x 4 patients). A total of 384 cell cultures (4 x 96 wells) were prepared, and the duplicate plates were grown

overnight at 37°C then frozen to -80°C before shipping. The samples were stored in dry ice and shipped to the GSC for high throughput sequencing.

4.7.3 PATIENT SEQUENCE ANALYSIS

The sequencing results from the Genome Sciences Center showed that some of the samples sent for sequencing were unable to obtain a high quality sequence and some samples were unable to obtain a sequencing product at all. Of the 384 samples sent for analysis, 307 could be sequenced. Therefore, an average of 19 samples out of a maximum of 24 could be sequenced per patient; this ranged from 14 to 23 sequences per patient, an average of 79.9% (Table 4.3). Each of the sequences was individually analyzed to verify a strong and clean sequence read in the chromatographs. Sequences that showed an unusual chromatograph read out (for example, double peaks, or a very weak signal) were not used for allelic alignments. The patient sequences were analyzed as stated in section 4.6.2. From the 16 patients analyzed, a total of twelve unique alleles were observed (Table 4.5).

Ideally several different sequences would have been obtained to assign each specific allele, but some allele sequences were limited (eg. In patient 38, DPB1*02:01:02 had 16 clear sequences for allele assignment, but DPB1*109:01 had only one clear sequence for allele assignment). The disproportional distribution of these alleles could be accounted to levels of allele-specific RNA transcripts present when the blood samples were collected (preferential

Table 4.5 Allele assignments for the patient samples with ACD to PPD.

Patient #	Allele 1	# of clear sequences	Allele 2	# of clear sequences
37	*02:01:02	12	*123:01	2
38	*02:01:02	16	*109:01	1
39	*04:01:01	8	*04:02	5
40	*04:01:01	15	*02:01:02	4
41	*04:02	8	*04:01:01	5
42	*17:01	13	*04:01:01	8
43	*04:01:01	20		
44	*04:01:01	23		
45	*116:01	10	*04:02	6
46	**57:01*	*2*	**04:02*	*1*
47	*04:01:01	16		
48	*03:01:01	14	*04:01:01	6
49	*104:01	6	*04:02	7
50	*01:01:01	10	*20:01:01	8
51	*04:01:01	18		
55	*02:01:02	12	*10:01	4

The allele assignments were made as described in section 4.6.2. There are twelve unique alleles identified in the patient study. Of the 16 patients in the study, four patients showed a homozygous genotype to HLA-DPB1*04:01:01. Patient 46 had the least number of constructs that were readable; the sequence files produced were questionable. The RNA profile for patient 46 show poor quality mRNA and PCR products, so the sequences for this patient were not used in any further analyses.

expression of one allele over another). The allelic distributions of the patient group were analyzed and displayed favored representation of HLA-DPB1*04:01:01 (present in 60% of patients and had an allelic frequency of 43.3%) (Table 4.6).

The allelic distribution of a large control population specific for HLA-DPB1 was required to use as a control population in order to identify any allelic distribution patterns present in the patient group. The control population used came from a study identifying population distributions of the HLA-DPB1 gene in a group of 200 Swedish Caucasian volunteers [8]. The allelic distribution of HLA-DPB1 is not evenly distributed, but instead follows ethnic-specific distribution patterns [9]. A Caucasian population expresses the allele HLA-DPB1*04:01 in roughly 40% of the patients analyzed, whereas a West African population predominantly presents HLA-DPB1*17:01 and *01:01 with allelic frequencies of 28% and 20% respectively [8, 10]. Ethnicity was not recorded from our patients, but the impression of the clinician was that this population was almost entirely of Caucasian descent.

To compare the control population directly with the patient results obtained in this study, discrepancies between the typing methods used in the two studies were identified and any possible mismatches of the typing results were detected. Since the control group sequences were subjected to the PCR-

Table 4.6 Allelic distributions of patient study in this study and patients with alleles.

HLA-DPB1 allele	# of patients with this allele present at least once	# Patients homozygous for this allele	# of times this allele occurred	Allelic frequency – patients in this study
*04:01:01	9/15 (60%)	4	13	43.3% (13/30)
*04:02	4/15 (26.6%)	0	4	13.3% (4/30)
*02:01:02	4/15 (26.6%)	0	4	13.3% (4/30)
*01:01:01	1/15 (6.6%)	0	1	3.3% (1/30)
*03:01:01	1/15 (6.6%)	0	1	3.3% (1/30)
*10:01	1/15 (6.6%)	0	1	3.3% (1/30)
*17:01	1/15 (6.6%)	0	1	3.3% (1/30)
*20:01:01	1/15 (6.6%)	0	1	3.3% (1/30)
*104:01	1/15 (6.6%)	0	1	3.3% (1/30)
*109:01	1/15 (6.6%)	0	1	3.3% (1/30)
*116:01	1/15 (6.6%)	0	1	3.3% (1/30)
*123:01	1/15 (6.6%)	0	1	3.3% (1/30)
Totals:			32	100%

The allele assignments for the patients were analyzed to interpret the allele distributions present in this population.

SSP method, and the patient sequences in the present study were subjected to the described methods (section 2 and above), discrepancies between the different methods had to be accounted for. To do this, the patient sequences obtained in this study were hypothetically subjected to the same methods as the control group sequences (PCR-SSP method; described in section 1.3.2). The control population data was collected in 2001, when only 84 alleles for HLA-DPB1 had been identified. As of January 2011, there were 145 recognized alleles for HLA-DPB1 [1, 2]. Of the twelve unique alleles identified in the present study, eight alleles would have found a direct match to their assigned allele using the PCR-SSP method. In order to identify how the four outstanding alleles (HLA-DPB1*104:01, *109:01, *116:01 and *123:01) would have been assigned with the PCR-SSP method, the SNPs that make these alleles unique were identified using their FASTA files from the Anthony Nolan Research Institute. These sequences were then individually aligned using ClustaW2 with alleles that were known in 2001 (HLA-DPB1*01:01 to *84:01). The aligned sequences were given an identity score when matched with the query sequence; any sequence that had a nucleotide mismatch with the query sequence had a reduced identity score. The alleles that had the top 5 identity scores were further analyzed by highlighting the SNPs that differed between the two sequences. If the SNPs locations in the aligned sequence matched the SNPs locations originally identified as unique in the query sequence and had the highest identity score, then the query sequence was assigned that allele's genotype (Table 4.7).

Table 4.7 Allelic distribution of control population and patients in this study.

HLA-DPB1 allele	Published allelic frequency – Caucasian population[†] (n=200)	Allelic frequency – patients in this study	p-values[‡]
*04:01:01	42.5%	43.3% (13/30)	1.0
*04:02	14.5%	16.6% (5/30)	0.783
*02:01:02	13.8%	16.6% (5/30)	0.779
*01:01:01	4.5%	3.3% (1/30)	1.0
*03:01:01	12.0%	6.6% (2/30)	0.544
*10:01	0.8%	3.3% (1/30)	0.706
*17:01	0.8%	3.3% (1/30)	0.706
*20:01:01	0.5%	3.3% (1/30)	1.0
*75:01	NR	3.3% (1/30)	-
Totals:		100%	

[†] *population data from Aldener-Cannava & Olerup (2001)*

[‡] P-values were calculated by Fisher's exact test, using SPSS 12.0

The allelic distributions from the patient study are compared directly with the control population. The p-values indicate that there is no significant statistical difference between the two populations.

4.8 References

1. Robinson, J., et al., IMGT/HLA database--a sequence database for the human major histocompatibility complex. Tissue Antigens, 2000. **55**(3): p. 280-7.
2. Robinson, J., et al., The IMGT/HLA database. Nucleic Acids Res, 2009. **37**(Database issue): p. D1013-7.
3. Larkin, M.A., et al., Clustal W and Clustal X version 2.0. Bioinformatics, 2007. **23**(21): p. 2947-8.
4. Altschul, S.F., et al., Basic local alignment search tool. J Mol Biol, 1990. **215**(3): p. 403-10.
5. Valks, R., et al., Contact dermatitis in hairdressers, 10 years later: patch-test results in 300 hairdressers (1994 to 2003) and comparison with previous study. Dermatitis, 2005. **16**(1): p. 28-31.
6. Zapolanski, T. and S.E. Jacob, para-Phenylenediamine. Dermatitis, 2008. **19**(3): p. E20-1.
7. Sieben, S., et al., Delayed-type hypersensitivity reaction to paraphenylenediamine is mediated by 2 different pathways of antigen recognition by specific alphabeta human T-cell clones. J Allergy Clin Immunol, 2002. **109**(6): p. 1005-11.
8. Aldener-Cannava, A. and O. Olerup, HLA-DPB1 typing by polymerase chain reaction amplification with sequence-specific primers. Tissue Antigens, 2001. **57**(4): p. 287-99.
9. Shammaa, D., et al., First report on HLA-DPB1 gene allelic distribution in the general Lebanese population. Genet Test Mol Biomarkers. **14**(2): p. 259-61.
10. Luger, A.M., et al., Molecular and serologic analysis of HLA genes and immunoglobulin allotypes in IgA nephropathy. Autoimmunity, 1994. **19**(1): p. 1-5.
11. Brownstein, M.J., J.D. Carpten, and J.R. Smith, *Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping.* Biotechniques, 1996. **20**(6): p. 1004-6, 1008-10.

CHAPTER 5

GENERAL CONCLUSIONS AND DISCUSSION

5.1 SUMMARY OF RESULTS

Developing an HLA typing procedure that is cost effective and novel presented with several difficult obstacles. Although the original plan for the procedure did not fully develop into a functional method, several of the steps developed were fruitful. These steps lead to the development of an efficient method of sequence based typing of HLA-DPB1.

5.2 SOLUTION HYBRIDIZATION

The solution hybridization procedure was successful at capturing large constructs (7kB) from single stranded M13 vectors. This procedure captured and purified specific constructs quickly and easily, without the need to screen complex cDNA populations. Invitrogen has perfected a solution hybridization method and provides a kit to capture constructs 2kb or larger (Dynabeads kilobaseBINDER). The development of this kit by Invitrogen validated the method described in this study and demonstrated that the method was useful and in demand. The main drawback in the described developed capturing procedure was the limited quantity of captured constructs. In order to ensure that the captured constructs were cloned by the mechanisms in the M13 vector, they required transformation into highly competent *E.coli* cells. A competency level of

1×10^{10} CFU/μg or greater for the cells was required in order for the captured molecules to be efficiently transformed. Due to limitations in the development of competent cells, only a competency level of 1×10^{8} CFU/μg for electrocompetent cells was accomplished.

With several limitations to the solution hybridization procedure identified, the decision was made to change the approach to the problem and use PCR amplification for cloning the genes of interest. The change to PCR amplification made the generation of HLA-DPB1 PCR products effortless, limited only by the quality of the RNA and subsequent cDNA used as the template.

5.3 RNA EXTRACTIONS

The RNA extraction methods outlined in sections 2.5, 2.6 and 2.7 were of great value to extract high quality RNA from various samples. The total RNA extraction method from human blood (section 2.7) was pivotal in the development of high quality cDNA and the acquisition of HLA-DPB1 sequences from the patient study conducted (Chapter 4) (Figure 4.2).

The procedure developed for RNA extraction enabled the development of the HLA typing procedure used in the patient study (section 4.7). RNA extraction is a well-established procedure available in kits provided by many suppliers, however, in-house development of the RNA extraction procedure allowed for a cost effective approach (section 2.5). The guanidine-thiocyanate phenol chloroform solution, commercially available as TRIzol (Invitrogen), is the best

way to acquire large quantities of RNA from samples. Production of high quality total RNA with TRIzol alone can be challenging because the requirement to precipitate the RNA to remove all traces of organic content is not ideal. Precipitating RNA produces a sticky and often tough pellet that is difficult to resuspend; this in turn may result in a loss of total RNA product. Also, possible contamination of TRIzol in the extracted RNA significantly reduces the quality of the final RNA product. Extracting large quantities of RNA with TRIzol enabled us to assess different RNA cleaning procedures to obtain the most effective cleaning method. The column based procedure used in the Qiagen RNeasy kits provided the most effective method to clean RNA samples without the need to precipitate the RNA. Combining these procedures together allowed us to extract a large amount of total RNA (200 µg/sample) of very high quality (between 1.9-2.0 for 28S/18S ratios) (Figure 2.2).

The extraction of high quality total RNA from human blood was emphasized by minimizing contaminants, such as haemoglobin, from blood. To minimize possible contamination from erythrocytes in the leukocytes, the blood samples were centrifuged to extract the buffy coat layer. The buffy coat was treated with erythrocyte lysis buffer to ensure all remaining erythrocytes were removed from the leukocytes. This allowed for RNA extraction to be completed without the presence of contaminating erythrocytes (Figure 2.4).

The leukocytes extracted from the blood samples were used to extract total RNA for analysis (Figure 2.4). The RNA extracted produced high quality cDNA with eAMV-RT. Using cDNA as the template, the PCR cloning of the HLA-DPB1 gene provided complete transcripts. The extraction of the leukocytes from blood also allowed for the storage of the cells at -80°C for future studies. The cell pellets were processed to extract the genomic DNA for genomic analysis of HLA-A, HLA-B and HLA-C (unpublished work, D. Zimmerman).

The ability to HLA type specific alleles for DPB1 is directly related to the quality of the RNA produced. For example, patient 46's sample was mishandled due to a pH error in the erythrocyte lysis buffer. This lead to an inefficient lysis of erythrocytes, increasing the handling time of the leukocytes and finally produced a poor quality cell pellet, in turn producing bad quality RNA profiles (Figure 2.5).

The total RNA extracted generated a cDNA product and a PCR product for HLA-DPB1. The PCR product for patient 46 was cloned and sequenced with all other samples. The consequence of the poor quality RNA was evident in the sequencing results for patient 46 (Table 4.3 and Table 4.5), as there were only three readable sequences, and, of those sequences, only two assignable sequences were made with low confidence (due to poor chromatographs). Given the poor sequencing results, this patient sample was omitted from the overall study.

Using a high fidelity DNA polymerase minimized the possibility of point mutations being generated during PCR amplification. It is important to note that point mutations could also be generated by the reverse transcriptase used (AMV-RT has an error rate of 2.7×10^{-5} errors/base [8]). These point mutations could results in improper assignment of specific alleles or wrongfully claiming the discovery of a new allele.

5.4 HLA AND DISEASE

The association of the human leukocyte antigens and disease has been very thoroughly studied. One of the biggest limitations in the field is the availability of a cost-effective procedure for HLA typing. There are several kits, including SBTexcellerator kits by Qiagen, and ALLSET GOLD HLA typing kits by Invitrogen, that provide low to high quality typing for certain HLA genes (A, B, C, DQB1, DRB1, and DPB1) but are generally restricted to exon 2 and 3 for class I genes and exon 2 for class II genes. The ability to clone HLA transcripts and sequence them entirely gives an enhanced depiction of HLA polymorphisms, allelic distributions and specific aspects of HLA associations with disease.

5.5 HLA-DPB1 AND ACUTE CONTACT DERMATITIS

The associations between HLA-DPB1 and disease have been sparse. The information on HLA-DPB1 is incomplete, with 145 alleles identified yet only 17 complete sequences available in the database. There is room for growth and discovery for HLA-DPB1. The depiction of this gene has garnered little interest in

the past, but has recently become an area of active research. HLA-DPB1 has been recently associated with insulin dependent diabetes mellitus (IDDM)[1], opticospinal multiple sclerosis [2] and berylliosis [3]. Blocking of HLA-DP receptors with specific antibodies led to a subdued activation response of TCCs by PPD (80% reduction in activation) [4]. The association of HLA-DP and susceptibility to acute contact dermatitis is worth investigating.

The population of patients with acute contact dermatitis to para-phenylenediamine (ACD to PPD) showed moderate to severe reactions. The sequences from these patients would give us a small window into any possible associations between HLA-DPB1 and susceptibility to ACD.

Of the 16 patients studied, 15 patients provided sequences that could be used for analysis. The allele HLA-DPB1*04:01:01 was predominately present in the samples studied (60% of patients), with 4 patients presenting as homozygous for the allele (Table 4.5). The patient population did not represent an evenly distributed population, but seemed to be predominately HLA-DPB1*04:01:01 populated. When compared with the control population, the distribution of the HLA-DPB1 alleles was not significantly different (Table 4.6). The allelic frequencies were compared to the control population by the Fisher's exact test on SSPS 12.0 with the help of Dr. Thomas Churchill, in the Department of Surgery at the University of Alberta. The p-values were examined to determine whether the null hypothesis can be rejected if the values were lower than 0.05. The p-

values we acquired were between 0.5 to 1.0, accepting the null hypothesis and showing that there was no direct distinction between the control population and the patient population. The sample size of our study affects the statistical power of the results obtained. In order to calculate the minimum sample size required to obtain statistical results that are truly significant, we can apply the data obtained from this study towards power analysis [7].

Work in the Elliott lab has collected 35 more samples from patients with ACD to PPD. This will increase the patient population to 50 and will increase the confidence of the study results. Future studies should also include a control population studied alongside the patient samples and subjected to the same methods.

These results indicate that there is no significant difference between the patient population and the control population and inform us that the procedure developed to analyze these patients worked as predicted. Although the study by *Sieben et al.* used a B7.21 monoclonal antibody to block the HLA-DP heterodimer from signaling between PBMCs and TCCs, the antibody was not HLA-DPB1 specific. The association between HLA-DP and ACD to PPD may be dependant on the alpha chain of the HLA-DP heterodimer, and future studies may determine if this is true.

The sequences recovered from the patient study were complete sequences from start codon to stop codon. The patient samples provided 12

unique sequences (Table 4.4), of which only six have complete sequence files in the HLA database [5, 6]. The procedure was able to complete most of the unknown sequences of several alleles, and in the future will enable the near completion of many other alleles with possible detection of new alleles. Since the ends of the cloned gene are composed of the primer sequences used for amplification, these nucleotides cannot be considered part of the coding sequence. There are several criteria that are required to submit a sequence to the HLA database and this will be completed in the future in the Elliott laboratory.

Although sequencing complete HLA-DPB1 transcripts allow us to gain more insight into the HLA-DPB1 gene, there are several regions outside of the coding sequences (introns, promoters, enhancers, etc.) that may contain valuable information about the expression and function of certain HLA-DPB1 alleles. Single nucleotide polymorphisms (SNPs) present in these regions may affect the overall HLA-DPB1 phenotype, which in turn could affect the gene's association with disease susceptibility or protection.

The described procedure can also be applied to other HLA genes (A, B, C, DQ (A/B), DR (A/B) and DPA1). Setting up universal primer pairs to accurately clone full transcripts of the other HLA genes is a challenging task that can be developed for future study in the Elliott lab. The allelic variability in other genes may make the development of one set of primers impossible, and may depend

on the development of several pairs of primers to fully clone these alleles (HLA-DRB1 has 873 alleles in total).

HLA typing has been a rapidly evolving field, with many new advances using complex software algorithms for deducing allelic assignments (SBTengine software by Qiagen) together with the accessibility of high throughput sequencing companies offering competitive prices. The solution hybridization protocols and the Lifecodes HLA-SSO typing kits (Gen-Probe Transplant Diagnostics) use Luminex technology to HLA type class I and class II HLA alleles from genomic samples. This protocol uses some of the principles outlined in the solution hybridization section (Chapter 3) to HLA type patient samples. The advancement of typing technology can only improve with time and hopefully will be more accessible and cost-effective in the future.

Automation of the procedures in this study could enable the ability to process and analyze a large number of samples for HLA-DPB1. Advancing the possibilities of this technique will allow the discovery of unfinished sequences as well as the discovery of new alleles in the HLA-DPB1 gene. The HLA typing procedures outlined in this study provide rapid, efficient sequence based typing of HLA-DPB1 and the framework for sequence based typing of other HLA genes.

5.6 References

1. Noble, J.A., et al., *The HLA class II locus DPB1 can influence susceptibility to type 1 diabetes.* Diabetes, 2000. **49**(1): p. 121-5.
2. Field, J., et al., *A polymorphism in the HLA-DPB1 gene is associated with susceptibility to multiple sclerosis.* PLoS One. **5**(10): p. e13454.
3. Richeldi, L., R. Sorrentino, and C. Saltini, *HLA-DPB1 GLUTAMATE-69 - A GENETIC-MARKER OF BERYLLIUM DISEASE.* Science, 1993. **262**(5131): p. 242-244.
4. Sieben, S., et al., *Delayed-type hypersensitivity reaction to paraphenylenediamine is mediated by 2 different pathways of antigen recognition by specific alphabeta human T-cell clones.* J Allergy Clin Immunol, 2002. **109**(6): p. 1005-11.
5. Robinson, J., et al., *IMGT/HLA database--a sequence database for the human major histocompatibility complex.* Tissue Antigens, 2000. **55**(3): p. 280-7.
6. Robinson, J., et al., *The IMGT/HLA database.* Nucleic Acids Res, 2009. **37**(Database issue): p. D1013-7.
7. MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structur modeling. *Psychological Methods, 1,* 130149.
8. Roberts, J.D., et al., *Fidelity of two retroviral reverse transcriptases during DNA dependent DNA synthesis in vitro.* Mol Cell Biol, 1989. **9**(2): p. 469-76.

APPENDIX

SUPPLEMENTAL FIGURE

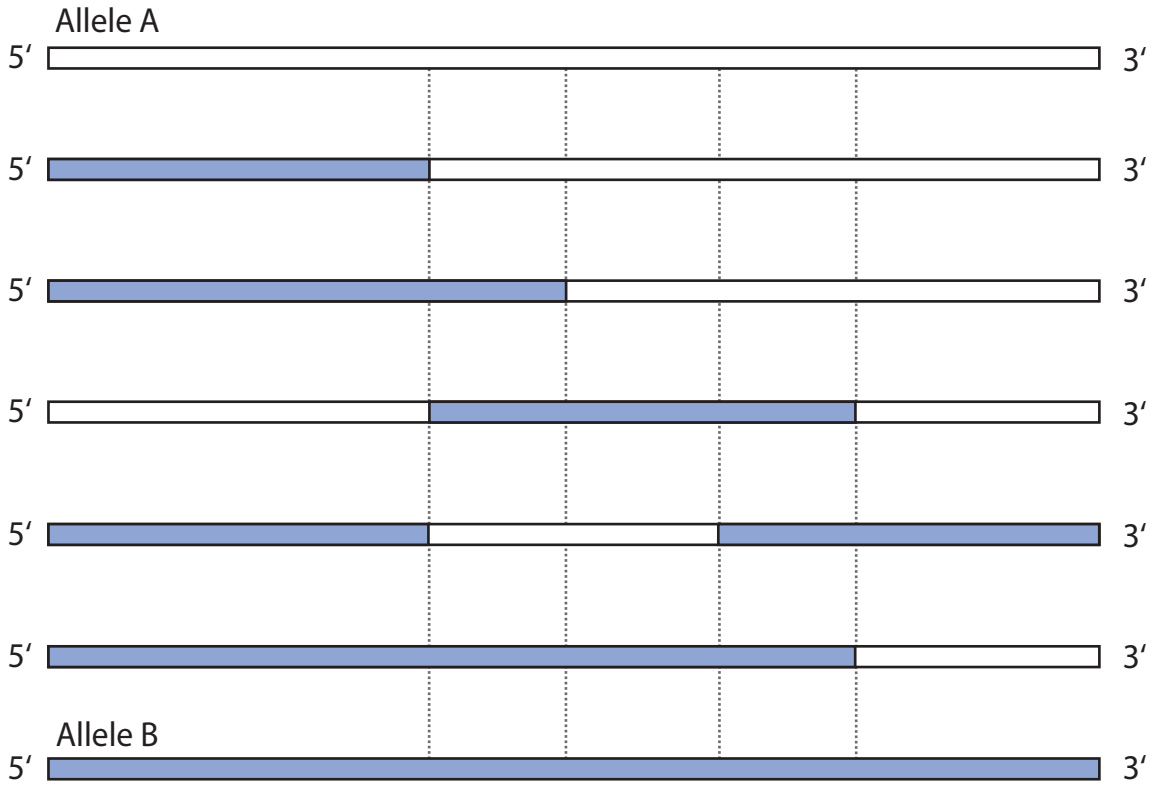


Figure A.1 Haplotype crossover. The sequences from patient 46 had different sequences that could not be assigned an allele haplotype. The sequences that were obtained are products of crossover between two alleles during the amplification. Haplotype crossovers can be produced from incomplete polymerase amplification of a given target, resulting in incomplete amplicons that act as primers during the next amplification cycle. This figure illustrates the possible crossover artifacts that can be created if crossing over occurs during PCR amplification.