Analysis of The DQ Family of The Human Leukocyte Antigens

Lu Shan
Computational Mol. Bio
1/17/01

-- Outline --

I. Brief Overview of the Human Leukocyte Antigens

Human leukocyte antigens, or HLA, are the human system of the major histocompatibility complex (MHC). These are glycoproteins expressed on the surface of every cell in the body. These molecules form part of a system of immune recognition, with the ability to distinguish "self" from "foreign" molecules. Their main role is to bind peptides from foreign antigens and present them to T cells to enable the specific destruction of "non-self" cells.[1]

The genetic components for the HLA protein molecules are located on the human chromosome 6p21 (the short arm). The overall organization of the human leukocyte antigens is shown in *Figure* 1. As shown in the figure, three classes of molecules (I, II, and III) have been identified. The class I and II molecules are encoded by multiple sets of different genes within the loci, but they display similar structure. In contrast, the class III region contains a diverse collection of genes, and the gene products show no established functional similarities.

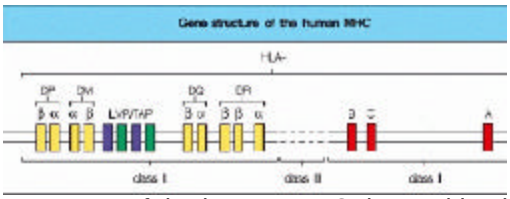

Figure 1. The gene structure of the human MHC, located in chromosome 6p21.[2]

[1] Roitt, I., Brostoff, J., and Male, D. Immunology. Barcelona, Spain:Mosby, 1996
[2] Current Biology, Ltd. Garland Publishing, 1997

HLA Class I antigens are further classified as HLA-A, -B and -C, according to the position of their encoding gene on the short arm of chromosome 6. The HLA I loci are highly polymorphic, corresponding to the molecules' ability to recognize a wide range of foreign substances. HLA class II molecules are heterodimeric glycoproteins consisting of an á- and â-chain as shown with the cartoon drawn in *Figure* 2. These class II proteins are further categorized as HLA-DR, -DQ and -DP. The family of proteins HLA-DQ is of the main interest in this paper.

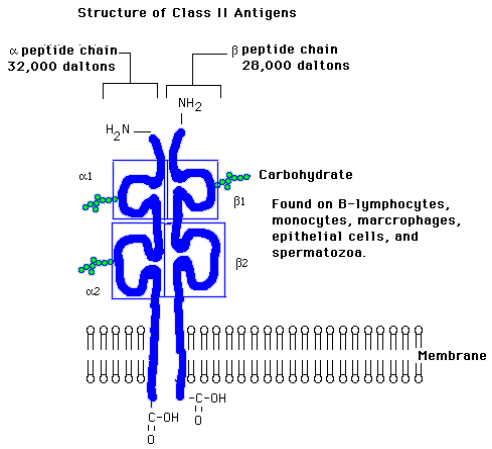

Figure 2. Cartoon of the structure of a class II HLA molecule. [3]

I.a. Functions and Characteristics of HLA-DQ Molecules

While the MHC class I proteins are expressed almost ubiquitously, the class II molecules are expressed mainly on specialized antigen presenting cells, such as B cells, macrophages and dendritic cells. The class II HLA molecules present peptide fragments for CD4+ T helper lymphocytes. Following synthesis of the á and â chains in the rough endoplasmic reticulum in the cell, both are transported through the Golgi compartment. They then join the endosomal/lysosomal compartment, where the foreign peptic fragments can be found and picked up, *en route* to the plasma membrane.[1]

The peptide binding site of the class II HLA molecules is the binding groove formed from the αβ heterodimer. This is illustrated by the crystal structure of HLA-DR1 bound to the peptide CLIP (see *Figure* 3). In the figure, the α chain is colored blue and the β chain is colored green. The CLIP peptide is colored green, fitting in the groove formed by the top of the α chain and the β chain. The binding groove in different HLA molecules contains variations in the amino acid sequence, giving rise to polymorphism.

[3] Natural Toxins Research Center, http://ntri.tamuk.edu/immunology/histocompatibility.html

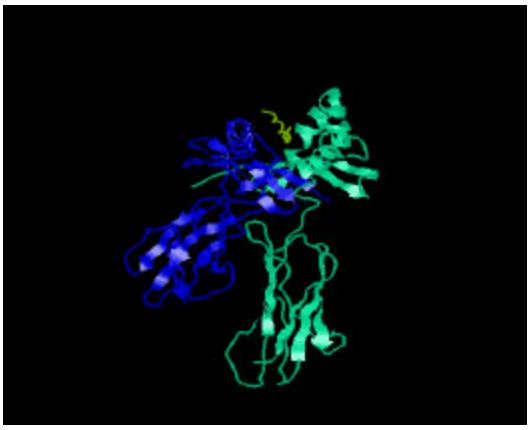

Figure 3. The crystal structure of HLA-DR1 bound with the CLIP peptide.[4]

The HLA-DQ molecules are unique compared to the rest of the MHC class II antigen-presenting molecules in several aspects. In HLA-DQ molecules, both polypeptide chains are polymorphic. In contrast to the á chain in HLA-DR molecules, the DQ á chain is more polymorphic. It contains a hypervariable loop between residues 48-56, one of the most polymorphic structures known. In addition to its higher level of polymorphism, the DQ molecules are expressed at a different level from the other class II molecules. Polymorphism is also found in the transcriptional control regions of the DQA (encoding for α chains) genes and DQB genes (encoding for β chains). It is thought that this could account for the functional differences observed among the different HLA-DQ alleles.[5]

More importantly, the human HLA-DQ molecules have been proposed to be involved in the pathogenesis of several human diseases, especially the T cell-mediated autoimmune diseases. It is been found that the HLA-DQ molecules are important in the pathogenesis of type 1 diabetes and Celiac Sprue (a gluten-sensitive gastrointestinal disease).[6] Particularly, the alleles of DQ genes that encode DQ8 (from alleles DQA1*0301 and DQB1*0302) and DQ2 (from alleles DQA1*0501 and DQB1*0201) confer the highest risk for both type 1 diabetes and Celiac Sprue. A recent report suggests that 95% of the celiac patients are with the genotype DQ2 and the remaining with DQ8.[7]

The investigation of the triggering mechanism in Celiac Sprue is a research interest of our lab, and my project currently looks into the natural modification of gliadin proteins (the wheat gluten proteins identified as the primary trigger of the autoimmune response). It is

[4] Swiss-PDB database. ID:1A6A
[5] Hjelmström, P. http://depts.washington.edu/rhwlab/dq/dq.html, 1996
[6] Wucherpfennig, K.W. *Current Opinion in Immunology*. 2001. **13**:650-656
[7] Sollid, L.M. *Annual Review of Immunology*. 2000. **18**:53-81

thought that the digested gliadin peptic fragments undergo selective deamidation (conversion of glutamine to glutamic acid) at the intestinal level by tissue transglutaminase. If that is the case, it is possible that the deamidated gliadin fragments gain the binding affinity to the DQ2 or DQ8 molecules and get displayed as foreign substances, causing T cells to attack the intestinal cells. Before jumping into the mechanistic studies of the deamidated gliadin binding with HLA-DQ2 molecules, an important piece of information can be gathered with the various techniques in bioinformatics. I am highly interested in comparing the binding motifs of the class II HLA-DQ molecules from the known alleles and their binding capabilities to both unmodified and modified gliadin peptides.

I.b. Nomenclature of the HLA Alleles

The HLA alleles are often written in the format of "HLA-DQB1*0201". The nomenclature is the following: a) "HLA" indicates it is a human leukocyte antigen gene; b) DQB1 represents the specific locus in the gene, i.e., DQB1; c) "0201" is the specific allele of the gene. In addition, the World Health Organization assigned the conventional nomenclature that, for example, names the DQ molecule that is encoded from the DQB1*0201 allele for the β chain by "DQ2"; the DQ molecules whose β chains are encoded from DQB1*0302 are called "DQ8". The totality of this information is compiled in the HLA dictionary.[8]

II. Searching for Human DQ Molecules and Performing Multiple-Sequence Alignment

II.a. BLAST Search

In order to find all the human DQ alleles currently known, PSI-BLAST searches (http://www.ncbi.nlm.nih.gov/BLAST/) are performed using either the nr database or swissprot based on a known DQA protein (PID g2135717). The statistically significant hits include human MHC class II antigens, as well as those of mice, bovine and rabbits. By checking the human antigens found from the search, only a very limited number of DQA alleles are there. DQA found from the nr and swissprot databases are organized in Table 1. Along with these proteins, DR and DX proteins are also picked up. Some of the DQB alleles are also picked up from the databases. It is somewhat disappointing that the search results do not seem to identify all of the DQA and DQB alleles from the BLAST searches. It is clear,

[8] Schreuder, G.M.Th., Hurley, C.K., Marsh, S.G.E., Lau, M., Maiers, M., Kollman, C., Noreen, H. *Tissue Antigens.* 1999. **54:**409-437

on the other hand, that swissprot has a much more complete compilation of the DQA family of proteins than the nr database.

PSI-BLAST in nr	DQA1*01, DQA1*02, DQA1*04
PSI-BLAST in swissprot	DQA1*05011, DQA1*02, DQA1*01, DQA1*05, DQA1*04, DQA1*06

Table 1. DQA alleles found in PSI-BLAST searches.

II.b. Web search

Turning my attention to web, I was able to find a more complete collection of DQA and DQB alleles compiled by the Anthony Nolan Research Institute. As of October 2001, a HLA database by the Anthony Nolan Research Institute contains 22 DQA alleles and 47 DQB alleles. These alleles are summarized in Table 2. This is a more satisfying result than the BLAST searches, though it is not clear why many of these alleles are not registered in databases such as swissprot.

HLA-DQA1		**HLA-DQB1**				
DQA1*01011	DQA1*0302	DQB1*0201	DQB1*0306	DQB1*05032	DQB1*0606	DQB1*0616
DQA1*01012	DQA1*0303	DQB1*0202	DQB1*0307	DQB1*0504	DQB1*0607	DQB1*0617
DQA1*01021	DQA1*0401	DQB1*0203	DQB1*0308	DQB1*06011	DQB1*0608	DQB1*0619
DQA1*01022	DQA1*05011	DQB1*03011	DQB1*0309	DQB1*06012	DQB1*0609	
DQA1*0103	DQA1*05012	DQB1*03012	DQB1*0310	DQB1*06013	DQB1*0610	
DQA1*01041	DQA1*0502	DQB1*0302	DQB1*0401	DQB1*0602	DQB1*06111	
DQA1*01042	DQA1*0503	DQB1*03032	DQB1*0402	DQB1*0603	DQB1*06112	
DQA1*0105	DQA1*0504	DQB1*03033	DQB1*05011	DQB1*06041	DQB1*0612	
DQA1*0106	DQA1*0505	DQB1*0304	DQB1*05012	DQB1*06042	DQB1*0613	
DQA1*0201	DQA1*06011	DQB1*03051	DQB1*0502	DQB1*06051	DQB1*0614	
DQA1*03011	DQA1*06012	DQB1*03052	DQB1*05031	DQB1*06052	DQB1*0615	

Table 2. The known alleles of HLA-DQ genes.[9]

II.c. Multiple-Sequence Alignment

With the sequence information of the known DQ alleles, multiple sequence alignment could be performed by ClustalW (on the Decypher machine). 22 DQA sequences and 47 DQB sequences are aligned, respectively. The resulted alignments are listed in Appendices I and II.

In the DQA alignment, the pairwise alignment scores range from 84% to 100% (percent identity), indicating a closely related family of proteins. The dendrogram is consisted of 4 major branches, DQA1*01 proteins, DQA1*02 and DQA1*03 proteins, DQA1*04 and DQA1*06 proteins, and DQA1*05 proteins. In the DQB alignment, the alignment scores range from 71% to 100%, much more diverse than the DQA family. The

[9] HLA database, Anthony Nolan Research Institute, http://www.ebi.ac.uk/imgt/hla/index.html, 2001

dendrogram also implies that there is much more dissimilarity in the protein pool: there are about 10-fold more branches and in many cases proteins that belong to one subgroup (such as DQB1*06) are separated out in the dendrogram. These dendrograms are shown in Figure 4.

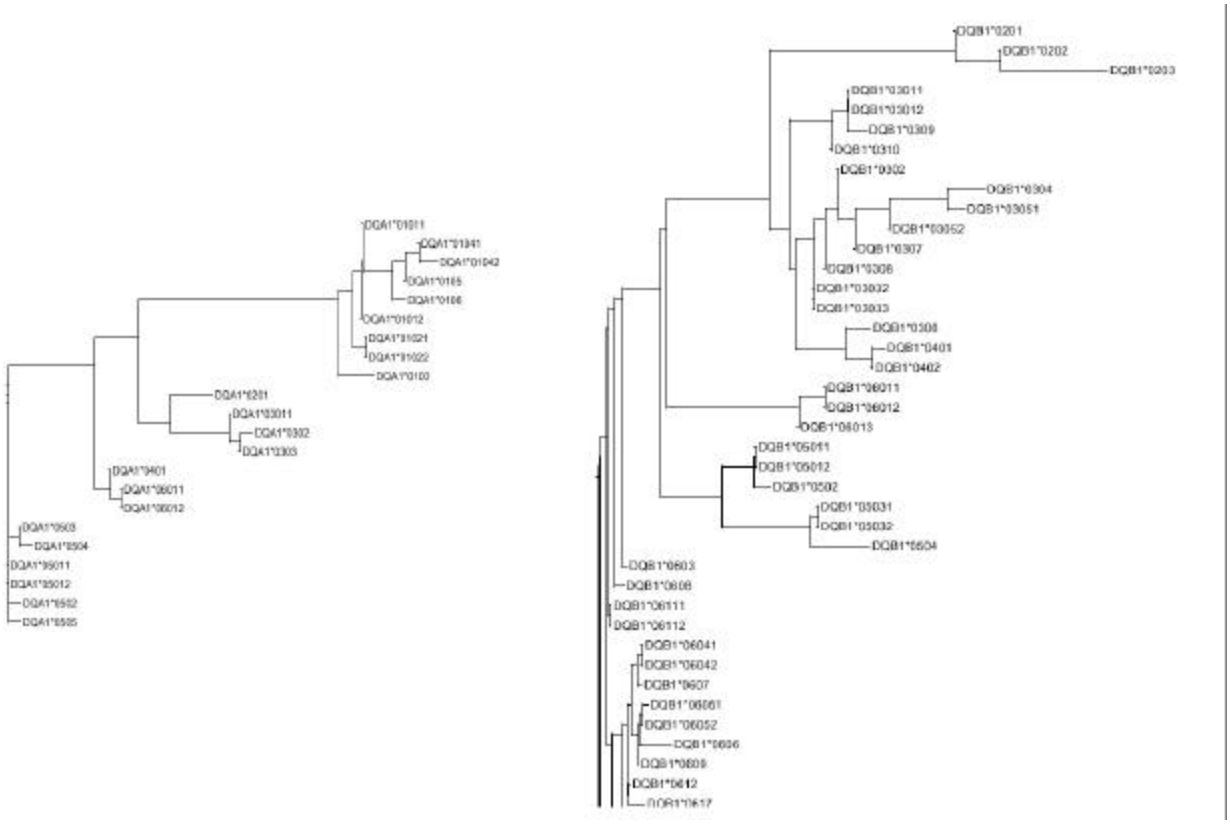

Figure 4. Dendrograms of DQA and DQB proteins

III. Analysis of DQ Allele Alignment and Motifs

III.a. Motifs in DQ Proteins

Emotif search (motif.stanford.edu/emotif) of the DQA1*0602 sequence returns 5 entries. All of the entries match to immunoglobulins and major histocompatibility proteins. A closer examination finds that only two signatures are found (IPB000495A and IPB000495B), with the patterns l.c....[filmv][fy]p..[ilv].[ilmv].w..[dn] and y.c.v.h..[fly]..p....w, respectively. Further eMotif scans using these patterns yield a wide range of immuglobulins, from class I HLA proteins to T cell receptor beta chains. Similar results for the DQA1 family of sequences indicate that these proteins possess the general function of an immunoglobulin but are probably not very specific in binding to substrates. On the other hand, eMotif search

with the DQB1*0201 yields 90 matches with scores over 99.3%, all of which are specifically labeled as class II histocompatibility antigen beta domain. Out of the 90 entries, 7 signatures are found (PF00969A~E and IPB000495A~B). Further eMotif scans give class II DQ beta proteins in human, rats, mice, pigs and etc. Having specific signatures for class II HLA antigen beta domain and matching to other DQ beta sequences, the DQB1 family seems to be much more specific than the DQA1 family.

BLOCKS+ search of DQB1*0201 retrieves 5 families, and the statistically significant ones are the following:

IPB000353 Class II histocompatibility antigen beta domain
IPB003006 Immunoglobulin and major histocompatibility proteins
IPB001039 Major histocompatibility complex proteins
IPB001003 MHC Class II, alpha chain, alpha-1

The first two families are similar to those found in the previous eMotif searches. The remaining families have lower combined expectation values. The statistically insignificant family is NADH-ubiquinone oxidoreductase chain, with an E-value of 0.73. In the IPB000353 family (HLA beta chain), 3 of 3 blocks are matched, tabulated below (Table 3) with the corresponding amino-acid locations. From the cartoon shown in Figure 2, it is reasonable to predict that the N-terminal block (amino acids 47-96) corresponds to the binding region of the beta chain. Similar searches with other DQB1 sequences yield similar results. In addition, BLOCKS+ searches using DQA1 sequences give the same families as the previous searches with DQB1 (in different orders).

Block	Location (amino acid)	E-Value
IPB000353A	47-96	2.2e-41
IPB000353B	125-174	3.5e-47
IPB000353C	176-230	1.8e-58

Table 3. Blocks matched for DQB1*0201 sequence

III.b. Analysis of DQ Protein Alignments

As shown in the multiple-sequence alignment, there are slight amino acid changes throughout both sets of alleles. In order to understand how the amino acid sequence variations correlate to the DQ polymorphism, as well as how they might impact the binding specificity of the protein products from the different alleles, I would like to look further into the alignment result and extract some useful information.

Based on the current knowledge of the binding motifs in MHC class II molecules, it is believed that the binding groove of the heterodimer contains 9 "pockets", each for the binding

of one amino acid in the bound peptide (since it usually contains 9 amino acids). The binding groove with the pockets is illustrated in *Figure* 4.

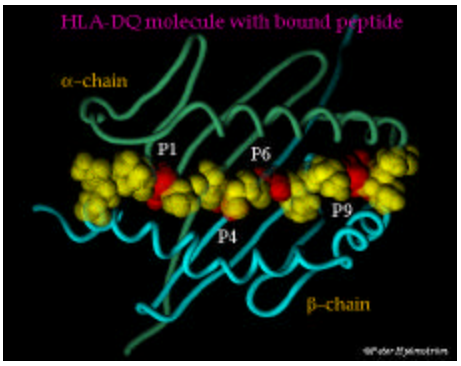

Figure 4. The binding groove of the DQ8 with a bound peptide.[5]

Furthermore, the residue positions for each pocket have been proposed for the DQ molecules and summarized in Table 4 below (with my corrections on the α chain residue numbering after examining). The residues for P7 and P8 in the α chain and P8 in the β chain are not specified because they are invariant in all known DQ sequences.

Pocket	P1	P2	P3	P4	P5	P6	P7	P8	P9
Reported residues in β chain	86, 87	77	74	13,26,28,71,74	70,71,74	9,30,70	30,47,67,70,71	-	9,30,37,38,57
Corrected residues in α chain	11, 34, 55, 56	11, 61	11, 25, 61, 64	11	61,64	69	-	-	74, 76

Table 4. The proposed residues in the binding pocket of DQ.[10] The "-" indicates the unlisted residues because they are invariant in all DQ alleles.

Based on the proposed residue positions for each binding pocket, I could examine the sequences and compare the amino acids in each binding pocket for every known allele protein product. The compilation of all of the α-chain contribution to the peptide binding environments is listed in Table 5.

Similarly, I examined the β-chain alleles and with the help of the multiple-sequence alignment, I could summarize which residues contribute to the binding pockets, as listed in Table 6.

[10] Baas, A., Gao, X., Chelvanayagam, G. *Immunogenetics*. 1999. **50**:8-15

	P1				P2		P3			P4	P5		P6	P9	
DQA1*01011	C	E	G	G	C	G	C	Y	G	R	C	G	R	A	I	M
DQA1*01012	C	E	G	G	C	G	C	Y	G	R	C	G	R	A	I	M
DQA1*01021	C	Q	G	G	C	G	C	Y	G	R	C	G	R	A	I	M
DQA1*01022	C	Q	G	G	C	G	C	Y	G	R	C	G	R	A	I	M
DQA1*0103	C	Q	G	G	C	G	C	Y	G	R	C	G	R	A	I	M
DQA1*01041	C	Q	G	G	C	G	C	Y	G	R	C	G	R	A	I	M
DQA1*01042	C	Q	G	G	C	G	C	Y	G	R	C	G	R	A	I	M
DQA1*0105	C	Q	G	G	C	G	C	Y	G	R	C	G	R	A	I	M
DQA1*0106	C	Q	G	-	C	G	C	Y	G	R	C	G	R	A	I	M
DQA1*0201	Y	Q	R	R	Y	F	Y	F	F	T	Y	F	T	L	I	L
DQA1*03011	Y	Q	R	R	Y	F	Y	Y	F	T	Y	F	T	L	I	V
DQA1*0302	Y	Q	R	R	Y	F	Y	Y	F	T	Y	F	T	L	I	V
DQA1*0303	Y	Q	R	R	Y	F	Y	Y	F	T	Y	F	T	L	I	V
DQA1*0401	Y	Q	R	-	Y	F	Y	Y	F	T	Y	F	T	L	I	L
DQA1*05011	Y	Q	R	-	Y	F	Y	Y	F	T	Y	F	T	L	S	L
DQA1*05012	Y	Q	R	-	Y	F	Y	Y	F	T	Y	F	T	L	S	L
DQA1*0502	Y	Q	R	-	Y	F	Y	Y	F	T	Y	F	T	L	S	L
DQA1*0503	Y	Q	R	-	Y	F	Y	Y	F	T	Y	F	T	L	S	L
DQA1*0504	Y	Q	R	-	Y	F	Y	Y	F	T	Y	F	T	L	S	L
DQA1*0505	Y	Q	R	-	Y	F	Y	Y	F	T	Y	F	T	L	S	L
DQA1*06011	Y	Q	R	-	Y	F	Y	F	F	T	Y	F	T	T	I	L
DQA1*06012	Y	Q	R	-	Y	F	Y	F	F	T	Y	F	T	T	I	L

Table 5. The residues in the peptide binding environment from the á alleles, compiled by examining the multiple-sequence alignment. The dash line "-" indicates the gap in the protein sequences.

	P1	P2	P3	P4					P5		P6		P7				P9			
DQB1*0201	E	L	R	A	G	L	S	K	A	R	K	A	Y	S	R	S	F	I	R	K	Y	S	I	V	A
DQB1*0202	E	L	R	A	G	L	S	K	A	R	K	A	Y	S	R	S	F	I	R	K	Y	S	I	V	A
DQB1*0203	E	L	R	A	G	L	S	K	A	R	K	A	Y	S	R	S	F	I	R	K	Y	S	I	V	D
DQB1*03011	E	L	T	E	A	Y	T	T	E	R	T	E	Y	Y	R	Y	Y	V	R	T	Y	Y	Y	A	D
DQB1*03012	E	L	T	E	A	Y	T	T	E	R	T	E	Y	Y	R	Y	Y	V	R	T	Y	Y	Y	A	D
DQB1*0302	E	L	T	E	G	L	T	T	E	R	T	E	Y	Y	R	Y	Y	V	R	T	Y	Y	Y	A	A
DQB1*03032	E	L	T	E	G	L	T	T	E	R	T	E	Y	Y	R	Y	Y	V	R	T	Y	Y	Y	A	D
DQB1*03033	E	L	T	E	G	L	T	T	E	R	T	E	Y	Y	R	Y	Y	V	R	T	Y	Y	Y	A	D
DQB1*0304	E	L	T	E	A	Y	T	T	E	R	T	E	Y	Y	R	Y	Y	V	R	T	Y	Y	Y	A	A
DQB1*03051	E	L	T	E	G	G	T	T	E	R	T	E	Y	Y	R	Y	Y	V	R	T	Y	Y	Y	A	A
DQB1*03052	E	L	T	E	G	G	T	T	E	R	T	E	Y	Y	R	Y	Y	V	R	T	Y	Y	Y	A	A
DQB1*0306	E	L	T	S	G	L	T	D	S	E	D	S	Y	Y	E	Y	Y	V	E	D	Y	Y	Y	A	D
DQB1*0307	E	L	T	E	G	L	T	T	E	R	T	E	Y	Y	R	Y	Y	V	R	T	Y	Y	Y	A	A
DQB1*0308	E	L	T	E	G	L	T	T	E	G	T	E	Y	Y	G	Y	Y	V	G	T	Y	Y	Y	A	A
DQB1*0309	E	L	T	E	A	Y	T	T	E	R	T	E	Y	Y	R	Y	Y	V	R	T	Y	Y	Y	A	D
DQB1*0310	E	L	T	E	A	Y	T	T	E	R	T	E	Y	Y	R	Y	Y	V	R	T	Y	Y	Y	A	D
DQB1*0401	E	L	T	S	G	G	T	D	S	E	D	S	F	Y	E	Y	Y	I	E	D	F	Y	Y	A	D
DQB1*0402	E	L	T	S	G	G	T	D	S	E	D	S	F	Y	E	Y	Y	I	E	D	F	Y	Y	A	D
DQB1*05011	A	Y	R	S	G	G	T	A	S	G	A	S	Y	H	G	H	Y	V	G	A	Y	H	Y	V	V
DQB1*05012	A	Y	R	S	G	G	T	A	S	G	A	S	Y	H	G	H	Y	V	G	A	Y	H	Y	V	V
DQB1*0502	A	Y	R	S	G	G	T	A	S	G	A	S	Y	H	G	H	Y	V	G	A	Y	H	Y	V	S
DQB1*05031	A	Y	R	S	G	G	T	A	S	G	A	S	Y	H	G	H	Y	V	G	A	Y	H	Y	V	D
DQB1*05032	A	Y	R	S	G	G	T	A	S	G	A	S	Y	H	G	H	Y	V	G	A	Y	H	Y	V	D
DQB1*0504	A	Y	R	S	G	G	T	D	S	E	D	S	Y	Y	E	Y	Y	I	E	D	Y	Y	Y	V	S
DQB1*06011	A	F	T	E	A	Y	T	T	E	R	T	E	L	Y	R	Y	Y	I	R	T	L	Y	D	V	D
DQB1*06012	A	F	T	E	A	Y	T	T	E	R	T	E	L	Y	R	Y	Y	I	R	T	L	Y	D	V	D
DQB1*06013	A	F	T	E	A	Y	T	T	E	R	T	E	L	Y	R	Y	Y	I	R	T	L	Y	D	V	D
DQB1*0602	A	F	T	E	G	L	T	T	E	G	T	E	F	Y	G	Y	Y	V	G	T	F	Y	Y	A	D
DQB1*0603	A	F	T	E	G	L	T	T	E	G	T	E	Y	H	G	H	Y	V	G	T	Y	H	Y	A	D
DQB1*06041	G	Y	T	E	G	L	T	T	E	R	T	E	Y	H	R	H	Y	V	R	T	Y	H	Y	A	V
DQB1*06042	G	Y	T	E	G	L	T	T	E	R	T	E	Y	H	R	H	Y	V	R	T	Y	H	Y	A	V
DQB1*06051	G	Y	T	E	G	L	T	T	E	R	T	E	Y	Y	R	Y	Y	V	R	T	Y	Y	Y	A	V
DQB1*06052	G	Y	T	E	-	L	T	T	E	R	T	E	-	Y	R	Y	Y	V	R	T	-	Y	Y	A	V
DQB1*0606	G	Y	R	A	-	L	T	T	A	R	T	A	-	Y	R	Y	Y	V	R	T	-	Y	Y	A	V
DQB1*0607	G	Y	T	E	G	L	T	T	E	R	T	E	Y	H	R	H	Y	V	R	T	Y	H	Y	A	D
DQB1*0608	A	F	T	E	G	L	T	T	E	G	T	E	Y	H	G	H	Y	V	G	T	Y	H	Y	A	V
DQB1*0609	G	Y	T	E	G	L	T	T	E	R	T	E	Y	Y	R	Y	Y	V	R	T	Y	Y	Y	A	V
DQB1*0610	A	F	T	E	G	L	T	T	E	G	T	E	F	Y	G	Y	Y	V	G	T	F	Y	Y	A	S
DQB1*06111	A	F	T	E	G	L	T	T	E	G	T	E	Y	Y	G	Y	Y	V	G	T	Y	Y	Y	A	D
DQB1*06112	A	F	T	E	G	L	T	T	E	G	T	E	Y	Y	G	Y	Y	V	G	T	Y	Y	Y	A	D
DQB1*0612	G	Y	T	E	G	L	T	T	E	G	T	E	F	Y	G	Y	Y	V	G	T	F	Y	Y	A	V
DQB1*0613	A	F	T	E	G	L	T	T	E	G	T	E	F	Y	G	Y	Y	V	G	T	F	Y	Y	A	V
DQB1*0614	A	F	T	E	G	L	T	T	E	G	T	E	F	H	G	H	Y	V	G	T	F	H	Y	A	D
DQB1*0615	G	Y	T	E	G	L	T	T	E	R	T	E	F	Y	R	Y	Y	V	R	T	F	Y	Y	A	D
DQB1*0616	A	F	T	E	G	L	T	T	E	G	T	E	F	Y	G	Y	Y	V	G	T	F	Y	Y	A	D
DQB1*0617	G	Y	T	E	G	L	T	A	E	G	A	E	F	H	G	H	Y	V	G	A	F	H	Y	A	V
DQB1*0618	A	F	T	E	G	L	T	T	E	R	T	E	F	Y	R	Y	Y	V	R	T	F	Y	Y	A	V
DQB1*0619	A	F	T	E	G	L	T	T	E	G	T	E	F	Y	G	Y	Y	V	G	T	F	Y	Y	A	D

Table 6. The residues in the peptide binding environment from the á alleles, compiled by examining the multiple-sequence alignment. The dash line "-" indicates the gap in the protein sequences.

The compilation of these results yields some interesting observations. First of all, there is a lot more polymorphism in the â alleles than the á ones. This can be observed by comparing the columns in the á table to the ones in the â table. For each residue position, there are mostly only 2 amino acid variations among the á alleles, while there are 3-4 different amino acids among the â alleles. In addition, there are more amino acids that have been identified to be involved in the binding pockets in the â chains than in the á chains. This indicates that the â chains might play a more important role than á in both the specific binding of peptide fragments and the polymorphism of the DQ molecules.

Secondly, by examining Table 5, it is easy to tell that the peptide binding pockets are clustered around the N-terminal residues (all of the identified residues are under #100). However, the multiple-sequence alignment data (Appendices I and II) reveal amino acid changes throughout the whole sequences. In particular, the C-terminal ends of the protein family are very different. Even though these differences might not be involved in the specificity of peptide binding, they might be important in determining which á and â chains actually preferentially form heterodimers since not all of the combinations are found. This could be an interesting topic for exploration later.

Thirdly, I am interested to look at some of the specific amino acid changes. From Table 6, it shows that the amino acid variations can bring about significant environmental changes to the binding pockets. For example, in the P1 binding environment, the amino acids vary from cysteine to tyrosine, from glutamic acid to glutamine and from glycine to arginine. The environment changes from polar to large hydrophobic, from neutral to positively charge or negatively charged. This drastic change accommodates the ability to bind to a large spectrum of foreign peptides.

IV. Comparison of DQ2 and DQ8 Molecules

IV.a. Alignment Comparison of DQ2 and DQ8 Beta Chains

Since Celiac Sprue is of my interest, I would like to look further into the DQ2 and DQ8 molecules, which are found to be in strong linkage with the disease. In order to compare the DQ2 and DQ8 beta chains (the beta chains are chosen because of their relative significance over the alpha chains), the DQA1*0201 and DQB1*0302 sequences are aligned using ALION (http://fold.stanford.edu/alion/).

Using the BLOSUM 62 matrix and the Smith-Waterman local alignment algorithm, the alignment finds 94% match in the two sequences (with 92% residue identity and no gaps). There are 12 mismatches in the sequences and 7 of them are found to be conservative substitutions.

The SeqWeb program “Pretty” has a benefit of aligning sequences in colors and constructing a consensus sequence. The alignment is generated with the same position-specific scoring matrix and is displayed below.

```
           1                                                    50
 DQ2_beta  MSWKKALRIP GGLRAATVTL MLSMLSTPVA EGRDSPEDFV YQFKGMCYFT
 DQ8_beta  MSWKKALRIP GGLRVATVTL MLAMLSTPVA EGRDSPEDFV YQFKGMCYFT
Consensus  MSWKKALRIP GGLR-ATVTL ML-MLSTPVA EGRDSPEDFV YQFKGMCYFT
           51                                                   100
 DQ2_beta  NGTERVRLVS RSIYNREEIV RFDSDVGEFR AVTLLGLPAA EYWNSQKDIL
 DQ8_beta  NGTERVRLVT RYIYNREEYA RFDSDVGVYR AVTPLGPPAA EYWNSQKEVL
Consensus  NGTERVRLV- R-IYNREE-- RFDSDVG--R AVT-LG-PAA EYWNSQK--L
           101                  120
 DQ2_beta  ERKRAAVDRV CRHNYQLELR
 DQ8_beta  ERTRAELDTV CRHNYQLELR
Consensus  ER-RA--D-V CRHNYQLELR
```

As seen in the consensus sequence, the “-“ lines indicate the positions of mismatches. An immediate observation is that several key substitutions from DQ2 to DQ8 involve the replacement smaller residues with tyrosine and replacement of a positive residue to a neutral residue. A more detailed residue comparison will be discussed in section IV.b.

A more graphical comparison of DQ2 and DQ8 beta chains is performed by Compare (http://pmgm2.stanford.edu/gcg-bin/seqweb.cgi) and shown in Figure 5. As shown below, the two sequences are very similar for the most part, yet small but perhaps significant distinctions are picked up by the alignment.

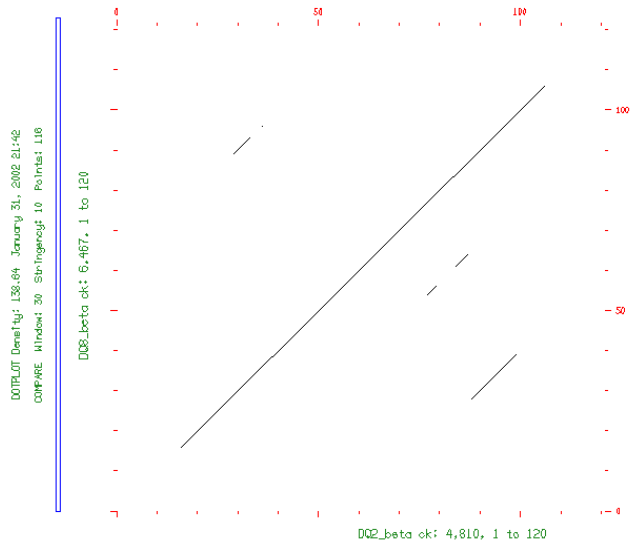

Figure 5. Graphical comparison of DQ2 and DQ8 beta sequences generated by “Compare”

IV.b. Further Residue Comparison of DQ2, DQ8 and a Control Sequence

Again, the alleles of the molecules I am interested in are DQA1*0501, DQB1*0201 (DQ2) and DQA1*0301, DQB1*0302 (DQ8). In addition, as a control, I choose to also look at a closely related DQ molecule (DQA1*0501, DQB1*0301), which does not confer susceptibility to Celiac Sprue.[7]

The binding environments of all three molecules are summarized in Table 7 with some of the differences colored in red.

	P1	P2	P3	P4	P5	P6	P7	P9
DQ2 á/â chains	Y,Q,R,-,E,L	Y,F,R	Y,Y,F,T,A	Y,G,L,S,K,A	F,T,R,K,A	L,Y,S,R	S,F,I,R,K	S,L,Y,S,I,V,A
DQ8 á/â chains	Y,Q,R,R,E,L	Y,F,T	Y,Y,F,T,E	Y,G,L,T,T,E	R,T,R,T,E	L,Y,Y,R	Y,Y,V,R,T	V,Y,Y,Y,A,A
Control á/â chains	Y,Q,R,-,E,L	Y,F,T	Y,Y,F,T,E	Y,A,Y,T,T,E	F,T,R,T,E	L,Y,Y,R	Y,Y,V,R,T	S,L,Y,Y,Y,A,D

Table 7. Comparison of the binding environments of DQ2, DQ8 and a control (DQA1*0501, DQB1*0301). The dash line "-" indicates a gap in the protein sequence during alignment.

From the compiled binding pocket environments of the three molecules, several observations can be drawn. First, the DQ2 molecule seems to show the most abundance in positively charged residues R and K and less negatively charged residues E. Compared to DQ8 and the control, DQ2 contains an extra arginine in P2 and extra lysines in P4, P5 and P7, replacing the threonine. In both P3 and P4, DQ2 contains an alanine in place of glutamic acid in DQ8 and the control. Secondly, although it seems that the DQ8 molecules shares tremendous similarity with the control molecule, there are still some differences to point out. DQ8 contains one more positively charged residue, in P5, with R instead of the F in the control. In P9, DQ8 contains one less negatively charged residue, with A instead of the D in the control. Overall speaking, the positively charged environment is the strongest in DQ2, then DQ8 and the weakest in the control molecule.

From the comparison above and based on the knowledge that Celiac Sprue occurs over 90% of the time with DQ2-inheritance, about 5% of the time in DQ8 inheritance, and 0% in people with only the control DQ molecules, it is reasonable to hypothesize that negatively charged peptides confer the preferential binding to the MHC class II molecules and cause the immunogenetic response (since DQ2 molecules constitute a highly positively charged environment). In deed, this hypothesis jibes well with the finding that anti-transglutaminase autoantibodies are uniquely expressed in Celiac Sprue patients. What transglutaminase can do is to deamidate the wheat gliadin peptides to create more glutamic acid. The negatively charged peptide product can in turn preferentially bind to DQ2, to a lesser degree to DQ8 and not to the control. The different levels of positive environments in DQ2 and DQ8 imply

that perhaps DQ8 requires the bound peptide to contain more negative residues. In addition, because of the positive environment in DQ2 and DQ8, the neutral, unmodified gliadin peptides do not bind to them.

Another observation is that the additional positively-charged residues in DQ2 come from the â allele. From comparing the â alleles in the DQ family it seems that the DQB1*02 alleles allocates the most positive charges. This is consistent with the previous postulation that the â chain contributes more to the specific binding of the peptides. It is also consistent with the fact that the World Health Organization decides to assign conventional names such as "DQ2" or "DQ8" merely after the â alleles.

V. Structural Alignment and Analysis of DQ

The structures of DQ proteins have remained to be an open field of research. From a 3motif (http://motif.stanford.edu/3motif/) search, only a long list of solved DR molecules is found. Very recently, however, the structure of DQ8 bound to an insulin peptide has been resolved by Don Wiley's group.[11] The structure is shown below in Figure 6. This breakthrough provides the basis for reliable protein structure prediction.

Figure 6. Crystal structure of DQ8 bound with an insulin peptide. The peptide is colored gray and shown on the top inside the groove. This structure is taken from SwissPDB (1JK8).

PlotStructure at SeqWeb provides both Chou-Flashman and Garnier-Robson methods of predicting the secondary structure. Both the DQB1*0201 and DQB1*0302 are put into the program, and displayed side by side in Figure 7. In the plots, helices are shown with a sine wave, beta-sheets with a sharp saw-tooth wave, turns with 180-degree turns, and coils with a dull saw-tooth wave. In these Chou-Flashman secondary structure predictions, it

[11] Lee, K. H., Wucherpfennig, K. W., Wiley, D. C. *Nat.Immunol.* 2001. **2**:501

is shown that the overall secondary structures of DQ2 and DQ8 beta chains are very similar. The slight variations are manifested in both the N-terminal ends and the C-termini. In the N-terminal region, in the second turn, a beta sheet in DQ8 replaces one region of the alpha helix in the DQ2 chain. A similar replacement is also in the third turn and the fourth turn. There are also local hydrophobicity/hydrophilicity changes.

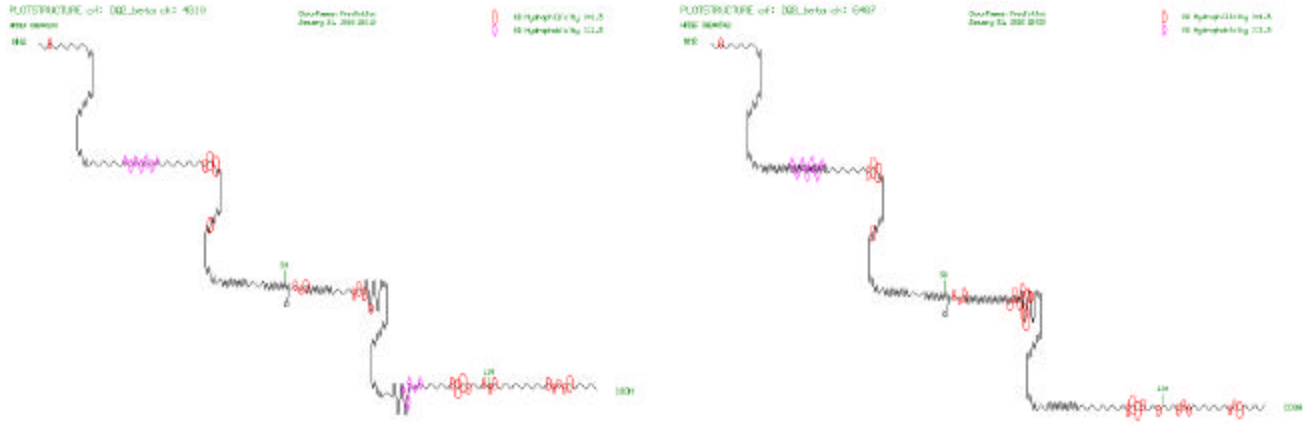

Figure 7. (a) DQ2 beta secondary structure (b) DQ8 beta secondary structure prediction prediction

In addition to making secondary structure predictions, it is also possible to align a sequence with a similar sequence of known structure. The FUGUE alignment server (http://www-cryst.bioc.cam.ac.uk/cgi-bin/cgiwrap/kenji/fugue/align.cgi) can help achieve this purpose. Entering the sequence of the DQ2 beta chain to compare with the recently solved DQ8 structure, the following alignment is retrieved. The FUGUE server is able to align the two sequences up and based on the alignment to predict where the beta sheets and alpha helices should also occur in the DQ2 unknown structure.

```
1jk8    (   2 )    vadhvASYGVNLyQsygpsGqYSHeFdgDEEFyVdlerketvWqlplFrr
13434              --------------------------------------------------
                      bbbbbbbbbbb    bbbbbbbb  bbbbbbb    bbbb  aaaaa

1jk8    (  51 )    fRrFdpqfAltNiavLkhNLniVikrSnstaAtnevPeVtVfskspvtlg
13434              --------------------------------------------------
                   a    aaaaaaaaaaaaaaaaaaaa           bbbbbb

1jk8    ( 101 )    qpNtLIÇlVdnIFPPvVnItWlsnghsvtegvsetsfLsksdhSFfKISy
13434              ------------------------------------------------MSW
                     bbbbbbbbbb     bbbbbb  bbb   bbb   bb     bbbbbb

1jk8    ( 151 )    Ltfl--PsddeiYdÇkVeHwgldepllkhwepespeDfVYQFKGmÇyFtn
13434              KKALRIPGGLRAATVTLMLSMLSTPVAEGRD--SPEDFVYQFKGMCYFTN
                   bbb         bbbbbb       bbbbb        bbbbbbbbbbb
```

```
1jk8    (  20 )    gterVrLVTrYIYnreeYArFdsdvgvYravtplGppAaeyWnsqkevle
13434              GTERVRLVSRSIYNREEIVRFDSDVGEFRAVTLLGLPAAEYWNSQKDILE
                      bbbbbbbbbb  bbbbbbb     bbb     aaaaaaaa  aaaaa

1jk8    (  70 )    rTraeldtVÇrhNyqLElrTTlqrrvePtVtIspsrtealnhhnlLvÇsV
13434              RKRAAVDRVCRHNYQLELRTTLQRRVEPTVTISPSRTEALNHHNLLVCSV
                   aaaaaaaa aaaaaaaaa  333     bbbbbb          bbbbbb

1jk8    ( 120 )    TdFYpaqikVrWfrndqeettgvvstplirNgdWtFQilVmLemtpqrgd
13434              TDFYPAQIKVRWFRNDQEETAGVVSTPLIRNGDWTFQILVMLEMTPQRGD
                   bbb     bbbbbb  bbb   bbb   bb     bbbbbbbb

1jk8    ( 170 )    vYtÇhVeHpSlqnpiivewrAqslvEalYlVcgErgg
13434              VYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGIGGFVLGLIFLGLGLII
                   bbbbbbb       bbbbbb
1jk8
13434              HHRSQKGLLH
```

A few more words on the solved DQ8 structure: Since this structure is only recently released, the motif is not yet classified for motif searches. A detailed view of the binding groove is shown in Figure 8.[6] The insulin peptide consists of two negatively charged residues (E) and it can be seen that in deed the glutamic acid residues in the P1 and P9 positions of the peptide are immersed in the positive (blue) pockets in the DQ8 molecules.

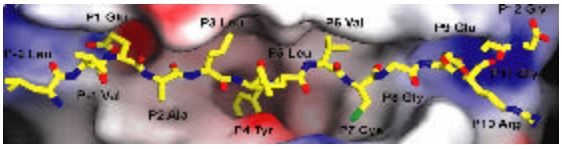

Figure 8. The binding groove of the DQ8 molecule with the insulin peptide. The positively charged pocket of DQ8 is colored blue and the negatively charged region is colored red.

VI. Summary

In summary, HLA class II molecules play an important role in the defense mechanism in our immune system. In particular, the DQ family, being very polymorphic, is a significant part of the major histocompatibility complex; however, it is also highly implicated in a series of autoimmune diseases such as Celiac Sprue and Type 1 diabetes. The tools of bioinformatics have given me the opportunity to perform detailed analysis on the DQ family and understand the similarities and differences among the protein sequences and structures. The database searches for the DQ family of proteins have yielded a number of results, but also show the current limitations in relying merely on the databases since sometimes they might not be updated. The multiple-sequence alignment and the motif analysis provide a

clearer picture of the range of polymorphism present in the DQ family. They also give insight into the similarity and distinctions within the family, as well as how the differences may contribute to the different functional roles played by DQA and DQB. Both statistical and residue comparisons among DQ2, DQ8 and another DQ molecule are made possible by a wide variety of methods and provides insight into the antigenic peptide binding in Celiac Sprue. In addition, secondary structure analysis is preformed and aided by a newly resolved crystal structure of peptide bound DQ8. This newly found structure also seems to be consistent with the functional analysis made earlier in paper. This exercise has revealed a lot more information about the HLA DQ family to me and enriched my understanding of how these molecules can contribute the binding of antigenic peptides. I foresee that future efforts involving detailed sequence alignments and epitope analyses of related food grain proteins such as oats and barley (which also have been found to trigger immune response in celiac patiens to some extent), although out of the scope of this paper, will be of great interest too.

Appendix I. DQA1 Alleles Alignment

DeCypher Results for: ClustalW Multiple Alignment

```
                                                  1234567890123456789012345678901234567
DQA1*01011      MILNKALLLGALALTTVMSPCGGEDIVADHVASCGVNLYQFYGPSGQYTHEFDGDEEFYV
DQA1*01012      MILNKALLLGALALTTVMSPCGGEDIVADHVASCGVNLYQFYGPSGQYTHEFDGDEEFYV
DQA1*01021      MILNKALLLGALALTTVMSPCGGEDIVADHVASCGVNLYQFYGPSGQYTHEFDGDEQFYV
DQA1*01022      MILNKALLLGALALTTVMSPCGGEDIVADHVASCGVNLYQFYGPSGQYTHEFDGDEQFYV
DQA1*0103       MILNKALLLGALALTTVMSPCGGEDIVADHVASCGVNLYQFYGPSGQFTHEFDGDEQFYV
DQA1*01041      MILNKALLLGALALTTMMSPCGGEGIVADHVASCGVNLYQFYGPSGQYTHEFDGDEEFYV
DQA1*01042      MILNKALLLGALALTTMMSPCGGEGIVADHVASCGVNLYQFYGPSGQYTHEFDGDEEFYV
DQA1*0105       MILNKALLLGALALTTMMSPCGGEGIVADHVASCGVNLYQFYGPSGQYTHEFDGDEEFYV
DQA1*0106       MILNKALLLGALALTTMMSPCGGEGIVADHVASCGVNLYQFYGPSGQYTHEFDGDEQFYV
DQA1*0201       MILNKALMLGALALTTVMSPCGGEDIVADHVASYGVNLYQSYGPSGQFTHEFDGDEEFYV
DQA1*03011      MILNKALMLGALALTTVMSPCGGEDIVADHVASYGVNLYQSYGPSGQYSHEFDGDEEFYV
DQA1*0302       MILNKALMLGALALTTVTSPCGGEDIVADHVASYGVNLYQSYGPSGQYSHEFDGDEEFYV
DQA1*0303       MILNKALMLGALALTTVMSPCGGEDIVADHVASYGVNLYQSYGPSGQYSHEFDGDEEFYV
DQA1*0401       MILNKALLLGALALTTVMSPCGGEDIVADHVASYGVNLYQSYGPSGQYTHEFDGDEQFYV
DQA1*05011      MILNKALMLGALALTTVMSPCGGEDIVADHVASYGVNLYQSYGPSGQYTHEFDGDEQFYV
DQA1*05012      MILNKALMLGALALTTVMSPCGGEDIVADHVASYGVNLYQSYGPSGQYTHEFDGDEQFYV
DQA1*0502       MILNKALMLGALALTTVMSPCGGEDIVADHVASYGVNLYQSYGPSGQYTHEFDGDEQFYV
DQA1*0503       MILNKALMLGALALTTVMSPCGGEDIVADHVASYGVNLYQSYGPSGQYTHEFDGDEQFYV
DQA1*0504       MILNKALMLGALALTTVMSPCGGEDIVADHVASYGVNLYQSYGLSGQYTHEFDGDEQFYV
DQA1*0505       MILNKALMLGTLALTTVMSPCGGEDIVADHVASYGVNLYQSYGPSGQYTHEFDGDEQFYV
DQA1*06011      MILNKALLLGALALTTVMSPCGGEDIVADHVASYGVNLYQSYGPSGQFTHEFDGDEQFYV
DQA1*06012      MILNKALLLGALALTTVMSPCGGEDIVADHVASYGVNLYQSYGPSGQFTHEFDGDEQFYV
                ********:**:*****: ******.******** ****** ** ***::*******:***

DQA1*01011      DLERKETAWRWPEFSKFGGFDPQGALRNMAVAKHNLNIMIKRYNSTAATNEVPEVTVFSK
DQA1*01012      DLERKETAWRWPEFSKFGGFDPQGALRNMAVAKHNLNIMIKRYNSTAATNEVPEVTVFSK
DQA1*01021      DLERKETAWRWPEFSKFGGFDPQGALRNMAVAKHNLNIMIKRYNSTAATNEVPEVTVFSK
DQA1*01022      DLERKETAWRWPEFSKFGGFDPQGALRNMAVAKHNLNIMIKRYNSTAATNEVPEVTVFSK
DQA1*0103       DLEKKETAWRWPEFSKFGGFDPQGALRNMAVAKHNLNIMIKRYNSTAATNEVPEVTVFSK
DQA1*01041      DLERKETAWRWPEFSKFGGFDPQGALRNMAVAKHNLNIMIKRYNSTAATNEVPEVTVFSK
DQA1*01042      DLERKETAWRWPEFSKFGGFDPQGALRNMAVAKHNLNIMIKRYNSTAATNEVPEVTVFSK
DQA1*0105       DLERKETAWRWPEFSKFGGFDPQGALRNMAVAKHNLNIMIKRYNSTAATNEVPEVTVFSK
DQA1*0106       DLERKEAAWRWPEFSKFGGFDPQGALRNMAVAKHNLNIMIKRYNSTAATNEVPEVTVFSK
```

```
DQA1*0201       DLERKETVWKLPLFHRLR-FDPQFALTNIAVLKHNLNILIKRSNSTAATNEVPEVTVFSK
DQA1*03011      DLERKETVWQLPLFRRFRRFDPQFALTNIAVLKHNLNIVIKRSNSTAATNEVPEVTVFSK
DQA1*0302       DLERKETVWQLPLFRRFRRFDPQFALTNIAVLKHNLNIVIKRSNSTAATNEVPEVTVFSK
DQA1*0303       DLERKETVWQLPLFRRFRRFDPQFALTNIAVLKHNLNIVIKRSNSTAATNEVPEVTVFSK
DQA1*0401       DLGRKETVWCLPVLRQFR-FDPQFALTNIAVTKHNLNILIKRSNSTAATNEVPEVTVFSK
DQA1*05011      DLGRKETVWCLPVLRQFR-FDPQFALTNIAVLKHNLNSLIKRSNSTAATNEVPEVTVFSK
DQA1*05012      DLGRKETVWCLPVLRQFR-FDPQFALTNIAVLKHNLNSLIKRSNSTAATNEVPEVTVFSK
DQA1*0502       DLGRKETVWCLPVLRQFR-FDRQFALTNIAVLKHNLNSLIKRSNSTAATNEVPEVTVFSK
DQA1*0503       DLGRKETVWCLPVLRQFR-FDPQFALTNIAVLKHNLNSLIKRSNSTAATNEVPEVTVFSK
DQA1*0504       DLGRKETVWCLPVLRQFR-FDPQFALTNIAVLKHNLNSLIKRSNSTAATNEVPEVTVFSK
DQA1*0505       DLGRKETVWCLPVLRQFR-FDPQFALTNIAVLKHNLNSLIKRSNSTAATNEVPEVTVFSK
DQA1*06011      DLGRKETVWCLPVLRQFR-FDPQFALTNIAVTKHNLNILIKRSNSTAATNEVPEVTVFSK
DQA1*06012      DLGRKETVWCLPVLRQFR-FDPQFALTNIAVTKHNLNILIKRSNSTAATNEVPEVTVFSK
                ** :**:.*  * : ::  ** * ** *:** ***** :*** *****************

DQA1*01011      SPVTLGQPNTLICLVDNIFPPVVNITWLSNGQSVTEGVSETSFLSKSDHSFFKISYLTFL
DQA1*01012      SPVTLGQPNTLICLVDNIFPPVVNITWLSNGQSVTEGVSETSFLSKSDHSFFKISYLTFL
DQA1*01021      SPVTLGQPNTLICLVDNIFPPVVNITWLSNGQSVTEGVSETSFLSKSDHSFFKISYLTFL
DQA1*01022      SPVTLGQPNTLICLVDNIFPPVVNITWLSNGQSVTEGVSETSFLSKSDHSFFKISYLTFL
DQA1*0103       SPVTLGQPNTLICLVDNIFPPVVNITWLSNGHAVTEGVSETSFLSKSDHSFFKISYLTFL
DQA1*01041      SPVTLGQPNTLICLVDNIFPPVVNITWLSNGQSVTEGVSETSFLSKSDHSFFKISYLTFL
DQA1*01042      SPVTLGQPNTLICLVDNIFPPVVNITWLSNGQSVTEGVSETSFLSKSDHSFFKISYLTFL
DQA1*0105       SPVTLGQPNTLICLVDNIFPPVVNITWLSNGQSVTEGVSETSFLSKSDHSFFKISYLTFL
DQA1*0106       SPVTLGQPNTLICLVDNIFPPVVNITWLSNGQSVTEGVSETSFLSKSDHSFFKISYLTFL
DQA1*0201       SPVTLGQPNTLICLVDNIFPPVVNITWLSNGHSVTEGVSETSFLSKSDHSFFKISYLTFL
DQA1*03011      SPVTLGQPNTLICLVDNIFPPVVNITWLSNGHSVTEGVSETSFLSKSDHSFFKISYLTFL
DQA1*0302       SPVTLGQPNTLICLVDNIFPPVVNITWLSNGHSVTEGVSETSFLSKSDHSFFKISYLTFL
DQA1*0303       SPVTLGQPNTLICLVDNIFPPVVNITWLSNGHSVTEGVSETSFLSKSDHSFFKISYLTFL
DQA1*0401       SPVTLGQPNTLICLVDNIFPPVVNITWLSNGHSVTEGVSETSFLSKSDHSFFKISYLTFL
DQA1*05011      SPVTLGQPNILICLVDNIFPPVVNITWLSNGHSVTEGVSETSFLSKSDHSFFKISYLTLL
DQA1*05012      SPVTLGQPNILICLVDNIFPPVVNITWLSNGHSVTEGVSETSFLSKSDHSFFKISYLTLL
DQA1*0502       SPVTLGQPNILICLVDNIFPPVVNITWLSNGHSVTEGVSETSFLSKSDHSFFKISYLTLL
DQA1*0503       SPVTLGQPNILICLVDNIFPPVVNITWLSNGHSVTEGVSETSFLSKSDHSFFKISYLTLL
DQA1*0504       SPVTLGQPNILICLVDNIFPPVVNITWLSNGHSVTEGVSETSFLSKSDHSFFKISYLTLL
DQA1*0505       SPVTLGQPNILICLVDNIFPPVVNITWLSNGHSVTEGVSETSFLSKSDHSFFKISYLTLL
DQA1*06011      SPVTLGQPNTLICLVDNIFPPVVNITWLSNGHSVTEGVSETSFLSKSDHSFFKISYLTFL
DQA1*06012      SPVTLGQPNTLICLVDNIFPPVVNITWLSNGHSVTEGVSETSFLSKSDHSFFKISYLTFL
                ********* ********************::*************************:*
```

```
DQA1*01011      PSADEIYDCKVEHWGLDQPLLKHWEPEIPAPMSELTETVVCALGLSVGLVGIVVGTVFII
DQA1*01012      PSADEIYDCKVEHWGLDQPLLKHWEPEIPAPMSELTETVVCALGLSVGLVGIVVGTVFII
DQA1*01021      PSADEIYDCKVEHWGLDQPLLKHWEPEIPAPMSELTETVVCALGLSVGLMGIVVGTVFII
DQA1*01022      PSADEIYDCKVEHWGLDQPLLKHWEPEIPAPMSELTETVVCALGLSVGLMGIVVGTVFII
DQA1*0103       PSADEIYDCKVEHWGLDQPLLKHWEPEIPAPMSELTETVVCALGLSVGLVGIVVGTVFII
DQA1*01041      PSADEIYDCKVEHWGLDQPLLKHWEPEIPAPMSELTETVVCTLGLSVGLVGIVVGTVFII
DQA1*01042      PSADEIYDCKVEHWGLDQPLLKHW----------------------------------
DQA1*0105       PSADEIYDCKVEHWGLDQPLLKHWEPEIPAPMSELTETVVCALGLSVGLVGIVVGTVFII
DQA1*0106       PSADEIYDCKVEHWGLDQPLLKHWEPEIPAPMSELTETVVCALGLSVGLVGIVVGTVFII
DQA1*0201       PSADEIYDCKVEHWGLDEPLLKHWEPEIPAPMSELTETVVCALGLSVGLVGIVVGTVLII
DQA1*03011      PSADEIYDCKVEHWGLDEPLLKHWEPEIPTPMSELTETVVCALGLSVGLVGIVVGTVLII
DQA1*0302       PSDDEIYDCKVEHWGLDEPLLKHWEPEIPTPMSELTETVVCALGLSVGLVGIVVGTVLII
DQA1*0303       PSDDEIYDCKVEHWGLDEPLLKHWEPEIPTPMSELTETVVCALGLSVGLVGIVVGTVLII
DQA1*0401       PSADEIYDCKVEHWGLDEPLLKHWEPEIPAPMSELTETVVCALGLSVGLVGIVVGTVFII
DQA1*05011      PSAEESYDCKVEHWGLDKPLLKHWEPEIPAPMSELTETVVCALGLSVGLVGIVVGTVFII
DQA1*05012      PSAEESYDCKVEHWGLDKPLLKHWEPEIPAPMSELTETVVCALGLSVGLVGIVVGTVFII
DQA1*0502       PSAEESYDCKVEHWGLDKPLLKHWEPEIPAPMSELTETVVCALGLSVGLVGIVVGTVFII
DQA1*0503       PSSEESYDCKVEHWGLDKPLLKHWEPEIPAPMSELTETVVCALGLSVGLVGIVVGTVFII
DQA1*0504       PSSEESYDCKVEHWGLDKPLLKHWEPEIPAPMSELTETVVCALGLSVGLVGIVVGTVFII
DQA1*0505       PSAEESYDCKVEHWGLDKPLLKHWEPEIPAPMSELTETVVCALGLSVGLVGIVVGTVFII
DQA1*06011      PSADEIYDCKVEHWGLDEPLLKHWEPEIPAPMSELTETVVCALGLSVGLVGIVVGTVFII
DQA1*06012      PSADEIYDCKVEHWGLDEPLLKHWEPEIPAPMSELTETVVCALGLSVGLVGIVVGTVFII
                ** :* ***********:******

DQA1*01011      QGLRSVGASRHQGPL
DQA1*01012      QGLRSVGASRHQGPL
DQA1*01021      QGLRSVGASRHQGPL
DQA1*01022      QGLRSVGASRHQGPL
DQA1*0103       QGLRSVGASRHQGPL
DQA1*01041      QGLRSVGASRHQGPL
DQA1*01042      ---------------
DQA1*0105       QGLRSVGASR-----
DQA1*0106       QGLRSVGASR-----
DQA1*0201       RGLRSVGASRHQGPL
DQA1*03011      RGLRSVGASRHQGPL
DQA1*0302       RGLRSVGASRHQGPL
DQA1*0303       RGLRSVGASRHQGPL
```

DQA1*0401	RGLRSVGASRHQGPL
DQA1*05011	RGLRSVGASRHQGPL
DQA1*05012	RGLRSVGASRHQGPL
DQA1*0502	RGLRSVGASRHQGPL
DQA1*0503	RGLRSVGASRHQGPL
DQA1*0504	RGLRSVGASRHQGPL
DQA1*0505	RGLRSVGASRHQGPL
DQA1*06011	RGLRSVGASRHQGPL
DQA1*06012	RGLRSVGASRHQGPL

Appendix II. DQB1 Alleles Alignment

DeCypher Results for: ClustalW Multiple Alignment

```
                                                        1234567890123456789012345678
DQB1*0201    MSWKKALRIPGGLRAATVTLMLSMLSTPVAEGRDSPEDFVYQFKGMCYFTNGTERVRLVS
DQB1*0202    MSWKKALRIPGGLRAATVTLMLSMLSTPVAEGRDSPEDFVYQFKGMCYFTNGTERVRLVS
DQB1*0203    MSWKKALRIPGGLRAATVTLMLSMLSTPVAEGRDSPEDFVYQFKGMCYFTNGTERVRLVS
DQB1*03011   MSWKKALRIPGGLRAATVTLMLSMLSTPVAEGRDSPEDFVYQFKAMCYFTNGTERVRYVT
DQB1*03012   MSWKKALRIPGGLRAATVTLMLSMLSTPVAEGRDSPEDFVYQFKAMCYFTNGTERVRYVT
DQB1*0302    MSWKKALRIPGGLRVATVTLMLAMLSTPVAEGRDSPEDFVYQFKGMCYFTNGTERVRLVT
DQB1*03032   MSWKKALRIPGGLRVATVTLMLAMLSTPVAEGRDSPEDFVYQFKGMCYFTNGTERVRLVT
DQB1*03033   MSWKKALRIPGGLRVATVTLMLAMLSTPVAEGRDSPEDFVYQFKGMCYFTNGTERVRLVT
DQB1*0304    MSWKKALRIPGGLRVATVTLMLAMLSTPVAEGRDSPEDFVYQFKAMCYFTNGTERVRYVT
DQB1*03051   MSWKKALRIPGGLRVATVTLMLAMLSTPVAEGRDSPEDFVYQFKGMCYFTNGTERVRGVT
DQB1*03052   -------------------------------------DFVYQFKGMCYFTNGTERVRGVT
DQB1*0306    -------------------------------------DFVYQFKGMCYFTNGTERVRLVT
DQB1*0307    ------------------------------------EDFVYQFKGMCYFTNGTERVRLVT
DQB1*0308    -------------------------------------DFVYQFKGMCYFTNGTERVRLVT
DQB1*0309    --------------------------------RDSPEDFVYQFKAMCYFTNGTERVRYVT
DQB1*0310    --------------------------------RDSPEDFVYQFKAMCYFTNGTERVRYVT
DQB1*0401    --------------------------------RDSPEDFVFQFKGMCYFTNGTELVRGVT
DQB1*0402    MSWKKALRIPGGLRVATVTLMLAMLSTPVAEGRDSPEDFVFQFKGMCYFTNGTERVRGVT
DQB1*05011   MSWKKSLRIPGDLRVATVTLMLAILSSSLAEGRDSPEDFVYQFKGLCYFTNGTERVRGVT
DQB1*05012   MSWKKSLRIPGDLRVATVTLMLAILSSSLAEGRDSPEDFVYQFKGLCYFTNGTERVRGVT
DQB1*0502    MSWKKSLRIPGDLRVATVTLMLAILSSSLAEGRDSPEDFVYQFKGLCYFTNGTERVRGVT
DQB1*05031   --------------------------------RDSPEDFVYQFKGLCYFTNGTERVRGVT
DQB1*05032   --------------------------------RDSPEDFVYQFKGLCYFTNGTERVRGVT
DQB1*0504    ----------------------------------------YQFKGLCYFTNGTERVRGVT
DQB1*06011   --------------------------------RDPPEDFVLQFKAMCYFTNGTERVRYVT
DQB1*06012   --------------------------------RDPPEDFVLQFKAMCYFTNGTERVRYVT
DQB1*06013   --------------------------------RDPPEDFVLQFKAMCYFTNGTERVRYVT
DQB1*0602    MSWKKALRIPGDLRVATVTLMLAMLSSLLAEGRDSPEDFVFQFKGMCYFTNGTERVRLVT
DQB1*0603    --------------------------------RDSPEDFVYQFKGMCYFTNGTERVRLVT
DQB1*06041   --------------------------------RDSPEDFVYQFKGMCYFTNGTERVRLVT
DQB1*06042   ----------------------------------SPEDFVYQFKGMCYFTNGTERVRLVT
DQB1*06051   --------------------------------RDSPEDFVYQFKGLCYFTNGTERVRLVT
```

```
DQB1*06052     ----------------------------------------------------TERVRLVT
DQB1*0606      ----------------------------------------------------TERVRLVT
DQB1*0607      ---------------------------------------YQFKGMCYFTNGTERVRLVT
DQB1*0608      ---------------------------------------YQFKGMCYFTNGTERVRLVT
DQB1*0609      --------------------------------RDSPEDFVYQFKGMCYFTNGTERVRLVT
DQB1*0610      ------------------------------------DFVFQFKGMCYFTNGTERVRLVT
DQB1*06111     ---------------------------------------YQFKGMCYFTNGTERVRLVT
DQB1*06112     ------------------------------------DFVYQFKGMCYFTNGTERVRLVT
DQB1*0612      MSWKKALRIPGDLRVATVTLMLAMLSSLLAEGRDSPEDFVYQFKGMCYFTNGTERVRLVT
DQB1*0613      MSWKKALRIPGDLRVATVTLMLAMLSSLLAEGRDSPEDFVFQFKGMCYFTNGTERVRLVT
DQB1*0614      MSWKKALRIPGDLRVATVTLMLAMLSSLLAEGRDSPEDFVFQFKGMCYFTNGTERVRLVT
DQB1*0615      MSWKKALRIPGDLRVATVTLMLAMLSSLLAEGRDSPEDFVFQFKGMCYFTNGTERVRLVT
DQB1*0616      MSWKKALRIPGDLRVATVTLMLAMLSSLLAEGRDSPEDFVFQFKGMCYFTNGTERVRLVT
DQB1*0617      MSWKKALRIPGDLRVATVTLMLAMLSSLLAEGRDSPEDFVYQFKGMCYFTNGTERVRLVT
DQB1*0618      MSWKKALRIPGDLRVATVTLMLAMLSSLLAEGRDSPEDFVYQFKGMCYFTNGTERVRLVT
DQB1*0619      MSWKKALRIPGDLRVATVTLMLAMLSSLLAEGRDSPEDFVFQFKGMCYFTNGTERVRLVT
                                                                ** ** *:

DQB1*0201      RSIYNREEIVRFDSDVGEFRAVTLLGLPAAEYWNSQKDILERKRAAVDRVCRHNYQLELR
DQB1*0202      RSIYNREEIVRFDSDVGEFRAVTLLGLPAAEYWNSQKDILERKRAAVDRVCRHNYQLELR
DQB1*0203      RSIYNREEIVRFDSDVGEFRAVTLLGLPDAEYWNSQKDILERKRAAVDRVCRHNYQLELR
DQB1*03011     RYIYNREEYARFDSDVEVYRAVTPLGPPDAEYWNSQKEVLERTRAELDTVCRHNYQLELR
DQB1*03012     RYIYNREEYARFDSDVEVYRAVTPLGPPDAEYWNSQKEVLERTRAELDTVCRHNYQLELR
DQB1*0302      RYIYNREEYARFDSDVGVYRAVTPLGPPAAEYWNSQKEVLERTRAELDTVCRHNYQLELR
DQB1*03032     RYIYNREEYARFDSDVGVYRAVTPLGPPDAEYWNSQKEVLERTRAELDTVCRHNYQLELR
DQB1*03033     RYIYNREEYARFDSDVGVYRAVTPLGPPDAEYWNSQKEVLERTRAELDTVCRHNYQLELR
DQB1*0304      RYIYNREEYARFDSDVEVYRAVTPLGPPAAEYWNSQKEVLERTRAELDTVCRHNYQLELR
DQB1*03051     RYIYNREEYARFDSDVGVYRAVTPLGPPAAEYWNSQKEVLERTRAELDTVCRHNYQLELR
DQB1*03052     RYIYNREEYARFDSDVGVYRAVTPLGPPAAEYWNSQKEVLERTRAELDTVCRHNYQLELR
DQB1*0306      RYIYNREEYARFDSDVGVYRAVTPLGPPDAEYWNSQKDILEEDRASVDTVCRHNYQLELR
DQB1*0307      RYIYNREEYARFDSDVGVYRVVTPLGPPAAEYWNSQKEVLERTRAELDTVCRHNYQLELR
DQB1*0308      RYIYNREEYARFDSDVGVYRAVTPLGPPAAEYWNSQKEVLEGTRAELDTVCRHNYQLELR
DQB1*0309      RYIYNREEYARFDSDVEVYRAVTPLGPPDAEYWNSQKEVLERTRAELDTVCRHNYQLELR
DQB1*0310      RYIYNREEYARFDSDVGVYRAVTPLGPPDAEYWNSQKEVLERTRAELDTVCRHNYQLELR
DQB1*0401      RYIYNREEYARFDSDVGVYRAVTPLGRLDAEYWNSQKDILEEDRASVDTVCRHNYQLELR
DQB1*0402      RYIYNREEYARFDSDVGVYRAVTPLGRLDAEYWNSQKDILEEDRASVDTVCRHNYQLELR
DQB1*05011     RHIYNREEYVRFDSDVGVYRAVTPQGRPVAEYWNSQKEVLEGARASVDRVCRHNYEVAYR
DQB1*05012     RHIYNREEYVRFDSDVGVYRAVTPQGRPVAEYWNSQKEVLEGARASVDRVCRHNYEVAYR
```

```
DQB1*0502       RHIYNREEYVRFDSDVGVYRAVTPQGRPSAEYWNSQKEVLEGARASVDRVCRHNYEVAYR
DQB1*05031      RHIYNREEYVRFDSDVGVYRAVTPQGRPDAEYWNSQKEVLEGARASVDRVCRHNYEVAYR
DQB1*05032      RHIYNREEYVRFDSDVGVYRAVTPQGRPDAEYWNSQKEVLEGARASVDRVCRHNYEVAYR
DQB1*0504       RYIYNREEYVRFDSDVGVYRAVTPQGRPSAEYWNSQKDILEEDRASVDRVCRHNYEVAYR
DQB1*06011      RYIYNREEDVRFDSDVGVYRAVTPQGRPDAEYWNSQKDILERTRAELDTVCRHNYEVAFR
DQB1*06012      RYIYNREEDVRFDSDVGVYRAVTPQGRPDAEYWNSQKDILERTRAELDTVCRHNYEVAFR
DQB1*06013      RYIYNREEDVRFDSDVGVYRAVTPQGRPDAEYWNSQKDILERTRAELDTVCRHNYEVAFR
DQB1*0602       RYIYNREEYARFDSDVGVYRAVTPQGRPDAEYWNSQKEVLEGTRAELDTVCRHNYEVAFR
DQB1*0603       RHIYNREEYARFDSDVGVYRAVTPQGRPDAEYWNSQKEVLEGTRAELDTVCRHNYEVAFR
DQB1*06041      RHIYNREEYARFDSDVGVYRAVTPQGRPVAEYWNSQKEVLERTRAELDTVCRHNYEVGYR
DQB1*06042      RHIYNREEYARFDSDVGVYRAVTPQGRPVAEYWNSQKEVLERTRAELDTVCRHNYEVGYR
DQB1*06051      RYIYNREEYARFDSDVGVYRAVTPQGRPVAEYWNSQKEVLERTRAELDTVCRHNYEVGYR
DQB1*06052      RYIYNREEYARFDSDVGVYRAVTPQGRPVAEYWNSQKEVLERTRAELDTVCRHNYEVGYR
DQB1*0606       RYIYNREEYARFDSDVGVYRAVTPQGRPVAEYWNSQKEVLERTRAAVDRVCRHNYEVGYR
DQB1*0607       RHIYNREEYARFDSDVGVYRAVTPQGRPDAEYWNSQKEVLERTRAELDTVCRHNYEVGYR
DQB1*0608       RHIYNREEYARFDSDVGVYRAVTPQGRPVAEYWNSQKEVLEGTRAELDTVCRHNYEVAFR
DQB1*0609       RYIYNREEYARFDSDVGVYRAVTPQGRPVAEYWNSQKEVLERTRAELDTVCRHNYEVGYR
DQB1*0610       RYIYNREEYARFDSDVGVYRAVTPQGRPSAEYWNSQKEVLEGTRAELDTVCRHNYEVAFR
DQB1*06111      RYIYNREEYARFDSDVGVYRAVTPQGRPDAEYWNSQKEVLEGTRAELDTVCRHNYEVAFR
DQB1*06112      RYIYNREEYARFDSDVGVYRAVTPQGRPDAEYWNSQKEVLEGTRAELDTVCRHNYEVAFR
DQB1*0612       RYIYNREEYARFDSDVGVYRAVTPQGRPVAEYWNSQKEVLEGTRAELDTVCRHNYEVGYR
DQB1*0613       RYIYNREEYARFDSDVGVYRAVTPQGRPVAEYWNSQKEVLEGTRAELDTVCRHNYEVAFR
DQB1*0614       RHIYNREEYARFDSDVGVYRAVTPQGRPDAEYWNSQKEVLEGTRAELDTVCRHNYEVAFR
DQB1*0615       RYIYNREEYARFDSDVGVYRAVTPQGRPDAEYWNSQKEVLERTRAELDTVCRHNYEVGYR
DQB1*0616       RYIYNREEYARFDSDVGVYRAVTPQGRPDAENWNSQKEVLEGTRAELDTVCRHNYEVAFR
DQB1*0617       RHIYNREEYARFDSDVGVYRAVTPQGRPVAEYWNSQKEVLEGARAELDTVCRHNYEVGYR
DQB1*0618       RYIYNREEYARFDSDVGVYRAVTPQGRPVAEYWNSQKEVLERTRAELDTVCRHNYEVAFR
DQB1*0619       RYIYNREEYARFDSDVGVYRAVTPLGRPDAEYWNSQKEVLEGTRAELDTVCRHNYEVAFR
                * ****** .****** :*.**  *   ** *****::**  ** :* ******::  *

DQB1*0201       TTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNDQEETAGVVSTPLI
DQB1*0202       TTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNGQEETAGVVSTPLI
DQB1*0203       TTLQRR--------PSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNG------------
DQB1*03011      TTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNDQEETTGVVSTPLI
DQB1*03012      TTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNDQEETTGVVSTPLI
DQB1*0302       TTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNDQEETTGVVSTPLI
DQB1*03032      TTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNDQEETTGVVSTPLI
DQB1*03033      TTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNDQEETTGVVSTPLI
```

DQB1*0304	TTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNDQEETTGVVSTPLI
DQB1*03051	TTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNDQEETTGVVSTPLI
DQB1*03052	TTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNDQEETTGVVSTPLI
DQB1*0306	TTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNDQEETTGVVSTPLI
DQB1*0307	TTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNDQEETTGVVSTPLI
DQB1*0308	TTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNDQEETTGVVSTPLI
DQB1*0309	TTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNDQEETTGVVSTPLI
DQB1*0310	TTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNDQEETTGVVSTPLI
DQB1*0401	TTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNDQEETTGVVSTPLI
DQB1*0402	TTLQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPAQIKVRWFRNDQEETTGVVSTPLI
DQB1*05011	GILQRRVEPTVTISPSRTEALNHHNLLICSVTDFYPSQIKVRWFRNDQEETAGVVSTPLI
DQB1*05012	GILQRRVEPTVTISPSRTEALNHHNLLICSVTDFYPSQIKVRWFRNDQEETAGVVSTPLI
DQB1*0502	GILQRRVEPTVTISPSRTEALNHHNLLICSVTDFYPSHIKVRWFRNDQEETAGVVSTPLI
DQB1*05031	GILQRRVEPTVTISPSRTEALNHHNLLICSVTDFYPSQIKVRWFRNDQEETAGVVSTPLI
DQB1*05032	GILQRRVEPTVTISPSRTEALNHHNLLICSVTDFYPSQIKVRWFRNDQEETAGVVSTPLI
DQB1*0504	GILQRRVEPTVTISPSRTEALNHHNLLICSVTDFYPSQIKVRWFRNDQEETAGVVSTPLI
DQB1*06011	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVRWFRNDQEETAGVVSTPLI
DQB1*06012	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVRWFRNDQEETAGVVSTPLI
DQB1*06013	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVRWFRNDQEETAGVVSTPLI
DQB1*0602	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVRWFRNDQEETAGVVSTPLI
DQB1*0603	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVRWFRNDQEETAGVVSTPLI
DQB1*06041	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI
DQB1*06042	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI
DQB1*06051	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI
DQB1*06052	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI
DQB1*0606	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI
DQB1*0607	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI
DQB1*0608	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI
DQB1*0609	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI
DQB1*0610	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI
DQB1*06111	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI
DQB1*06112	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI
DQB1*0612	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI
DQB1*0613	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI
DQB1*0614	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI
DQB1*0615	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI
DQB1*0616	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI
DQB1*0617	GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI

```
DQB1*0618      GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI
DQB1*0619      GILQRRVEPTVTISPSRTEALNHHNLLVCSVTDFYPGQIKVQWFRNDQEETAGVVSTPLI
                 ****        ************:********.:***:****.

DQB1*0201      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGIGGFVL
DQB1*0202      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGIGGFVL
DQB1*0203      ---DWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGIGGFVL
DQB1*03011     RNGDWTFQILVMLEMTPQHGDVYTCHVEHPSLQNPITVEWRAQSESAQSKMLSGIGGFVL
DQB1*03012     RNGDWTFQILVMLEMTPQHGDVYTCHVEHPSLQNPITVEWRAQSESAQSKMLSGIGGFVL
DQB1*0302      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQNPIIVEWRAQSESAQSKMLSGIGGFVL
DQB1*03032     RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQNPIIVEWRAQSESAQSKMLSGIGGFVL
DQB1*03033     RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQNPIIVEWRAQSESAQSKMLSGIGGFVL
DQB1*0304      RNGDWTFQILVMLEMTPQHGDVYTCHVEHPSLQNPITVEWRAQSESAQSKM---------
DQB1*03051     RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQNPIIVEWRAQSESAQSKM---------
DQB1*03052     RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQNPIIVEWRAQSESAQSKM---------
DQB1*0306      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQNPIIVEWRAQSESAQSKM---------
DQB1*0307      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQNPIIVEWRAQSESAQSKM---------
DQB1*0308      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQNPIIVEWRAQSESAQSKM---------
DQB1*0309      RNGDWTFQILVMLEMTPQHA-VYTCHVEHPSLQNPITVEWRAQSESAQSKMLSGIGGFVL
DQB1*0310      RNGDWTFQILVMLEMTPQHGDVYTCHVEHPSLQNPITVEWRAQSESAQSKMLSGIGGFVL
DQB1*0401      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQNPIIVEWRAQSESAQSKMLSGIGGFVL
DQB1*0402      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQNPIIVEWRAQSESAQSKMLSGIGGFVL
DQB1*05011     RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*05012     RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*0502      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*05031     RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*05032     RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*0504      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*06011     RNGDWTFQILVMLEMTPQHGDVYTCHVEHPSLQSPITVEWRAQSESAQNKMLSGIGGFVL
DQB1*06012     RNGDWTFQILVMLEMTPQHGDVYTCHVEHPSLQSPITVEWRAQSESAQNKMLSGIGGFVL
DQB1*06013     RNGDWTFQILVMLEMTPQHGDVYTCHVEHPSLQSPITVEWRAQSESAQNKMLSGIGGFVL
DQB1*0602      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*0603      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*06041     RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*06042     RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*06051     RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*06052     RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*0606      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
```

```
DQB1*0607      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*0608      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*0609      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*0610      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*06111     RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*06112     RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*0612      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*0613      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*0614      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*0615      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*0616      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*0617      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*0618      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
DQB1*0619      RNGDWTFQILVMLEMTPQRGDVYTCHVEHPSLQSPITVEWRAQSESAQSKMLSGVGGFVL
                  ***************:. ************.** ************.**

DQB1*0201      GLIFLGLGLIIHHRSQKGLLH--------
DQB1*0202      GLIFLGLGLIIHHRSQKGLLH--------
DQB1*0203      GLIFLGLGLIIHHRSQKGLLH--------
DQB1*03011     GLIFLGLGLIIHHRSQKGLLH--------
DQB1*03012     GLIFLGLGLIIHHRSQKGLLH--------
DQB1*0302      GLIFLGLGLIIHHRSQKGLLH--------
DQB1*03032     GLIFLGLGLIIHHRSQKGLLH--------
DQB1*03033     GLIFLGLGLIIHHRSQKGLLH--------
DQB1*0304      ----------------------------
DQB1*03051     ----------------------------
DQB1*03052     ----------------------------
DQB1*0306      ----------------------------
DQB1*0307      ----------------------------
DQB1*0308      ----------------------------
DQB1*0309      GLIFLGLGLIIHHRSQKGLLH--------
DQB1*0310      GLIFLGLGLIIHHRSQKGLLH--------
DQB1*0401      GLIFLGLGLIIHHRSQKGLLH--------
DQB1*0402      GLIFLGLGLIIHHRSQKGLLH--------
DQB1*05011     GLIFLGLGLIIRQRSRKGLLH--------
DQB1*05012     GLIFLGLGLIIRQRSRKGLLH--------
DQB1*0502      GLIFLGLGLIIRQRSRKGLLH--------
DQB1*05031     GLIFLGLGLIIRQRSRKGPQGPPPAGLLH
```

DQB1*05032 GLIFLGLGLIIRQRSRKGPQGPPPAGLLH
DQB1*0504 GLIFLGLGLIIRQRSRKGPQGPPPAGLLH
DQB1*06011 GLIFLGLGLIIRQRSQKGPQGPPPAGLLH
DQB1*06012 GLIFLGLGLIIRQRSQKGPQGPPPAGLLH
DQB1*06013 GLIFLGLGLIIRQRSQK----------
DQB1*0602 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*0603 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*06041 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*06042 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*06051 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*06052 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*0606 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*0607 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*0608 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*0609 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*0610 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*06111 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*06112 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*0612 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*0613 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*0614 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*0615 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*0616 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*0617 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*0618 GLIFLGLGLIIRQRSQKGLLH--------
DQB1*0619 GLIFLGLGLIIRQRSQKGLLH--------