Introduction

The major histocompatibility complex (MHC) plays a fundamental role in cellular immune responses. The MHC class I molecules are responsible for binding peptides derived from intracellular proteins, forming a peptide-MHC composite ligand on the cell surface. These ligands can be sampled by circulating T lymphocytes (1). Once the T cell becomes activated due to binding of the ligand, the cell differentiates into cytotoxic T lymphocytes (CTL) which then destroy virally-infected cells that are presenting the same peptide on their cell surface (2).

The MHC class I molecules consists of two separate chains. The alpha (α) chain and the β-2-microglobulin chain (figure 1). The two chains and peptide fragment form a cell surface complex to which the lymphocyte receptors bind. The crystallized MHC molecule in figure 1 contains only the extracellular portion of the α chain as well as the β-2-microglobulin chain and a peptide fragment bound by the molecule.

Because the α chain contains the portion of the molecule responsible for binding and presenting the peptide fragments, it will be the focus of the characterization of the molecule. The α chain consists of five distinct domains; three extracellular domains designated α1, α2, and α3, a transmembrane domain, and a cytoplasmic domain.

The α1 and α2 domains contain the positions that contribute to the binding pockets for the peptide fragments and the lymphocyte receptors. The binding groove is broken down into six distinct pockets based on chemical and physical characteristics. The most important pockets for peptide fragment binding are the B and the F pockets. The amino acids that contribute to these pockets are responsible for anchoring the fragment into the groove. The B pocket consists of eleven amino acids, the majority of which are located in the α1 domain(figure 2).

The F pocket is responsible for anchoring the carboxyl end of the peptide fragment. This pocket is made up of ten amino acids, predominantly from the α2 domain(figure 3).

Figure 1. Crystallized structure of HLA-A locus molecule.

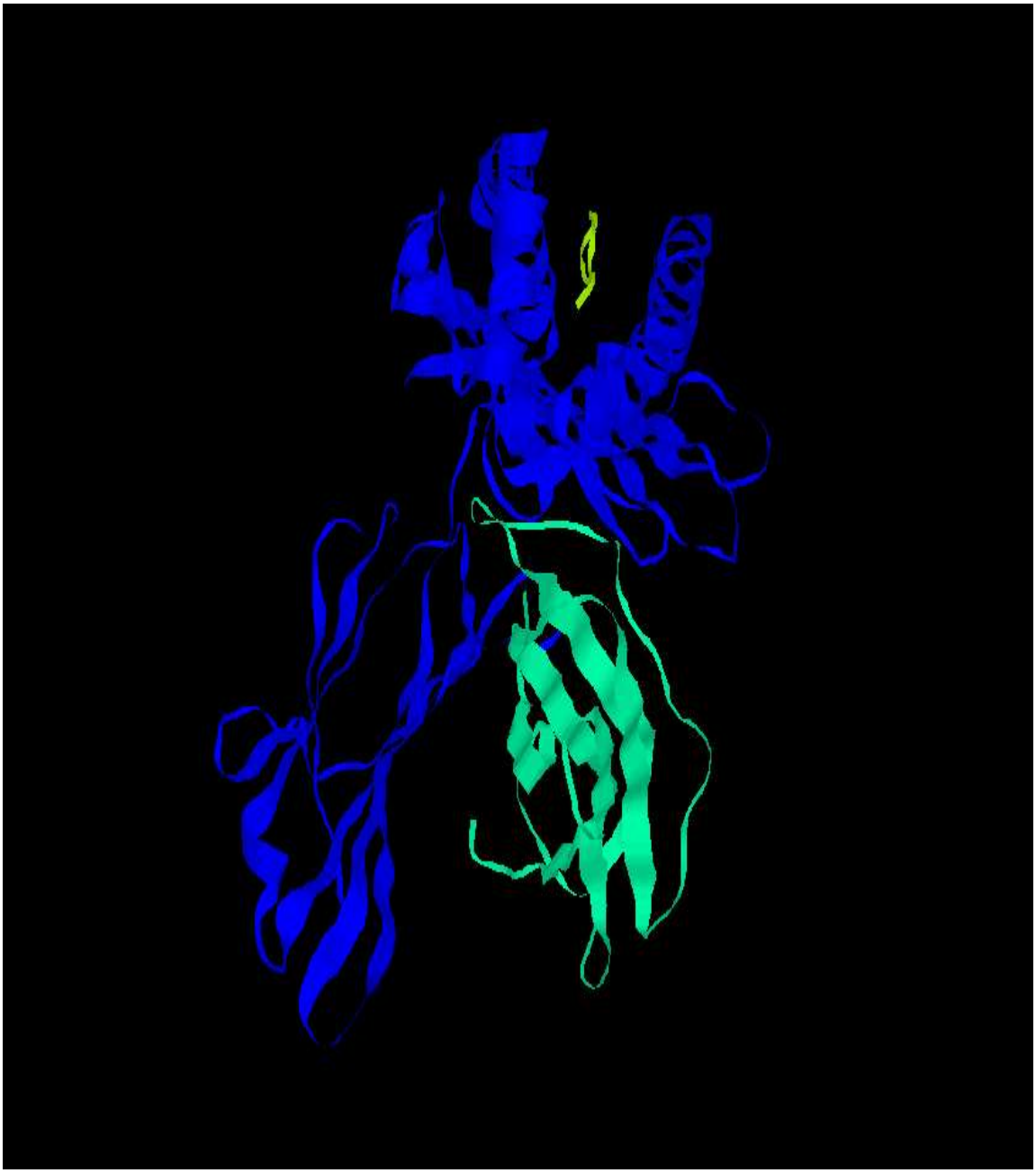


Figure 1: A three dimensional ribbon image of the crystallized structure derived from an MHC class I molecule from the HLA-A locus(PDB: 1a1m). The α chain is shaded blue, the peptide fragment bound to the groove is green, and the β-2-microglobulin chain is turquoise.

Figure 2. Crystallized structure of the binding groove in an HLA-A molecule.

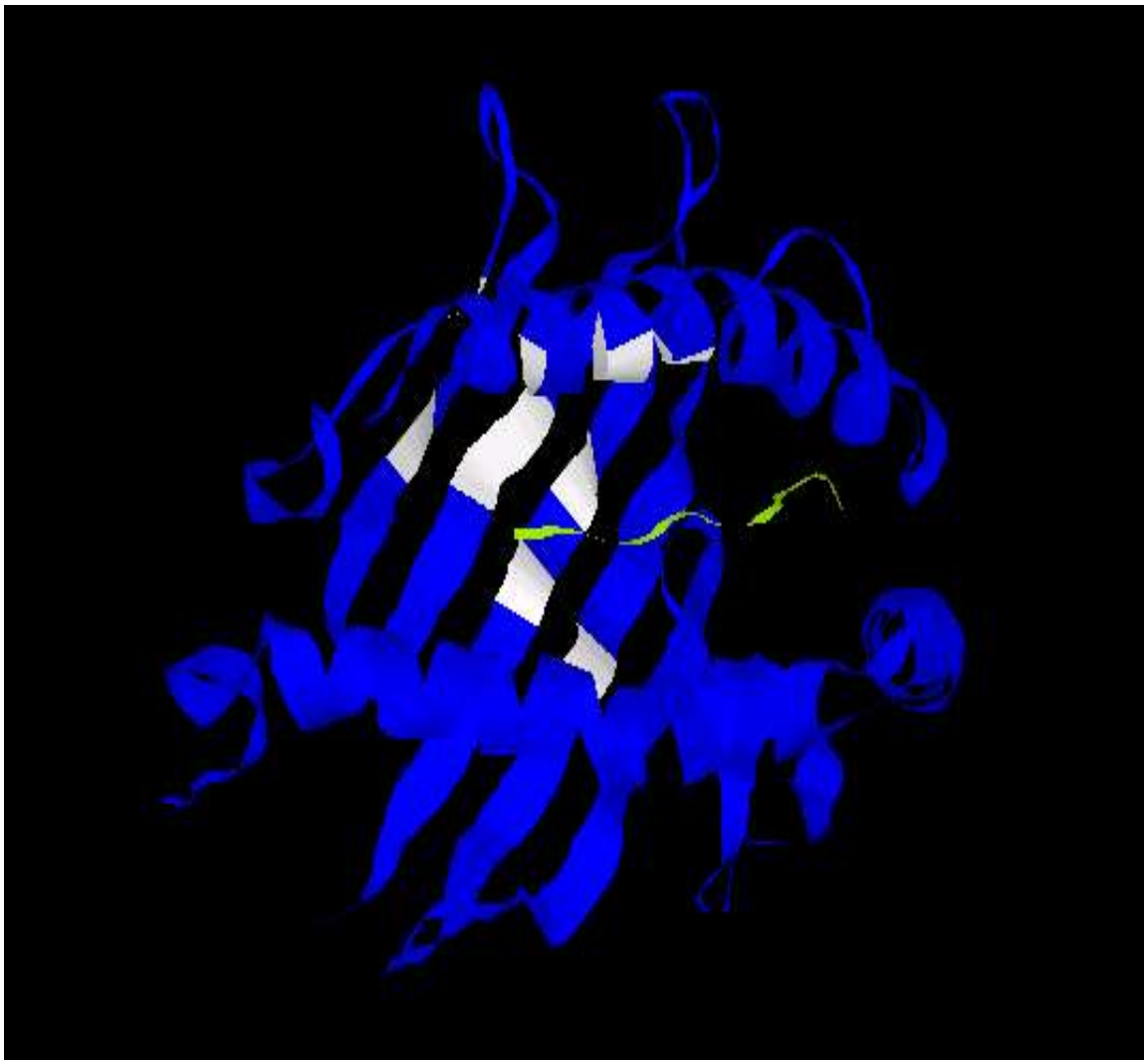


figure 2: The α1 and α2 domains of a class I MHC molecule (PDB: 1a1m). The 11 amino acid positions contributing to the B pocket are colored in white while the remainder of the amino acids are blue. The peptide fragment bound by the molecule is indicated by the green ribbon structure.

Figure 3. Crystallized structure of binding groove in an HLA-A molecule.

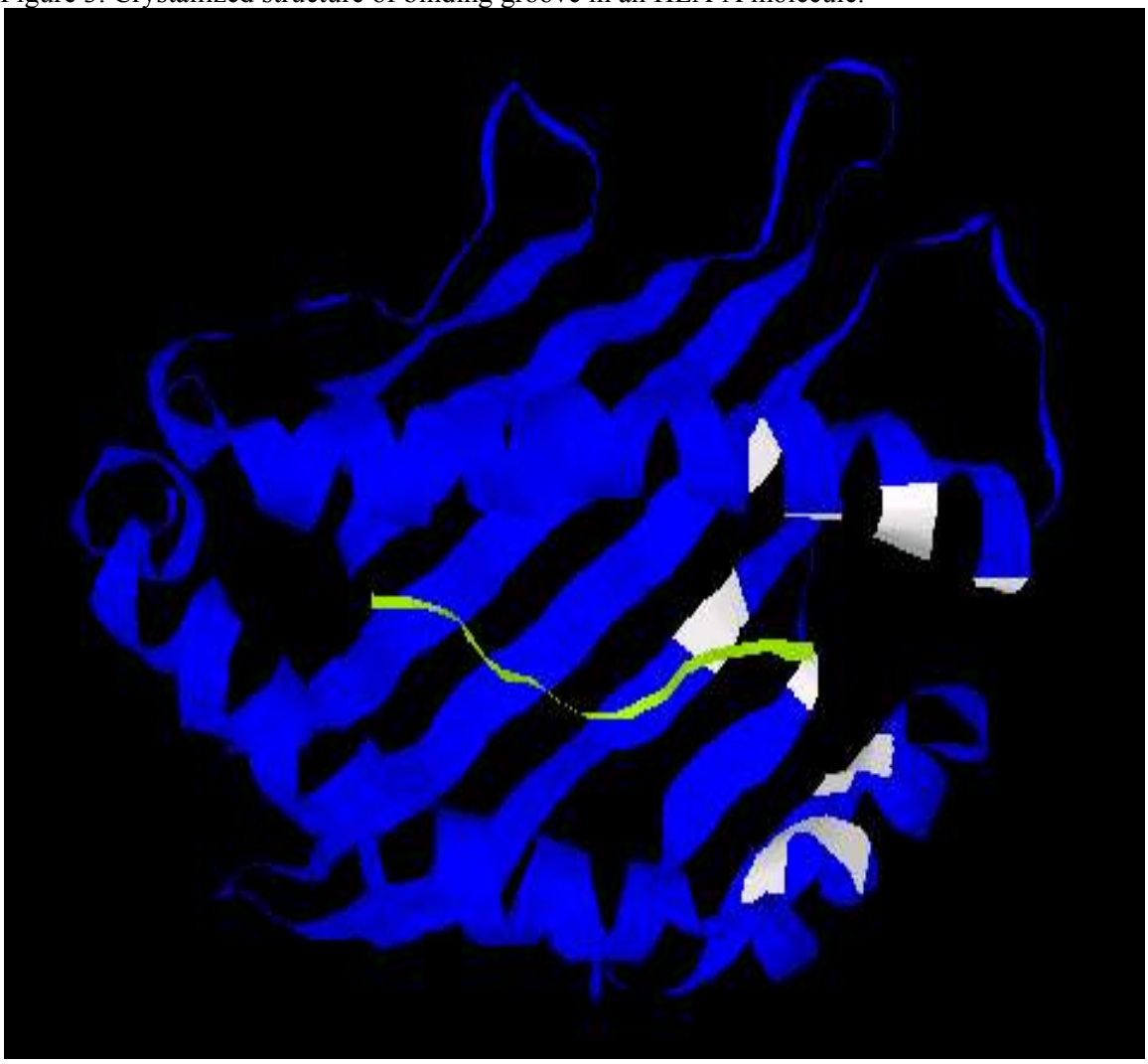


figure 3. The α1 and α2 domains of a class I MHC molecule (PDB: 1a1m). The amino acids contributing to the F pocket of the binding groove are colored white while the remainder of the α1 and α2 domains are colored blue. The peptide fragment anchored by the binding pocket is represented by the green ribbon structure.

Unlike most genes, MHC molecules exhibit positive selection for nucleotide substitutions that result in a change in the amino acid encoded by the codon(10). The majority

of the point mutations occur at the codons for the peptide binding residues; however, the entire gene is often (in evolutionary time) subject to gene duplication events as well as recombination with other similar MHC genes. Such expansion is most apparent in the haplotypes of the Class I MHC for *M. mulatta* characterized by the Clinical Research Division at the Fred Hutchinson Cancer Research Center (FHCRC). Figure 4 is a modified version of the haplotype map generated by the FHCRC. The expansion of the Class I MHC is predominantly in the B-locus region, however there has been a duplication of the A-locus region in *M. mulatta* as well.

These mechanisms are responsible for the enormous amount of diversity present within the MHC loci. In humans, although there are three loci that bind peptides and present them to $CD8^+$ T lymphocytes, most of the diversification is due to the polymorphism of the loci. This expansion is most apparent in humans where over 560 B locus alleles have been described. Furthermore, in certain species such as *Macaca mulatta*, there has been repeated duplication of the B locus resulting in as many as 19 paralogous genes (figure 4). Also in the *M. mulatta*, the A locus has undergone duplication to form two distinct A-like genes, each producing functional molecules that are expressed by the individual in *M. mulatta* (16). The expansion of the B locus however is on a much larger scale than the duplication observed in the A locus. While not all of the 19 B-like genes are functional, representative alleles for several have been sequenced and characterized (3). Other loci such as the E and the F, which are considered to be non-classical MHC molecules, are much more conserved than the A and B loci and present a restricted set of peptides to the TCR (3).

The *M. mulatta* species is one of the most widely used animal models being used for the study of infectious diseases (3). Because *M. mulatta* is such a common animal model for viral and other infectious disease studies, the MHC has been very well characterized (6). Unfortunately it is becoming increasingly difficult to obtain *M. mulatta* for research use. To overcome these difficulties, attempts are being made to establish the *Macaca fascicularis* as a new model for clinical

Figure 4. Haplotype map for Human and Macaque MHC class I region.

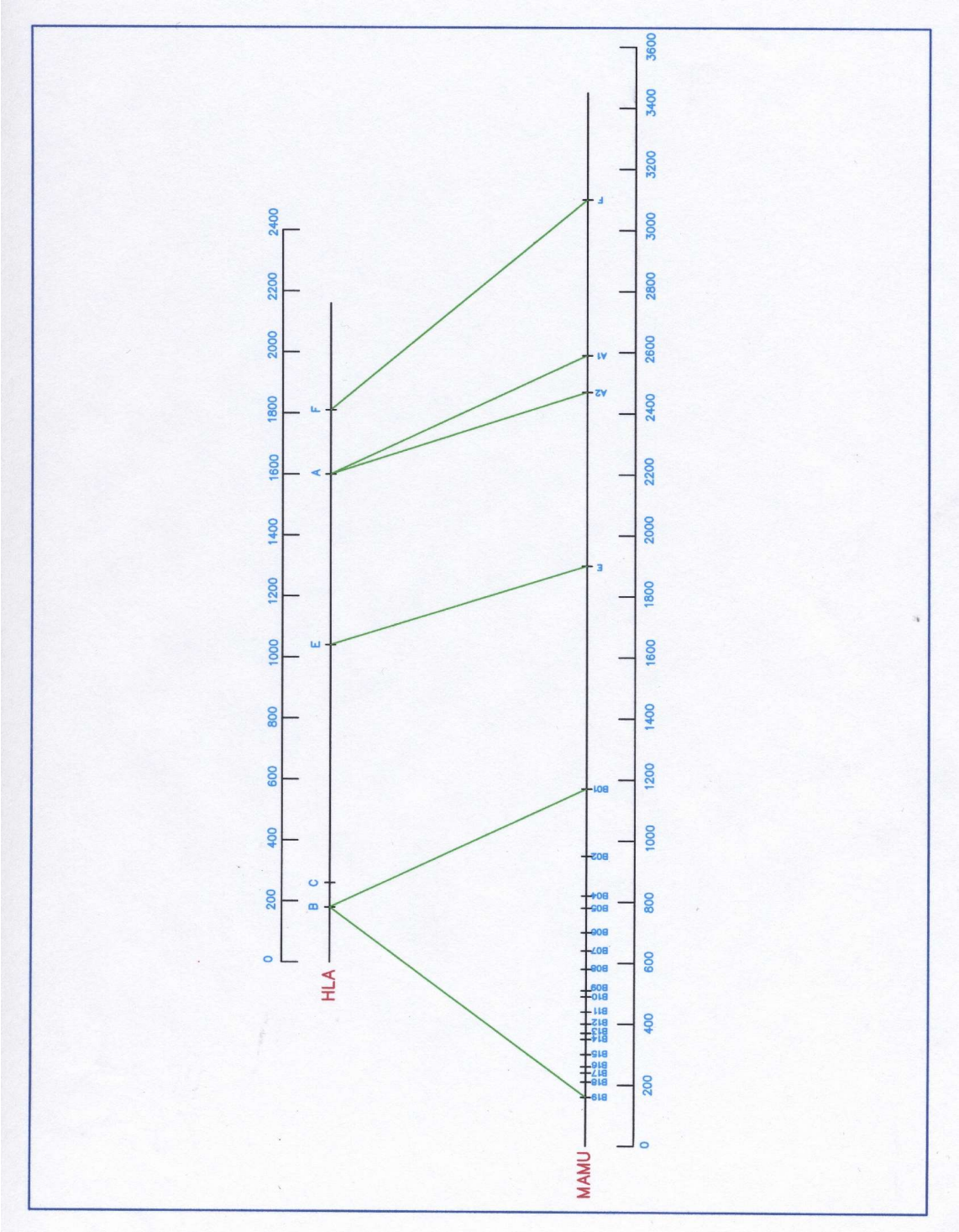


figure 4. Graphic representation of the MHC class I region for Human (HLA) and *M. mulatta (*Mamu). The scale indicates the length of the regions in kilobases. The green lines serve as an aid to indicate comparable regions between the haplotypes.

studies involving Simian Immunodeficiency Virus (SIV) as well as other pathogens such as *Bacillus anthracis*. In order for the *M. fascicularis* to become a viable animal model for such studies, the MHC genes of the *M. fascicularis* must be characterized.

Materials and Methods

Animals

The RNAs from *M. fascicularis* were received from Anita Trichel, DVM, at the University of Pittsburgh. The RNAs were extracted from peripheral blood lymphocytes with the exception of animal #13402, which came from tissue. All animals were healthy at the time of extraction. The sex and age of the animals are not known.

Table 1. Relationship between Animal ID numbers and clones obtained from that individual

Animal #	***Clone Series***	***locus***	***length(nt)***
13102	9-1 to 9-16	Mafa-B	1060
13102	13-1 to 13-6	Mafa-E	1048
13202	10-1 to 10-16	Mafa-E	1048
13202	11-1 to 11-16	Mafa-F	1018
13402	16-1 to 16-16	Mafa-B	1060
19102	15-1 to 15-16	Mafa-B	1060

Table 1. Lists the Animal ID# from the RNA samples, the clone series obtained from each sample, the putative locus, and the length of the clone nucleotide sequence

cDNA clones

The RNAs were converted to cDNAs using the Invitrogen superscript kit (Invitrogen, Carlsbad CA). The cDNA synthesis reaction mixture contained 5 ug RNA, 1 ul 10mMdNTP, 1 ul oligodT, and nanopure water (ddH_2O) to bring the mixture to a final volume of 10 ul. The reaction mixture was then incubated at 65^{o}C for 5 minutes, then placed on ice for a minimum of

one minute. After being on ice for the appropriate amount of time, 9 ul of cocktail consisting of 10x RT buffer, 25 mM $MgCl_2$, 0.1 M DTT, and Rnase Out was added to each sample. After the cocktail was added the samples were incubated at 92°C for 2 minutes. After the incubation, 1 ul of reverse transcriptase was added to each tube. The samples were then mixed and incubated at 42°C for 50 minutes. The reactions were terminated at 70°C for 15 minutes. After pulse spinning the samples, 1 ul of Rnase H was added to each tube and samples were incubated at 37°C for 20 minutes. The cDNAs were then refrigerated until further use.

PCR Experiments

The PCR experiments were set up using 2 ul of cDNA per amplification, 1 ul each of forward and reverse primer at a concentration of 10 mM, 1 ul of 10 mM dNTP mix, 5 ul of 25 mM $MgCl_2$, 5 ul of 10x Gold Buffer, 34.8 ul of nanopure H_2O, and 0.2 ul of Amplitaq gold (Invitrogen, Carlsbad CA). The PCR amplification was carried out with an initial ten minutes at 95°C; followed by 30 cycles consisting of 30 seconds at 94°C, 30 seconds at 55°C, 2 minutes 15 seconds at 72°C; and the final stage at 72°C for seven minutes. After the cycles had been completed, the samples were stored at 4°C.

The PCR products were gel purified using a 1.8% agarose/TAE gel. Electrophoresis was conducted at 90volts for 1.5 hours. The band was excised from the gel and the DNA was eluted using geneclean spin filters (Qbiogene, Irvine CA). The purified PCR DNA products were ligated into pCR4-TOPO Vector (Invitrogen, Carlsbad CA) and transformed into *E. coli*TOP10. Transformants were plated on LB plates with kanamycin.

Sequencing

Small scale *Escherichia coli* cultures were established using *E. coli* cells that had been transformed with the PCR4-TOPO vector containing the cDNA insert. The cultures were grown in 2 ml LB broth containing kanamycin at a concentration of 10 μg/ml. After approximately 20

hours of incubation at 37°C, the small scale cultures were used to inoculate 100 ml of LB broth containing kanamycin at a concentration of 10 ug/ml. The large scale cultures were incubated at 37°C for 24 hours to bring the *E. coli* cells to a saturated concentration. After incubation, the *E. coli* cultures were stored at 4°C until the plasmid purification could be performed.

Plasmids were isolated and purified from the *E. coli* cells using the Qiagen Maxi kit protocol and the Qiagen HiSpeed Maxi kit protocol (Qiagen, Valencia CA). The protocols were followed with the following substitution. After the debris from cell lysis was removed via centrifugation, the supernatant was poured through miracloth (EMD biosciences, CA) to remove the remaining cell lysate.

After purification of the plasmids, the purity of the DNA was assessed using the ratio of absorbance values at 260 nm and 280 nm. For pure DNA, the ratio of A260:A280 is 1.8; however, for sequencing purposes, any ratio between 1.7 and 1.9 is acceptable. Because the DNA concentration is directly proportional to absorbance, the concentration of the DNA sample can be determined by multiplying the absorbance reading at 260 nm with the extinction coefficient for DNA (50 ug/ml has A260 = 1.0). The DNA was then adjusted to a concentration of 250ug/ml by adding sterile nanopure water.

For samples that were less than the desired concentration, the following protocol was used. The samples were first treated with 0.1 volume of 3 M sodium acetate and 2.2 volumes of 95% ethanol to precipitate the DNA. The precipitate was centrifuged at 14,000 rpm for 15 minutes to pellet the DNA. After the centrifugation, the sodium acetate/ethanol solution was removed and the pelleted DNA was washed with 200 ul of 70% ethanol. The solution was centrifuged at 14,000 rpm for two minutes to repellet the DNA. Once the centrifugation was completed, the ethanol was removed and the DNA pellet was dried using vacuum dessication. The pellet of DNA was then redissolved to the proper volume using 0.5 volumes of TE and 0.5 volumes of ddH_2O.

Once both the purity and the concentration of the DNA were verified and adjusted, gel

electrophoresis was used to ensure that the clones were the full length of the expected molecule. Restriction digests for each clone were set up using EcoRI to excise the clone from the plasmid. The restriction digest contained 4 ul of the sample DNA at a concentration of 250 ug/ml, 2 ul of 10x EcoRI buffer, and 1 ul of the EcoRI restriction enzyme (10 units/ul). The total volume for the digest was brought to 20 ul with 13 ul ddH_2O. The restriction digest was incubated at 37°C for one hour. Both the undigested sample and the EcoRI-digested material were electrophoresed using a 1% agarose gel at 80 volts for approximately 1.5 hours. A 100-base pair ladder was electrophoresed along with the samples to assess the size of the fragments produced by the digest. The clones were expected to have a length of approximately 1200 base pairs after being excised from the plasmid, and any clones that were significantly truncated were excluded from sequencing.

For each full length clone, cycle sequencing reactions were set up using 2.5 ul of the template DNA from the clone as well as 4 ul terminator ready reaction mix (Applied Biosystem, Foster City CA), 4 ul halfBD extender (Genetix, Boston MA), and 4 ul sequencing primer at a concentration of 0.8 pmole/ul. Because of the length of the insert, internal primers (Table 2) were used in conjunction with the standard external primers. The internal primers used were derived from realized sequence information. The reactions were heated to 96°C for 2 minutes, and then amplification was conducted for 25 cycles as follows: denature for 10 seconds at 96°C, anneal for 5 seconds at 50°C, and extend for 4 minutes at 60°C. Upon completion, the samples were held at 4°C.

Table 2. Primers used for amplification and sequencing of MHC class I cDNAs from *M. fascicularis*

Amplification Primers

Primer	Binding region	Sequence
5' MAS	Sense exon 1(10-34) A locus	AATTCATGGCGCCCCGAACCCTCCTCCTGG
3' MAS	Antisense exon 8 (2)-3'UTR(18) A locus	CTAGACCACACAAGGCGGCTGTCTCAC
5' MBS	Sense exon 1(10-34) B locus	AATTCATGGCGCCCCGAACCCTCCTCCTGC
3' MBS	Antisense exon 8 (2)-UTR(18) B locus	CTAGACCACACAAGACAGTTGTCTCAG
NA1START[b]	Sense exon 2 (1-23) All Loci	GCTCYCACTCCWTGARGTATTTC
A2END	Antisense exon 3 (248-272) All Loci	GCGCTGCAGCGTCTCCTTCCCGTTC
A3MID	Antisense exon 4 (85-104) All Loci	CCAGGTCAGTGTGATCTCCG

Sequencing Primers

Primer	Binding region	Sequence
MAFA1S	Sense	CTACAACCAGAGCGAGG
MAFA1A	Antisense	CCTCGCTCTGGTTGTAG
MAFA2A	Sense	CACCACAGCTGCCCACTTC
MAFA2S	Antisense	GAAGTGGGCAGCTGTGGTG

Table 2. Amplification primers were obtained from the research done on *M. mulatta (*6).
[a] Sequencing primers were generated from partial sequences of *M. fascicularis*
[b] Mixed bases in the above primers are as follows; Y =C or T, R = A or G, and W = A or T.

The cleanup for the cycle sequencing reactions was done by adding 2 ul of 3 M sodium acetate (pH 5.3) and 50 ul 95% ethanol to the reaction mix product. The tube was mixed and then placed on ice for 15 minutes. After sitting on ice, the solution was centrifuged at 14,000 rpm for 30 minutes. Upon completion of the centrifugation, as much of the sodium acetate/ethanol as possible was removed. The pellet containing the reaction product was washed using 250 ul of 70% ethanol and then centrifuged for 1 minute at 14,000 rpm. The ethanol was then aspirated off and the pellet was dried using a speed vacuum concentrator. Once the pellet was completely dried, it was suspended in 25 ul of TSR (template suppression reagent). The sample was then heated at 95^{o}C for 2 minutes. The sample was then placed in a small DNA sample tube to be sequenced.

The samples were sequenced using an ABI Prism 310 Genetic Analyzer (Applied Biosystem, Foster City CA). The fragments generated by the sequencing reactions were aligned using the suite of fragment assembly programs (FAS) available in the Wisconsin GCG package to produce complete double-stranded nucleotide sequences for the genes. Each cycle sequencing reaction for a given primer was done twice in order to verify the nucleotide positions.

Synonymous and Non-Synonymous Substitution Rates

In order to calculate the non-synonymous and synonymous substitution rates in the Macaque class I MHC molecules, a multiple sequence alignment of the nucleotide sequences was generated using the pileup program in the Wisconsin GCG Package. The resulting alignment was imported into MEGA2 (11) for substitution rate analysis. The positions for each of the regions, (antigen-binding positions from α1 and α2 domains, the remainder of the positions in the α1 and α2 domains, and the α3 domain), were selected using the sequence editing portion of the software. The substitution rates for those positions were calculated using the Jukes-Cantor statistical method available in the MEGA2 package.

Nomenclature

In general, the nomenclature for MHC molecules is fairly standardized. For the majority of the organisms, including the Macaques, the standard nomenclature is used. The first part of the molecule designation is derived from the organism name. The first two characters of the genus and the first two characters of the species are combined to create a four letter abbreviation of the organism. Therefore, the standard abbreviation for every molecule derived from *Macaca fascicularis* would have a designation that begins with Mafa. After the organism has been established in the name, the locus that the particular allele belongs to is added to the name so as to provide as much information about the particular sequence as possible. In the case of a B-locus MHC class I molecule from *M. fascicularis*, the designation would be Mafa-B. The last

part of the designation is number that refers to the specific allele, such as Mafa-B*01, which would be the first of the B-locus molecules named from *M. fascicularis.* Following the nomenclature patterns, the other members of the Macaque family, *M. mulatta* and *M. nemestrina* are designated using Mamu and Mane respectively. One of the most notable exceptions to this standard nomenclature is the Human MHC molecules. Instead of using the four letter abbreviation for the species, Humans use the designation 'HLA', short for human leukocyte antigen.

Wisconsin GCG Package

The Wisconsin GCG is a software package that combines programs that are grouped according to function. These programs provide the ability to conduct sequence database searches, DNA fragment assembly, protein analysis, multiple sequence analysis, and phylogenetic analysis, among many others. Below each of the program groups that were used are discussed in more detail.

The Fragment Assembly System (FAS) in the Wisconsin GCG package is a suite of programs that facilitate the assembly, alignment and editing of overlapping sequence fragments. The GelStart program is the initial program for creating an assembly project. GelStart creates a database for the project as well as the necessary directories for storage of the fragments and the alignments produced. After the project has been instantiated, the sequence fragments may be entered using the GelEnter program. This program accepts the sequence data from existing files as well as from the terminal keyboard. Sequences that have been entered into the project using GelEnter can no longer be edited without using the FAS programs. The actual alignment of the sequence fragments is carried out by GelMerge. GelMerge recognizes overlaps between sequence fragments and attempts to align them to construct contiguous sequences (contigs). As more sequence fragments are aligned using GelMerge, the program attempts to align the fragments and contigs into larger assemblies, eventually resulting in a complete sequence based

on the original fragments.

The GelAssemble program acts as a contig editor that allows the user to review and edit the alignments generated by GelMerge. GelAssemble also allows for manual alignments, enabling fragments without sufficient overlap for GelMerge to function, to be aligned by the user. The editing function also allows the user to resolve any disparities in the alignments of fragments. The FAS also provides a graphical representation of the contigs using GelView allowing for a schematic view of how all of the contigs are aligned in the project. This is most useful in assessing the completeness of the fragment assembly that was generated.

To carry out pairwise sequence analysis, several programs are provided which allow different types of alignment to occur based on the algorithm used. Because of the data being analyzed, only one of the pairwise analysis programs provided was appropriate. For the optimal alignment of two entire sequences, the Gap module is provided. The Gap module uses the Needleman-Wunsch algorithm to align two entire sequences that maximizes the number of matches while minimizing the number of gaps necessary to create the alignment The Gap module works best when the sequences being aligned are similar in length in order to help minimize the number of gaps necessary to generate the alignment.

The module in the Wisconsin package used to generate the Multiple Sequence Alignments (MSAs) is the PileUp module. The PileUp module computes the multiple sequence alignment by completing successive pairwise alignments between sequences. The most similar sequences are aligned first, and then each sequence aligned against a consensus sequence from the previous alignment. These alignments are based on the same algorithm that is used in the Gap module for pairwise alignments. PileUp will also provide a tree structure that shows the clustering relationships that it uses in order to compute the final multiple sequence alignment.

The Wisconsin package provides several programs in addition to PileUp to display multiple sequence alignments, Pretty being one such program. Pretty works both on the traditional MSA files as well as rich sequence alignments (RSA). The Pretty module takes

existing MSAs and allows the user to make several types of modifications. Some of these modifications include the ability to present a consensus sequence of the alignment, as well as the ability to mask every position in each sequence that agrees with the consensus. Such modifications make it much easier to distinguish differences between sequences in the MSA.

The suite of phylogenetic modules provided by the Wisconsin package provide two ways to generate phylogenetic trees using a MSA. The two main programs available to create a phylogenetic tree from data are the Neighbor-Joining method (17), and the Parsimony method (18). In order to compensate for possibilities such as multiple substitutions occurring at a given position over time, several statistical algorithms for calculating differences are available. Each of the above programs attempt to construct phylogenies in a distinct fashion. The Neighbor-Joining module uses distances for phylogenetic analysis. This method generates a distance table from all possible pairwise comparisons and constructs a phylogenetic tree based on the table. The phylogenetic tree is constructed using the number of differences between the sequences as its branch length between two sequences.

The PaupSearch module is based on the parsimony method for phylogenetic analysis. While the distance method generates all pairwise alignments and calculates the number of differences between sequences, the parsimony method attempts to construct phylogenies by finding the tree with the fewest number of changes necessary to get from a given sequence to any other. The parsimony method also incorporates optimality criteria in order to resolve on the solution. Because this is such a computationally intensive process, there are three possible algorithms that can be used in order to minimize the time spent generating the trees. One pitfall of the the Neighbor-Joining method is that it will arbitrarily break the tree in order to resolve on one solution. While you always resolve on one tree when using the Neighbor-Joining method, it is quite possible, and often very common to obtain many equally parsimonious trees using the Parsimony method.

Pocket Extraction

A crucial aspect of analyzing the Class I MHC molecules in *M. fascicularis* relies on the ability to compare the amino acid composition of the B and F pockets in the binding groove of the molecule. A small perl script was designed to automate the extraction of the sequences (Appendix 2). This program has been designed to work both on unix systems, as well as via web form. It relies on FASTA formatted files of the amino acid sequences as the data format. The choice to use the FASTA format was two-fold. The first being FASTA is a standard file format that all major databases support, so the information is readily available. The second reason for using FASTA is the ease at which it can be parsed due to the simplicity of the file structure.

The pocket extraction program works by searching the amino acid sequence until it finds the start of the α1 domain. The five amino acids surrounding the start site are very highly conserved, all five being present in ~95% of all sequences presently characterized. The other 5% or so have single amino acid differences almost always at the first amino acid position used in the comparison. The reason for using all five of the positions instead of using the four that are conserved in all sequences is that using five positions minimizes the chance of matching another site with the same amino acid sequence.

Based on where the start site of the α1 domain is located, the program then extracts the amino acids at the hard coded positions for the B and F binding pockets (Table 3).

Table 3 Amino acid positions in the mature protein contributing to the B and F binding pockets.

B Pocket Positions	***F Pocket Positions***
7	77
9	80
24	81
25	84
34	95
45	116
63	123
66	143

B Pocket Positions	*F Pocket Positions*
67	146
70	147
99	

Table 3 The amino acid positions that contribute to the B and F binding pockets. Amino acid positions are counted with position one being the first amino acid in the α1 domain.

This information is then returned to the user in a readable and usable format.

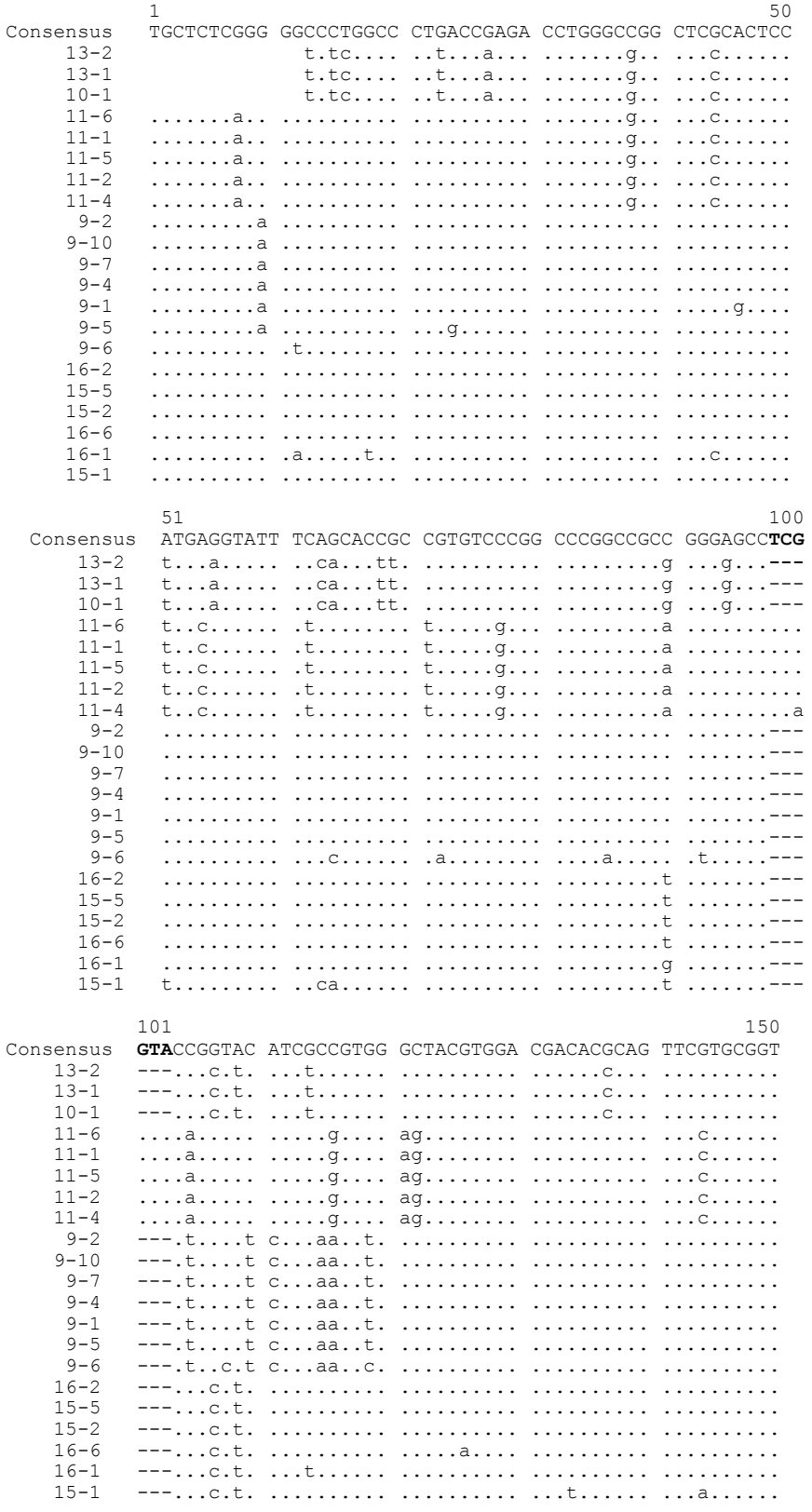
Peptide Binding Prediction

In order to determine what peptide fragments from a given protein could be bound and presented by the MHC class I molecules in *M. fascicularis*, the MAPPP (MHC-I Antigenic Peptide Processing Prediction) tool was used (12). MAPPP uses two distinct algorithms to predict the fragments that will be generated by a given amino acid sequence. For the largest set of peptide fragments, PaproC uses the algorithm that is designed to predict fragments based on the affinity of the positions when exposed to the active site of the proteasome (13). The second program, FragPredict, attempts to predict amino acid fragments by analyzing the transfer energy of the side-chains on the fragment(14).

RESULTS

After the sequences had been obtained for the *M. fascicularis* MHC class I molecules. The nucleotide sequences were aligned in an MSA using the pileup method in the Wisconsin GCG package (figure 5). In the nucleotide alignment, periods (.) indicate agreement with the consensus sequence at that position.

figure 5. Nucleotide MSA for clones from *M. fascicularis*.



```
          151                                                 200
Consensus TCGACAGCGA CGCCGAGAGT CCGAGGATGG AGCCGCGGGC GCCGTGGGTG
    13-2  at........ .....c.... .a........ .......... ..........
    13-1  at........ .....c.... .a........ .......... ..........
    10-1  at........ .....c.... .a........ .......... ..........
    11-6  .......... .....c..t. .......... .......... ..........
    11-1  .......... .....c..t. .......... .......... ..........
    11-5  .......... .....c..t. .......... .......... ..........
    11-2  .......... .....c..t. .......... .......... ..........
    11-4  .......... .....c..t. .......... .......a.. ..........
     9-2  .......... .......... .......... .......... ..........
    9-10  .......... .......... .......... .......... ..........
     9-7  .......... .......... .......... .......... ..........
     9-4  .......... .......... .......... .......... ..........
     9-1  .......... .......... .......... .......... ..........
     9-5  .......... .......... .......... .......... ..........
     9-6  .......... ..t....... .......... .a...t.... .......a..
    16-2  .......... .......... .......... .......... ..g....a.a
    15-5  .......... .......... .......... .......... ..g....a.a
    15-2  ..a....... .......... .......... .......... ..g....a.a
    16-6  .......... .......... .......... .......... ..g....a.a
    16-1  .......... .......... .......... ....t..... .......a.a
    15-1  .......... .......... .....aga.. .......... ...c....c.

          201                                                 250
Consensus GAGCAGGAGG GGCCGGAGTA TTGGGACCGG AACACACGGA ACGCCAAGG-
    13-2  .......... .......... ........a. g.g....... g.....g..a
    13-1  .......... .......... ........a. g.g....... g.....g..a
    10-1  .......... .......... ........a. g.g....... g.....g..a
    11-6  .....a.... .....cg... ......g... .c....g..t .........c
    11-1  .....a.... .....c.... ......g... .c....g..t .........c
    11-5  .....a.... .....c.... ......g... .c....g..t .........c
    11-2  .....a.... .....c.... ......g... .c....g..t .........c
    11-4  .....a.... .....c.... ......g... .c....g..t ........ac
     9-2  .......... .......... .......... .......... ..t......t
    9-10  .......... .......... .......... .......... ..t......t
     9-7  .......... .......... .......... .......... ..t......t
     9-4  .......... .......... .......... .......... ..t......t
     9-1  .......... .......... .......... .......... ..t......t
     9-5  .......... .......... .......... .......... ..t......t
     9-6  .......... .......... ......aga. g.g....... ga.......c
    16-2  .......... .......... ......aga. g.g....... .a.......a
    15-5  .......... .......... ......aga. g.g....... .a.......a
    15-2  .......... .......... ......aga. g.g....... .a.......a
    16-6  .......... .......... ......aga. g.g....... .a.......a
    16-1  .......... .......... ......aga. gcg....... t........c
    15-1  .......... .......... ......aga. c.g....... t........g

          251                                                 300
Consensus CACCGCACAG ACTTACCGAG TGAGCCTGGG GAACCTGCGC GGCTACTACA
    13-2  .......... ....t..... ...a.....a ..c....... ..........
    13-1  .......... ....t..... ...a.....a ..c....... ..........
    10-1  .......... ....t..... ...a.....a ..c....... ..........
    11-6  ..a.....g. ...g...... ..gc....a. ...g....t. ct.cg.....
    11-1  ..a.....g. ...g...... ..gc....a. ...g....t. ct.cg.....
    11-5  ..a.....g. ...g...... ..gc....a. ...g....t. ct.cg.....
    11-2  ..a.....g. ...g...... ..gc....a. ...g....t. ct.cg.....
    11-4  ..a.....g. ...g...... ..gc....a. ...g....t. ct.cg.....
     9-2  .......... ...c...... .......... .......... ..........
    9-10  .......... ...c...... .......... .......... ..........
     9-7  .......... ...c...... .......... .......... ..........
     9-4  .......... ...c...... .......... .......... ..........
     9-1  .......... ...c...... .......... .......... ..........
     9-5  .......... ...c...... .......... .......... ..........
     9-6  ..a....... ....t..... g..a....c. ..c.gc..t. c.........
    16-2  .g........ t.c.t..... ..g....... ..t....... ..........
    15-5  .g........ t.c.t..... ..g....... ..t....... ..........
    15-2  .g........ t.c.t..... ..g....... ..t....... ..........
    16-6  .g........ t.c.t..... ..g....... ..t....... ..........
    16-1  .cg....... ...g...... ..ga...... ..c....... ..........
    15-1  ..a....... g..c....g. g..a....c. ........t. c.........
```

```
          301                                                        350
Consensus ACCAGAGCGA GGCCGGGTCT CACACCTTCC AG-GGATGTA CGGCTGCGAC
    13-2  .......... .......... ..t...c... ..t.....c. t.........
    13-1  .......... .......... ..t...c... ..t.....c. t.........
    10-1  .......... .......... ..t...c... ..t.....c. t.........
    11-6  .......... .......... ......c... ..g.a...a. ..........
    11-1  .......... .......... ......c... ..g.a...a. ..........
    11-5  .......... .......... ......c... ..g.a...a. ..........
    11-2  .......... .......... ......c... ..g.a...a. ..........
    11-4  .......... .......... ......c... ..g.a...a. ..........
     9-2  ........a. .......... .......a.. ..a....... ..........
    9-10  ........a. .......... .......a.. ..a....... ..........
     9-7  ........a. .......... .......a.. ..a....... ..........
     9-4  ........a. .......... .......a.. ..a....... ..........
     9-1  ........a. .......... .......a.. ..a....... ..........
     9-5  ........a. .......... .......a.. ..a....... ..........
     9-6  .......... ...g...... ......c... ..t....... t.........
    16-2  .......... ...g...... .......a.. ..t.....gt t......a..
    15-5  .......... ...g...... .......a.. ..t.....gt t......a..
    15-2  .......... ...g...... .......a.. ..t.....gt t......a..
    16-6  .......... ...g...... .......a.. ..t.....gt t......a..
    16-1  .......... ...g...... ......a... ..ac...... ..........
    15-1  .......... ..gg...... .....ag... ..aca..... ..........

          351                                                        400
Consensus CTGGGGCCCG ACGGGCGCCT CCTCCGCGGG TATCACCAGT ACGCCTACGA
    13-2  ......t... ........t. .......... ...g.a.... t.........
    13-1  ......t... ........t. .......... ...g.a.... t.........
    10-1  .....at... ........t. .......... ...g.a.... t.........
    11-6  a......... ....a..... .......... .........c ..........
    11-1  a......... ....a..... .......... .........c ..........
    11-5  a......... ....a..... .......... .........c ..........
    11-2  a......... ....a..... .......... .........c ..........
    11-4  a......... ....a..... .......... .........c ..........
     9-2  .......... .......... .......... ...t...... ..........
    9-10  .......... .......... .......... ...t...... ..........
     9-7  .......... .......... .......... ...t...... ..........
     9-4  .......... .......... .......... ...t...... ..........
     9-1  .......... .......... .......... ...t...... ..........
     9-5  .......... .......... .......... ...t...... ..........
     9-6  .....a.... .......... .......... .......... t.........
    16-2  g....a..a. ......a... .......... .......... t.........
    15-5  g....a..a. ......a... .......... .......... t.........
    15-2  g....a..a. ......a... .......... .......... t.........
    16-6  g....a..a. ......a... .......... .......... t.........
    16-1  .....a.... .......... .......... ..c.gg...g ..........
    15-1  g....a.... .......... .........c ...g.a.... t.........

          401                                                        450
Consensus CGGCAAGGAT TACATCGCCC TGAACGAGGA CCTGCGCTCC TGGACCGCCG
    13-2  .......... ..tc..a... ....t..... .......... ...t....g.
    13-1  .......... ..tc..a... ....t..... .......... ...t....g.
    10-1  .......... ..tc..a... ....t..... .......... ...t....g.
    11-6  .......... ......t... .......... .......... ..........
    11-1  .......... ......t... .......... .......... ..........
    11-5  .......... ......t... .......... .......... ..........
    11-2  .......... ......t... .......... .......... ..........
    11-4  .......... ......t... .......... .......... ..........
     9-2  .......... .......... .......... .......... ........g.
    9-10  .......... .......... .......... .......... ........g.
     9-7  .......... .......... .......... .......... ........g.
     9-4  .......... .......... .......... .......... ........g.
     9-1  .......... .......... .......... .......... ........g.
     9-5  .......... .......... .......... .......... ........g.
     9-6  .......... .......... .......... .......... ..........
    16-2  .......... .......... .....a.... .......... ..........
    15-5  .......... .......... .....a.... .......... ..........
    15-2  .......... .......... .....a.... .......... ..........
    16-6  .......... ...g...... .....a.... .......... ..........
    16-1  .......... .......... .......... .......... ..........
    15-1  .....g.... .......... .......... .......... ..........
```

```
          451                                                        500
Consensus  CGGACACGGC GGCTCAGAAC ACCCAGCGCA AGTGGGAGGC GGCCGGTGAG
    13-2   t......... ......a.t. t..g...aa. ...caa.t.a t.g.tc....
    13-1   t......... ......a.t. t..g...aa. ...caa.t.a t.g.tc....
    10-1   t......... ......a.t. t..g...aa. ...caa.t.a t.g.tc....
    11-6   .........t a....g..t. .........t tc.at..... a.ag.aat.t
    11-1   .........t a....g..t. .........t tc.at..... a.ag.aat.t
    11-5   .........t a....g..t. .........t tc.at..... a.ag.aat.t
    11-2   .........t a....g..t. .........t tc.at..... a.ag.aat.t
    11-4   .........t a....g..t. .........t tc.at..... a.ag.aat.t
     9-2   .a........ .......... .......a.. .......... ........t.
    9-10   .a........ .......... .......a.. .......... ........t.
     9-7   .a........ .......... .......a.. .......... ........t.
     9-4   .a........ .......... .......a.. .......... ........t.
     9-1   .a........ .......... .......a.. .......... ........t.
     9-5   .a........ .......... .......a.. .......... ........t.
     9-6   ......t... .......... .......... .......... ....c...c.
    16-2   g.....t... .......... .......... .......... ..t.......
    15-5   g.....t... .......... .......... .......... ..t.......
    15-2   g.....t... .......... .......... .......... ..t.......
    16-6   g.....t... .......... .......... .......... ..t.......
    16-1   ......t... .......... .......... .........g ..a.c..t.t
    15-1   ......t... .......... .......... .......... ....c...ca

          501                                                        550
Consensus  GC-GAGCAGT -CAGAGCCTA CCTGGAGGGC ACGTGCGTGG AGTGGCTCCG
    13-2   ..t.....cc ag........ ......a.a. ..a....... .........a
    13-1   ..t.....cc ag........ ......a.a. ..a....... .........a
    10-1   ..t.....cc ag........ ......a.a. ..a....... .........a
    11-6   ..a...g... t...ga.... .......... ga....c... ...t......
    11-1   ..a...g... t...ga.... .......... ga....c... ...t......
    11-5   ..a...g... t...ga.... .......... ga....c... ...t......
    11-2   ..a...g... t...ga.... .......... ga....c... ...t......
    11-4   ..a...g... t...ga.... .......... ga....c... ...t......
     9-2   ..g..a...a ag........ .........g .......... ..........
    9-10   ..g..a...a ag........ .........g .......... ..........
     9-7   ..g..a...a ag........ .........g .......... ..........
     9-4   ..g..a...a ag........ .........g .......... ..........
     9-1   ..g..a...a ag........ .........g .......... ..........
     9-5   ..g..a...a ag........ .........g .......... ..........
     9-6   ..g....... gg......c. g......... .a........ ..........
    16-2   ..a....g.. t......... .g........ cg........ ..........
    15-5   ..a....g.. t......... .g........ cg........ ..........
    15-2   ..a....g.. t......... .g........ cg........ ..........
    16-6   ..a....g.. t......... .g........ cg........ ..........
    16-1   ..g....g.. t......... .......... cg........ ......c...
    15-1   ..g...a..g a......... ...c.....g c......... ..........

          551                                                        600
Consensus  CAGATACCTG GAGAACGGGA AGGAGACGCT GCAGCGCGCG GATCCCCCAA
    13-2   .......... .....t.... .......... .......t.a ..a.......
    13-1   .......... .....t.... .......... .......t.a ..a.......
    10-1   .......... .....t.... .......... .......t.a ..a.......
    11-6   .......... .......... .......... a........a .....t....
    11-1   .......... .......... .......... a........a .....t....
    11-5   .......... .......... .......... a........a .....t....
    11-2   .......... .......... .......... a........a .....t....
    11-4   .......... .......... .......... a........a .....t....
     9-2   ....c..... .......... .......... .......... ..........
    9-10   ....c..... .......... .......... .......... ..........
     9-7   ....c..... .......... .......... .......... ..........
     9-4   ....c..... .......... .......... .......... ..........
     9-1   ....c..... .......... .......... .......... ..........
     9-5   ....c..... .......... .......... .......... ..........
     9-6   .......... .......... .......... .......... ..........
    16-2   .......... .......... .......... .......... ..........
    15-5   .......... .......... .......... .......... ..........
    15-2   .......... .......... .......... .......... ..........
    16-6   .......... .......... .......... .......... ..........
    16-1   .......... .......... .......... .......... ..........
    15-1   .......... .......... .......... .......... ..........
```

```
          601                                                       650
Consensus AGACACACGT GACCCACCAC CCCGTCTCTG ACCATGAGGC CACCCTGAGG
    13-2  .......... ........g. .......... .t........ ..........
    13-1  .......... .......... .......... .t........ ..........
    10-1  .......... .......... .......... .t........ ..........
    11-6  ..g....... tg........ .a.a...... ...g...... ..........
    11-1  ..g....... tg........ .a.a...... ...g...... ..........
    11-5  ..g....... tg........ .a.a...... ...g...... ..........
    11-2  ..g....... tg........ .a.a...... ...g...... ..........
    11-4  ..g....... tg........ ...a...... ...g...... ..........
     9-2  .......... .........t ..t....... .......... .........a
    9-10  .......... .........t ..t....... .......... .........a
     9-7  .......... .........t ..t....... .......... .........a
     9-4  .......... .........t ..t....... .......... .........a
     9-1  .......... .........t ..t....... .......... .........a
     9-5  .......... .........t ..t....... .......... .........a
     9-6  .......... .........t .......... .......... ..........
    16-2  .......... .......... .......... .......... ..........
    15-5  .......... .......... .......... .......... ..........
    15-2  .......... .......... .......... .......... ..........
    16-6  .......... .......... .......... .......... ..........
    16-1  .......... .......... .......... .......... ..........
    15-1  .......... .......... .......... .......... ..........

          651                                                       700
Consensus TGCTGGGCCC TGGGCTTCTA CCCTGCGGAG ATCACACTGA CCTGGCAGCG
    13-2  .......... .......... t......... ......g... ..........
    13-1  .......... .......... t......... ......g... ..........
    10-1  .......... .......... t......... ......g... ..........
    11-6  .......... .......... ......c... .......... ..........
    11-1  .......... .......... ......c... .......... ..........
    11-5  .......... .......... ......c... .......... ..........
    11-2  .......... .......... ......c... .......... ..........
    11-4  .......... .......... ......c... .......... ..........
     9-2  .......t.. .......... .......... .......... ..........
    9-10  .......t.. .......... .......... .......... ..........
     9-7  .......t.. .......... .......... .......... ..........
     9-4  .......t.. .......... .......... .......... ..........
     9-1  .......t.. .......... .......... .......... ..........
     9-5  .......t.. .......... .......... .......... ..........
     9-6  .......... .......... .......... .......... ..........
    16-2  .......... .......... .......... .......... ..........
    15-5  .......... .......... .......... .......... ..........
    15-2  .......... .......... .......... .......... ..........
    16-6  .......... .......... .......... .......... ..........
    16-1  .......... .......... .......... .......... ..........
    15-1  .......... .......... .......... .......... ..........

          701                                                       750
Consensus GGATGGGGAG GACCAAACTC AGGACACGGA GCTTGTGGAG ACCAGGCCAG
    13-2  .......... .....g..c. .......... .......... ........t.
    13-1  .......... .....g..c. .......... .......... ........t.
    10-1  .......... .....g..c. .......... ....t..... ........t.
    11-6  ...c...... ..a..g..c. .......... .......... ........t.
    11-1  ...c...... ..a..g..c. .......... .......... ........t.
    11-5  ...c...... ..a..g..c. .......... .......... ........t.
    11-2  ...c...... ..a..g..c. .......... .......... ........t.
    11-4  ...a...... ..a..g..c. .......... .......... ........t.
     9-2  .......... .......... .......... .......... ..........
    9-10  .......... .......... .......... .......... ..........
     9-7  .......... .......... .......... .......... ..........
     9-4  .......... .......... .......... .......... ..........
     9-1  .......... .......... .......... .......... ..........
     9-5  .......... .......... ..ag...... .......... ..........
     9-6  .......... ..a....... .......c.. .......... ..........
    16-2  .......... .......... .......c.. .......... ..........
    15-5  .......... .......... .......c.. .......... ..........
    15-2  .......... .......... .......c.. .......... ..........
    16-6  .......... .......... .......c.. .......... ..........
    16-1  .......... ..g....... .......c.. .......... ..........
    15-1  .......... .......... .......c.. ...c...... ..........
```

```
          751                                                       800
Consensus CAGGAGATGG AACCTTCCAG AAGTGGGGAG CTGTGGTGGT GCCTTCTGGA
    13-2  ....g..... ...t...... .......c.. .......... a.........
    13-1  ....g..... ...t...... .......c.. .......... a.........
    10-1  ....g..... ...t...... .......c.. .......... a.........
    11-6  ....g..... .......... .......c.. .......... ......c...
    11-1  ....g..... .......... .......c.. .......... ......c...
    11-5  ....g..... .......... .......c.. .......... ......c...
    11-2  ....g..... .......... .......c.. .......... ......c...
    11-4  ....g..... .......... .......c.. .......... ......c...
     9-2  t......... .......... .......... .......... ..........
    9-10  t......... .......... .......... .......... ..........
     9-7  t......... .......... .......... .......... ..........
     9-4  t......... .......... .......... .......... ..........
     9-1  t....a.... .......... .......... .......... ..........
     9-5  t......... .......... .......... .......... ..........
     9-6  t......... .......... .......... .......... ..........
    16-2  g......... .......... .......... .......... ..........
    15-5  g......... .......... .......... .......... ..........
    15-2  g......... .......... .......... .......... ..........
    16-6  g......... .......... .......... .......... ..........
    16-1  g......... .......... .......... .......... ..........
    15-1  g......... .......... .......... .......... ..........

          801                                                       850
Consensus GAAGAGCAGA GATACACGTG CCATGTGCAG CACGAGGGGC TGCCGGAGCC
    13-2  ..g....... .......... .......... ..t....... .a..c.....
    13-1  ..g....... .......... .......... ..t....... .a..c.....
    10-1  ..g....... .......... .......... ..t....... .ag.c.....
    11-6  ..g....... .......... .........a .....a.... ....cc....
    11-1  ..g....... .......... .........a .....a.... ....cc....
    11-5  ..g....... .......... .........a .....a.... ....cc....
    11-2  ..g....... .......... .........a .....a.... ....cc....
    11-4  ..g....... .......... .........a .....a.... ....cc....
     9-2  .......... .......... .......... ...c...... ..........
    9-10  .......... .......... .......... ...c...... ..........
     9-7  .......... .......... .......... ...c...... ..........
     9-4  .......... .......... .......... ...c...... ......g...
     9-1  .......... .......... .......... ...c...... ..........
     9-5  .......... .......... .......... ...c...... ..........
     9-6  .......... .......... t......... ...c...... ..........
    16-2  .......... .......... .......... ..t....... ..........
    15-5  .......... .......... .......... ..t....... ..........
    15-2  .......... .......... .......... ..t....... ..........
    16-6  .......... .......... .......... ..t....... ..........
    16-1  .......... .......... .......... ..t....... ..........
    15-1  .......... .......... .......... ..t....... ..........

          851                                                       900
Consensus CCTCACCCTG AGATGGGAGC CATCTTCCCA GTCCACCATC CCCATCGTGG
    13-2  ..gag..... .......... .g.......g .......... ..........
    13-1  ..gag..... .......... .g.......g .......... ..........
    10-1  .......... .......... .g........ .......... ..........
    11-6  .......... .........t .g.....t.. .c........ .....t....
    11-1  .......... .........t .g.....t.. .c........ .....t....
    11-5  .......... .........t .g.....t.. .c........ .....t....
    11-2  .......... .........t .g.....t.. .c.....g.. .....t....
    11-4  .......... .........t .g.....t.. .c........ .....t....
     9-2  .......... .......... .......... .......... ..........
    9-10  .......... .......... .......... .......... ..........
     9-7  .......... .......... .......... ....g..... ..........
     9-4  .......... .......... .......... .......... ..........
     9-1  .......... .......... .......... .......... ..........
     9-5  .......... .......... .......... .......... ..........
     9-6  .......... .......... .......... .......... ..........
    16-2  .......... .......... .......... .......... ..........
    15-5  .......... .......... .......... .......... ..........
    15-2  .......... .......... .......... .......... ..........
    16-6  .......... .......... .......... .......... ..........
    16-1  .......... .......... .......... .......... ..........
    15-1  .......... .......... .......... .......... ..........
```

```
          901                                                       950
Consensus GCATCGTTGC TGGCCTGGCT GTCCTAGCAG TTGTGGTCAC TGGAGCTGTG
    13-2  .....a.... ........t. c....t.g.. c......... ..........
    13-1  .....a.... ........t. c....t.g.. c......... ..........
    10-1  .....a.... ........t. c....t.g.. c......... ..........
    11-6  .......... .......... .......... .......t.. ..........
    11-1  .......... .......... .......... .......t.. ..........
    11-5  .......... .......... .......... .......t.. ..........
    11-2  .......... .......... .......... .......t.. ..........
    11-4  .......... .......... .......... .......... ..........
     9-2  ....t..... .......... .......... .......... c.........
    9-10  ....t..... .......... .......... .......... c.........
     9-7  ....t..... .......... .......... .......... c.........
     9-4  ....t..... .......... .......... .......... c.........
     9-1  ....t..... .......... .......... .......... c.........
     9-5  ....t..... .......... .......... .......... c.........
     9-6  ....t..... .......... .......... .......... c.........
    16-2  .........t .......... .......... .......... ..........
    15-5  .........t .......... .......... .......... ..........
    15-2  .........t .......... .......... .......... ..........
    16-6  .........t .......... .......... .......... ..........
    16-1  ....t....t .......... .......... .......... c.........
    15-1  ....t....t ....t..... .......... .......... c.........

          951                                                      1000
Consensus GTCGCTGCTG TGATGTGGAG GAGGAAGAGC TCAGGTGGAA AAGGAGGGAG
    13-2  ..t.t..... .......... .......... ......a... ..........
    13-1  ..t.t..... .......... .......... ......a... ..........
    10-1  ..t.t..... .......... .......... ......a... ..........
    11-6  .......... .a........ ...a...... ....a.a... .ca.......
    11-1  .......... .a........ ...a...... ....a.a... .ca.......
    11-5  .......... .a........ .......... ....a.a... .ca.......
    11-2  .......... .a........ ...a...... ....a.a... .ca.......
    11-4  .......... .a.......a .......... ....a.a... .ca.......
     9-2  .......... .......... .......... .......... ..........
    9-10  .......... .......... .......... .......... ..........
     9-7  .......... .......... .......... .......... ..........
     9-4  .......... .......... .......... .......... ..........
     9-1  .......... .......... .......... .......... ..........
     9-5  .......... .......... .......... .......... ..........
     9-6  .......... .......... .......... .......... ..........
    16-2  .......... .......... .......... c......... ..........
    15-5  .......... .......... .......... .......... ..........
    15-2  .......... .......... .......... .......... ..........
    16-6  .......... .......... .......... .......... ..........
    16-1  .......... .......... .......... .......... ..........
    15-1  .......... .......... .......... .......... ...c......

          1001                                                     1050
Consensus CTACTCTCAG GCTGCGTCCA GCGACAGTGC CCAGGGCTCT GATGTGTCTC
    13-2  .......... ...t...gt. ........a. ......a... ....a.....
    13-1  .......... ...t...gt. ........a. ......a... ....a.....
    10-1  .......... ...tt..gt. ........a. ......a... ....a.....
    11-6  .......... c..a.aat
    11-1  .......... c..a.aat
    11-5  .......... c..a.aat
    11-2  .......... c..a.aat
    11-4  .......... c..a.aat
     9-2  .......... .......... .......... .......... ..........
    9-10  .......... .......... .......... .......... ..........
     9-7  .......... .......... .......... .......... ..........
     9-4  .......... .......... .......... .......... ..........
     9-1  .......... .......... .......... .......... ..........
     9-5  .......... .......... .......... .......... ..........
     9-6  .......... .......... .......... .......... ..........
    16-2  .......... .......... .......... .......... ..........
    15-5  .......... .......... .......... .......... ..........
    15-2  .......... .......... .......... .......... ..........
    16-6  .......... .......... .......... .......... ..........
    16-1  .......... .......... at........ .......... ..........
    15-1  .......... .......... at........ .......... ..........
```

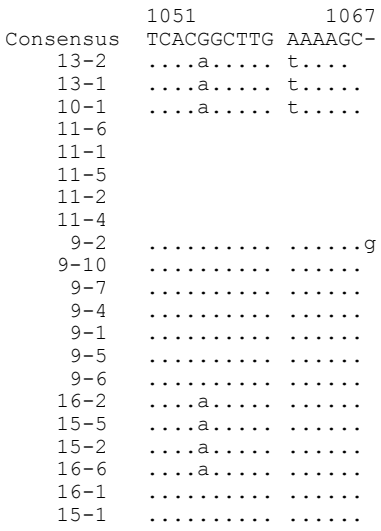


figure 5. Multiple sequence alignment of the nucleotide sequences of *M. fascicularis* MHC class I mRNAs. The sequences are compared to a consensus sequence derived from the data. The Mafa-E locus alleles all contain a six nucleotide size polymorphism. The polymorphism is indicated using bold font for the positions. The gaps present in the other sequences at this polymorphism are demarcated using dashes (-). Dashes(-) that appear in the consensus sequence indicate positions that do not resolve to a single consensus character.

After the MSA for the nucleotide sequences had been generated, each sequence that had been realized was then translated into the prospective amino acid sequence for that molecule. The protein sequences were then aligned into an MSA using the pileup program available in the Wisconsin GCG package (figure 6).

figure 6. Amino acid MSA for *M. fascicularis* sequences.

Leader Peptide

```
        -12          -1
consensus LS GALALTETWA
13-2           S...K...
13-1           S...K...
10-1           S...K...
11-5       .. ..........
11-1       .. ..........
11-2       .. ..........
11-6       .. ..........
11-4       .. ..........
9-7        .. E.........
9-2        .. E.........
9-10       .. E.........
9-4        .. E.........
9-5        .. E....A....
9-1        .. E.........
16-6       .. ..........
15-5       .. ..........
16-2       .. ..........
15-2       .. ..........
16-1       .. .T.S......
9-6        .. .S........
15-1       .. ..........
```

α 1

```
                                                                                                                    1
60
consensus GSHSMRYFST AVSRPGRREP RYRYIAVGYV DDTQFVRFDS DAESPRMEPR APWVEQEGPE
13-2      ....LK..H. S......GG. --.F.S.... .......Y.. ..A.Q..... ..........
13-1      ....LK..H. S......GG. --.F.S.... .......Y.. ..A.Q..... ..........
10-1      ....LK..H. S......GG. --.F.S.... .......Y.. ..A.Q..... ..........
11-5      ....L..... .......... .......E.. .....L.... ..AI...... .........Q
11-1      ....L..... .......... .......E.. .....L.... ..AI...... .........Q
11-2      ....L..... .......... .......E.. .....L.... ..AI...... .........Q
11-6      ....L..... .......... .......E.. .....L.... ..AI...... .........R
11-4      ....L..... .......... Q......E.. .....L.... ..AI...... .........Q
9-7       .......... .......... --W.LE.... .......... .......... ..........
9-2       .......... .......... --W.LE.... .......... .......... ..........
9-10      .......... .......... --W.LE.... .......... .......... ..........
9-4       .......... .......... --W.LE.... .......... .......... ..........
9-5       .......... .......... --W.LE.... .......... .......... ..........
9-1       ..R....... .......... --W.LE.... .......... .......... ..........
16-6      .......... .......W.. --.F.....M .......... .......... .R.I......
15-5      .......... .......W.. --.F...... .......... .......... .R.I......
16-2      .......... .......W.. --.F...... .......... .......... .R.I......
15-2      .......... .......W.. --.F...... ........N. .......... .R.I......
16-1      .......... .......G.. --.F.S.... .......... .......... ...I......
9-6       ........T. .M...D.... --WHLE.... .......... .V.......W ...M......
15-1      ....L...H. .......W.. --.F...... .....M.... ......E... ...A......


          61                                  90
consensus YWERETRNAK DTAQTFRVSL GNLRGYYNQS EA
13-2      ..DQ...S.R ........N. ET........ ..
13-1      ..DQ...S.R ........N. ET........ ..
10-1      ..DQ...S.R ........N. ET........ ..
11-5      ....T.GY.. AN.R.D..A. RK.LLR.... ..
11-1      ....T.GY.. AN.R.D..A. RK.LLR.... ..
11-2      ....T.GY.. AN.R.D..A. RK.LLR.... ..
11-6      ....T.GY.. AN.R.D..A. RK.LLR.... ..
11-4      ....T.GY.. TN.R.D..A. RK.LLR.... ..
9-7       ..D.N...S. V....H.... .......... K.
9-2       ..D.N...S. V....H.... .......... K.
9-10      ..D.N...S. V....H.... .......... K.
9-4       ..D.N...S. V....H.... .......... K.
9-5       ..D.N...S. V....H.... .......... K.
9-1       ..D.N...S. V....H.... .......... K.
16-6      ...E...K.. .A..S...G. .I........ ..
15-5      ...E...K.. .A..S...G. .I........ ..
16-2      ...E...K.. .A..S...G. .I........ ..
15-2      ...E...K.. .A..S...G. .I........ ..
16-1      ...EA..I.. AR...D..D. .T........ ..
9-6       ...E...R.. AN.....GN. RTALR..... ..
15-1      ...EQ..I.. GN..AH.GN. R..LR..... .G
```

α 2

```
          91
150
consensus GSHTYQWMYG CDLGPDGRLL RGYHQFAYDG KDYIALNEDL RSWTAADTAA QNTQRKWEAA
13-2      ....L...H. ....S...F. ...E...... ...LT..... ...S.V.... .ISEQ.SNDG
13-1      ....L...H. ....S...F. ...E...... ...LT..... ...S.V.... .ISEQ.SNDG
10-1      ....L...H. ....S...F. ...E...... ...LT..... ...S.V.... .ISEQ.SNDG
11-5      ....L.G.N. ..M....... .....H.... ....S..... ........V. RI...FY..E
11-1      ....L.G.N. ..M....... .....H.... ....S..... ........V. RI...FY..E
11-2      ....L.G.N. ..M....... .....H.... ....S..... ........V. RI...FY..E
11-6      ....L.G.N. ..M....... .....H.... ....S..... ........V. RI...FY..E
11-4      ....L.G.N. ..M....... .....H.... ....S..... ........V. RI...FY..E
9-7       ......R... .......... ...Y.Y.... .......... .......... ....H.....
9-2       ......R... .......... ...Y.Y.... .......... .......... ....H.....
9-10      ......R... .......... ...Y.Y.... .......... .......... ....H.....
9-4       ......R... .......... ...Y.Y.... .......... .......... ....H.....
9-5       ......R... .......... ...Y.Y.... .......... .......... ....H.....
9-1       ......R... .......... ...Y.Y.... .......... .......... ....H.....
16-6      ........V. .NV....H.. .......... ...V...K.. .....G.M.. .........V
15-5      ........V. .NV....H.. .......... .......K.. .....G.M.. .........V
16-2      ........V. .NV....H.. .......... .......K.. .....G.M.. .........V
15-2      ........V. .NV....H.. .......... .......K.. .....G.M.. .........V
16-1      ....I.T... .......... ...R.D.... .......... .......M.. .........GD
9-6       ....L..... .......... .......... .......... .......M.. ..........
15-1      ....V.T... ..V....... ...E...... R......... .......M.. ..........
```

```
        151                                 182
consensus GEAEQFRAYL EGTCVEWLRR YLENGKETLQ RA
13-2      S...HQ.... .D......H. .......... .S
13-1      S...HQ.... .D......H. .......... .S
10-1      S...HQ.... .D......H. .......... .S
11-5      EY..E..T.. ..E.L.L... .......... ..
11-1      EY..E..T.. ..E.L.L... .......... ..
11-2      EY..E..T.. ..E.L.L... .......... ..
11-6      EY..E..T.. ..E.L.L... .......... ..
11-4      EY..E..T.. ..E.L.L... .......... ..
9-7       .V...K.... .......... H......... ..
9-2       .V...K.... .......... H......... ..
9-10      .V...K.... .......... H......... ..
9-4       .V...K.... .......... H......... ..
9-5       .V...K.... .......... H......... ..
9-1       .V...K.... .......... H......... ..
16-6      ....R....V ..R....... .......... ..
15-5      ....R....V ..R....... .......... ..
16-2      ....R....V ..R....... .......... ..
15-2      ....R....V ..R....... .......... ..
16-1      RY..R..... ..R....P.. .......... ..
9-6       RA...W..Q. ..K....... .......... ..
15-1      RA..KD.... ..P....... .......... ..
```

α 3

```
        183
242
consensus DPPKTHVTHH PVSDHEATLR CWALGFYPAE ITLTWQRDGE DQTQDTELVE TRPAGDGTFQ
13-2      E........R .......... .......... ..V....... .......... ..........
13-1      E......... .......... .......... ..V....... .......... ..........
10-1      E......... .......... .......... ..V....... .........L. ..........
11-5      ....A..A.. HI..R..... .......... .......... E......... ..........
11-1      ....A..A.. HI..R..... .......... .......... E......... ..........
11-2      ....A..A.. HI..R..... .......... .......... E......... ..........
11-6      ....A..A.. HI..R..... .......... .......... E......... ..........
11-4      ....A..A.. .I..R..... .......... .......E.. E......... ..........
9-7       .......... .......... ..V....... .......... .......... ...V......
9-2       .......... .......... ..V....... .......... .......... ...V......
9-10      .......... .......... ..V....... .......... .......... ...V......
9-4       .......... .......... ..V....... .......... .......... ...V......
9-5       .......... .......... ..V....... .......... ....S..... ...V......
9-1       .......... .......... ..V....... .......... .......... ...V.N....
16-6      .......... .......... .......... .......... .......... ...G......
15-5      .......... .......... .......... .......... .......... ...G......
16-2      .......... .......... .......... .......... .......... ...G......
15-2      .......... .......... .......... .......... .......... ...G......
16-1      .......... .......... .......... .......... E......... ...G......
9-6       .......... .......... .......... .......... E......... ...V......
15-1      .......... .......... .......... .......... .......... ...G......
```

```
       243                              274
consensus KWGAVVVPSG EEQRYTCHVQ HEGLPEPLTL RW
13-2      ..A....... .......... .......RA. ..
13-1      ..A....... .......... .......RA. ..
10-1      ..A....... .......... ....A..... ..
11-5      ..A....... .......... .....Q.... ..
11-1      ..A....... .......... .....Q.... ..
11-2      ..A....... .......... .....Q.... ..
11-6      ..A....... .......... .....Q.... ..
11-4      ..A....... .......... .....Q.... ..
9-7       .......... .......... .Q........ ..
9-2       .......... .......... .Q........ ..
9-10      .......... .......... .Q........ ..
9-4       .......... .......... .Q...G.... ..
9-5       .......... .......... .Q........ ..
9-1       .......... .......... .Q........ ..
16-6      .......... .......... .......... ..
15-5      .......... .......... .......... ..
16-2      .......... .......... .......... ..
15-2      .......... .......... .......... ..
16-1      .......... .......... .......... ..
9-6       .......... .......... .Q........ ..
15-1      .......... .......... .......... ..
```

Transmembrane domain

```
        275                                        313
consensus EPSSQSTIPI VGIVAGLAVL AVVVTGAVVA AVMWRRKSS
13-2      ....R..... ...I...VL. GA.......V .........
13-1      ....R..... ...I...VL. GA.......V .........
10-1      .......... ...I...VL. GA.......V .........
11-5      .S...P.... .......... .......... .........
11-1      .S...P.... .......... .......... .........
11-2      .S...P.V.. .......... .......... .........
11-6      .S...P.... .......... .......... .........
11-4      .S...P.... .......... .......... ....K....
9-7       ......A... .......... .......... .........
9-2       .......... .......... .......... .........
9-10      .......... .......... .......... .........
9-4       .......... .......... .......... .........
9-5       .......... .......... .......... .........
9-1       .......... .......... .......... .........
16-6      .......... ....V..... .......... .........
15-5      .......... ....V..... .......... .........
16-2      .......... ....V..... .......... ........P
15-2      .......... ....V..... .......... .........
16-1      .......... ....V..... .......... .........
9-6       .......... .......... .......... .........
15-1      .......... ....V..... .......... .........
```

Cytoplasmic domain

```
        314                          340
consensus GGKGGSYSQA ASSDSAQGSD VSLTACK*
13-2      .R........ SC...T.... E......*
13-1      .R........ SC...T.... E......*
10-1      .R........ LC...T.... E......*
11-5      DRNR.....P T*
11-1      DRNR.....P T*
11-2      DRNR.....P T*
11-6      DRNR.....P T*
11-4      DRNR.....P T*
9-7       .......... .......... .....*
9-2       .......... .......... .....*
9-10      .......... .......... .....*
9-4       .......... .......... .....*
9-5       .......... .......... .....*
9-1       .......... .......... .....*
16-6      .......... .......... .....*
15-5      .......... .......... .....*
16-2      .......... .......... .....*
15-2      .......... .......... .....*
16-1      .......... ..N....... .....*
9-6       .......... .......... .....*
15-1      ...A...... ..N....... .....*
```

figure 6. Alignment of predicted amino acid sequences of *M. fascicularis* MHC class I molecules. The sequences are compared to a consensus sequence derived from the data. The Mafa-E locus molecules all contain a two amino acid size polymorphism in the α 1 domain. The polymorphism is indicated using bold font for the positions. The numbering of the sequences was adjusted to keep the positions of the domains consistent with similar research. The gaps are indicated using dashes (-) and the stop codons are indicated by an asterisk.

The MSA was broken apart into fragments demarcating the domains of the MHC class I

molecule. The protein sequences in the MSA were used to construct a phylogenetic tree of the sequences. In order to determine the putative locus of each sequence, each molecule was compared against known sequences using BLAST. Once the prospective loci of the molecules had been determined, representative sequences of the loci from closely related sequences were incorporated into the phylogenetic tree (figure 7).

The B-locus alleles in similar species have been shown to be very diverse in the peptide fragments that are bound. Because of this, it was necessary to incorporate a larger representative group of B-locus alleles for comparisons between the *M. fascicularis* molecules and molecules from other similar species (figure 8).

The B and F binding pockets in the antigen binding site tend to be the most diverse regions of the molecule in regards to their amino acid composition. The amino acids contributing to the aforementioned binding pockets were extracted from the amino acid sequences derived for each B-locus molecule (Table 4). The variability of the regions is very apparent in the table for both the B and the F pockets.

Figure 7. Phylogenetic tree of of MHC class I sequences from Macaques.

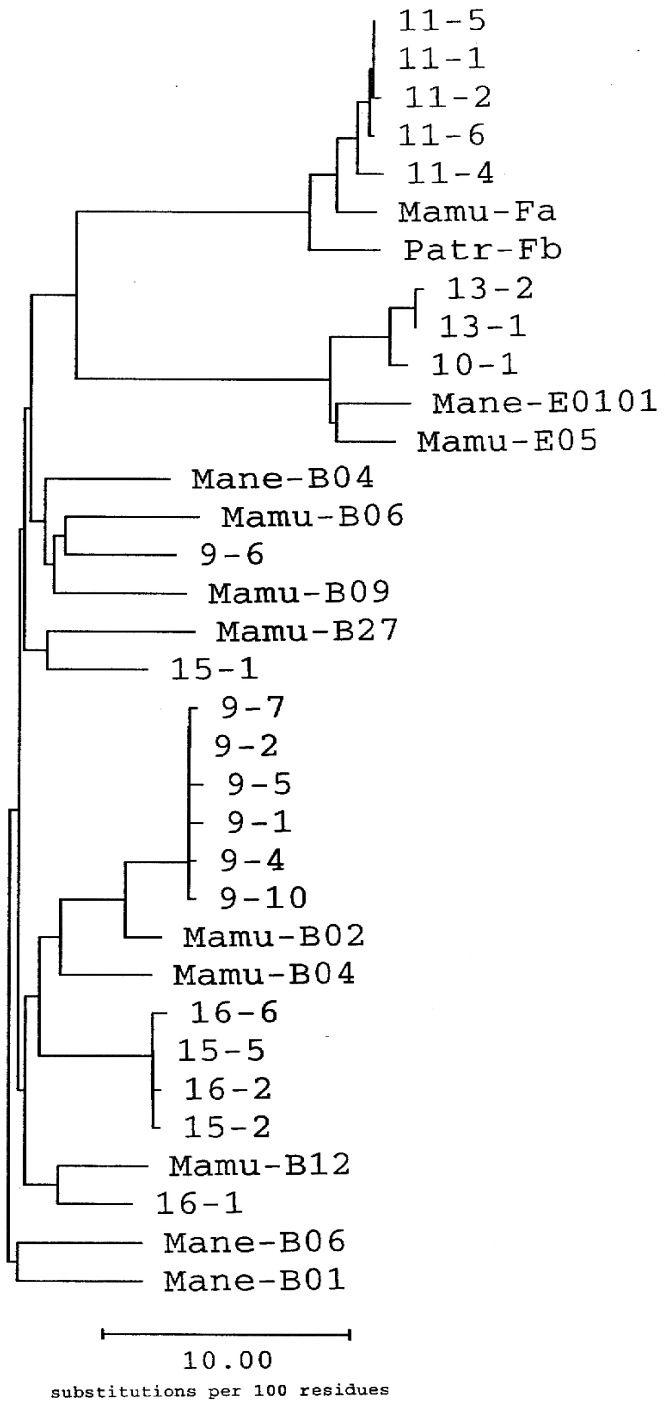


figure 7. Phylogenetic analysis of the amino acid sequences derived from the MHC class I molecules for *M. fascicularis*. Mafa, Mane, and Patr(Pan troglodytes) molecules were used to identify the prospective locus for the sequences from *M. fascicularis*.

Figure 8. Phylogenetic tree of B locus MHC sequences from Macaques

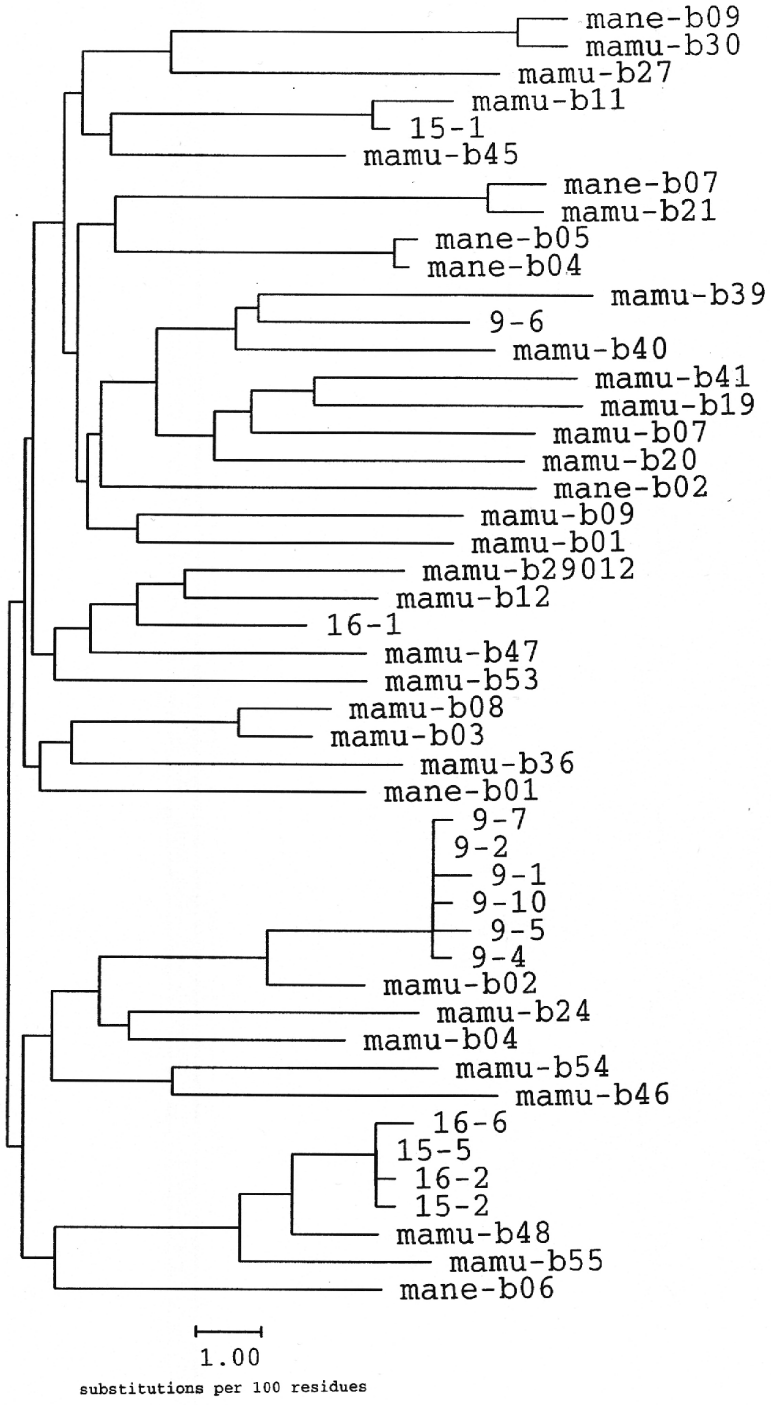


figure 8. Phylogenetic analysis of putative B-locus alleles from *M. fascicularis* compared to B-locus alleles from *M. mulatta (*Mamu) and *M. nemestrina (*mane). Tree was constructed using the Neighbor joining method with uncorrected distances.

Table 4. B and F binding pockets for B-locus alleles from *M. fascicularis*

	B Pocket positions											F Pocket positions								
Protein	7	9	24	25	34	45	63	66	67	70	99	77	80	81	84	95	116	123	143	146 147
9-2	Y	S	E	V	V	M	N	N	S	T	Y	S	N	L	Y	Y	Y	Y	T	K W
16-1	-	-	S	-	-	-	A	I	A	R	-	D	T	-	-	I	D	-	-	- -
16-2	-	-	A	-	-	-	E	K	A	A	V	G	I	-	-	-	F	-	-	- -
9-6	-	T	-	-	-	-	E	R	A	N	-	N	T	A	-	L	F	-	-	- -
15-1	-	H	A	-	M	E	Q	I	A	N	-	N	-	-	-	V	F	-	-	- -

Table 4: The B and F pockets of the binding grove in the B-locus MHC Class I alleles from *M. fascicularis*. The positions contributing to the pockets were obtained from similar analysis done on closely related species(8). Dashes (-) indicate conservation of amino acid present in 9-2.

The same position extraction was done with the E and F-locus molecules (Table 5).

Table 5. B and F binding pockets for E and F locus alleles from *M. fascicularis*

	B Pocket positions											F Pocket positions									
Protein	7	9	24	25	34	45	63	66	67	70	99	77	80	81	84	95	116	123	143	146	147
10-1	Y	H	S	V	V	M	E	S	A	T	H	N	T	L	Y	L	F	Y	S	K	S
Mamu-E05	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mane-E0201	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Mane-E0501	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11-1	Y	S	Y	I	Q	P	E	T	G	K	G	R	L	R	L	H	H	K	R	Q	R
Mamu-F	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table 5: The B and F pockets of the binding grove in the E and F locus MHC Class I alleles from *M. fascicularis* and other Macaques. The positions contributing to the pockets were obtained from similar analysis done on closely related species(8). Dashes (-) indicate identity.

While the B and F pockets from the B-locus alleles have a large number of differences in amino acids, the aforementioned pockets in the E- and F-locus alleles are completely conserved, not only within *M. fascicularis*, but across species as well.

It has been established that in humans there is overdominant selection in the classical

MHC Class I molecules (10, 1). Extensive use of non-synonymous and synonymous substitution rates have been used to analyze the 57 amino acid positions that are involved with the recognition of processed foreign antigens (10). Similar analysis of the hominoid MHC class I molecules have supported the hypothesis of overdominant selection for these amino acid positions (9). It is also important to note that in most eukaryotic genes studied, the number of nucleotide substitutions between polymorphic alleles is very small, usually between 0.0001 and 0.02 per nucleotide position. Also, most of these substitutions are synonymous, resulting in no change in amino acid sequence. Therefore, the substitution rate pattern observed is unique to polymorphic alleles in the MHC loci (10). While it was suspected that a similar pattern for the non-synonymous and synonymous substitution rates would be observed in *M. fascicularis*, there has not been any such analysis to date. Therefore, the substitution rates for *M. fascicularis* have been calculated to verify that it is indeed the case that the substitution rate pattern mimics that seen in hominoids (Table 6).

Table 6. Synonymous (ds) and non-synonymous (dn) substitution rates in different regions of MHC Class I alleles

	Antigen Binding site[a] (54 codons)		alpha 1 and 2 domains[b] (128 codons)		alpha 3 domain (92 codons)	
	ds	dn	ds	dn	ds	dn
	[c]					
Mafa-B	12.6 ± 4.2	23.8 ± 3.9	6.1 ± 1.5	3.4 ± 0.7	4.8 ± 2.0	1.2 ± 0.5
Mafa-E	0.0 ± 0.0	0.0 ± 0.0	0.4 ± 0.4	0.4 ± 0.2	0.0 ± 0.0	0.4 ± 0.3
Mafa-F	0.0 ± 0.0	0.0 ± 0.0	0.7 ± 0.7	0.0 ± 0.0	1.0 ± 1.0	1.6 ± 0.7
Mamu-B (36)[d]	14.2 ± 3.2	26.5 ± 4.1	7.6 ± 1.5	4.3 ± 0.7	8.4 ± 2.0	1.9 ± 0.5
Mane-B (18)[d]	13.5 ± 3.3	26.6 ± 4.1	7.8 ± 1.3	4.7 ± 0.7	9.2 ± 1.8	2.0 ± 0.4
HLA-A[e]	6.2 ± 2.6	13.7 ± 2.3	2.9 ± 1.3	1.8 ± 0.5	9.4 ± 3.0	1.5 ± 0.6
Gogo-A[e]	5.4 ± 2.6	14.8 ± 2.3	8.2 ± 2.1	3.5 ± 0.7	2.8 ± 1.5	2.3 ± 0.7
Patr-A[e]	0.7 ± 1.0	7.2 ± 1.6	2.7 ± 1.3	0.3 ± 0.2	2.8 ± 1.6	0.6 ± 0.4
Hominoid-A[e]	4.8 ± 1.6	14.5 ± 1.4	5.8 ± 1.1	2.1 ± 0.4	7.9 ± 2.0	2.1 ± 0.6

Table 6: Synonymous and non-synonymous nucleotide substitution rates for selected codons contributing to the predicted peptide.

[a] 54 codons in the α1 and α2 domains encoding positions that comprise the antigen binding site (8).

[b] The 128 positions of the α1 and α2 domains that do not contribute to the antigen binding site.

[c] Jukes-Cantor estimate of mean nucleotide substitutions and standard errors presented as percentages.

[d] The number of Samples used in the calculations of the Synonymous and Non-synonymous substitution rates for the specified locus.

[e] Results from similar analysis done using hominoid A locus alleles(9).

The substitution rates were calculated for *M. fascicularis* sequences for both the classical class I MHC molecules (B-locus) and the non-classical class I MHC Molecules (loci E and F). Because the nonclassical MHC molecules bind a very restricted set of peptides, it is expected that the non-synonymous substitution rates would be close to if not lower than the synonymous substitution rates. For both the E and F locus alleles, we see no substitutions that effect the antigen binding site. It is important to note that there is a very low substitution rate for the remainder of the α 1 and α 2 domains as well as the α 3 domains.

In order to facilitate the usage of the extraction program for the B and F binding programs, the perl script was converted into a web form (figure 10). By using this format, the program is available to a much larger user base than the original program. The results (figure 11) of the extraction are returned to the user in a table listing the amino acid position and the amino acid that is at that position in the sequence that was entered.

Further analysis of the B and F binding pockets for the B-locus alleles was carried out by using the Neighbor Joining method to determine the percent divergence of the molecules. The resulting distances were used to create a phylogenetic tree representation of the data (figure 9).

Based on the comparisons done to determine the segregation patterns of the HLA-B alleles in relation to the MHC-B in macaques, it was shown that HLA-B*27 and 15-1 clone were the most closely related clone sequence and HLA molecule. Because of this, comparisons between the B and F binding pockets for the HLA-B*27 molecules and the 15-1 clone were carried out (figure 12).

Figure 9. Phylogenetic tree of B and F binding pocket sequences from Macaques

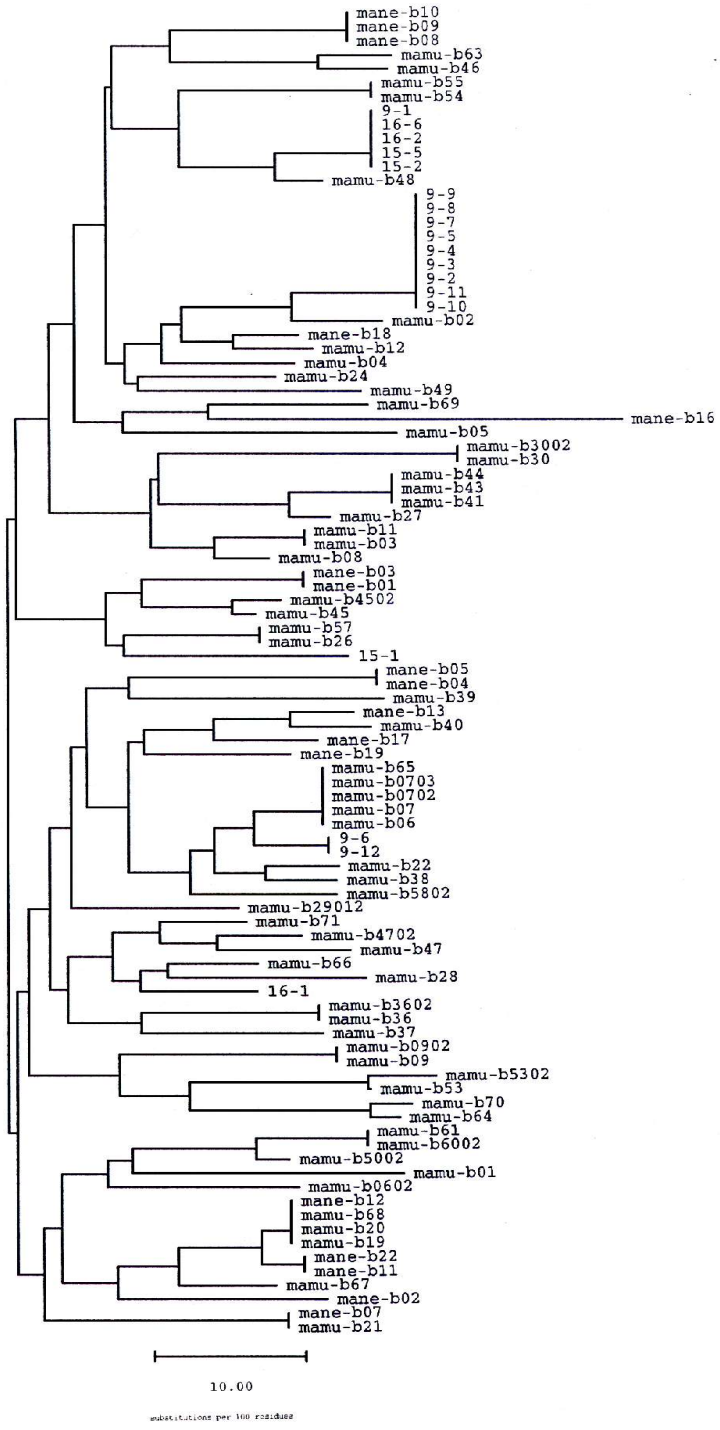


Figure 9. phylogenetic representation of the percent divergence between the B and F peptide-binding pockets for the MHC class I B-locus molecules in Macaques. Analysis was done using the Neighbor-Joining method.

Figure 10. Pocket Extractor Query

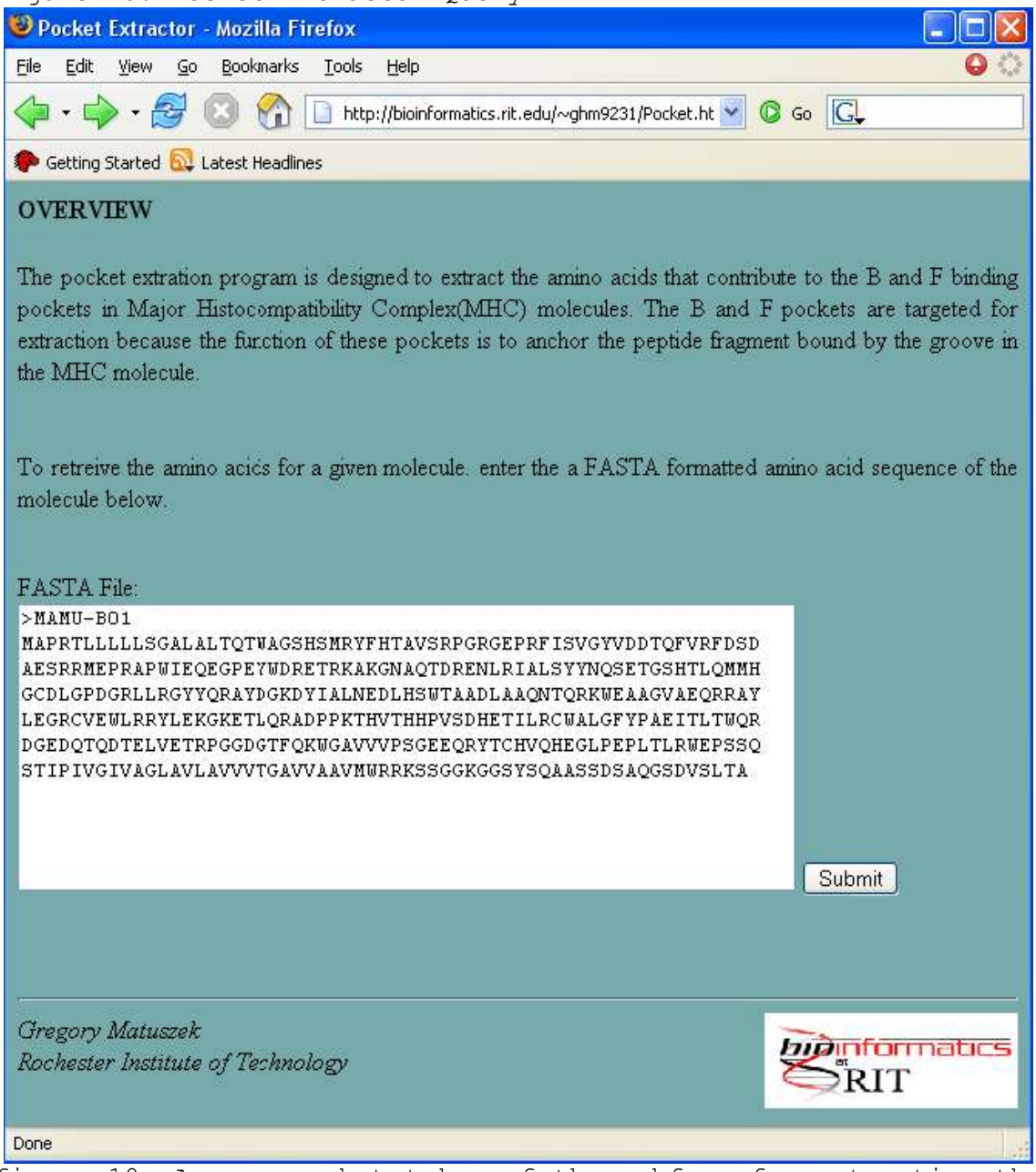


figure 10. A screen shot taken of the webform for extracting the B and F binding pockets from an MHC class I molecule.

Figure 11 Pocket Extractor Results

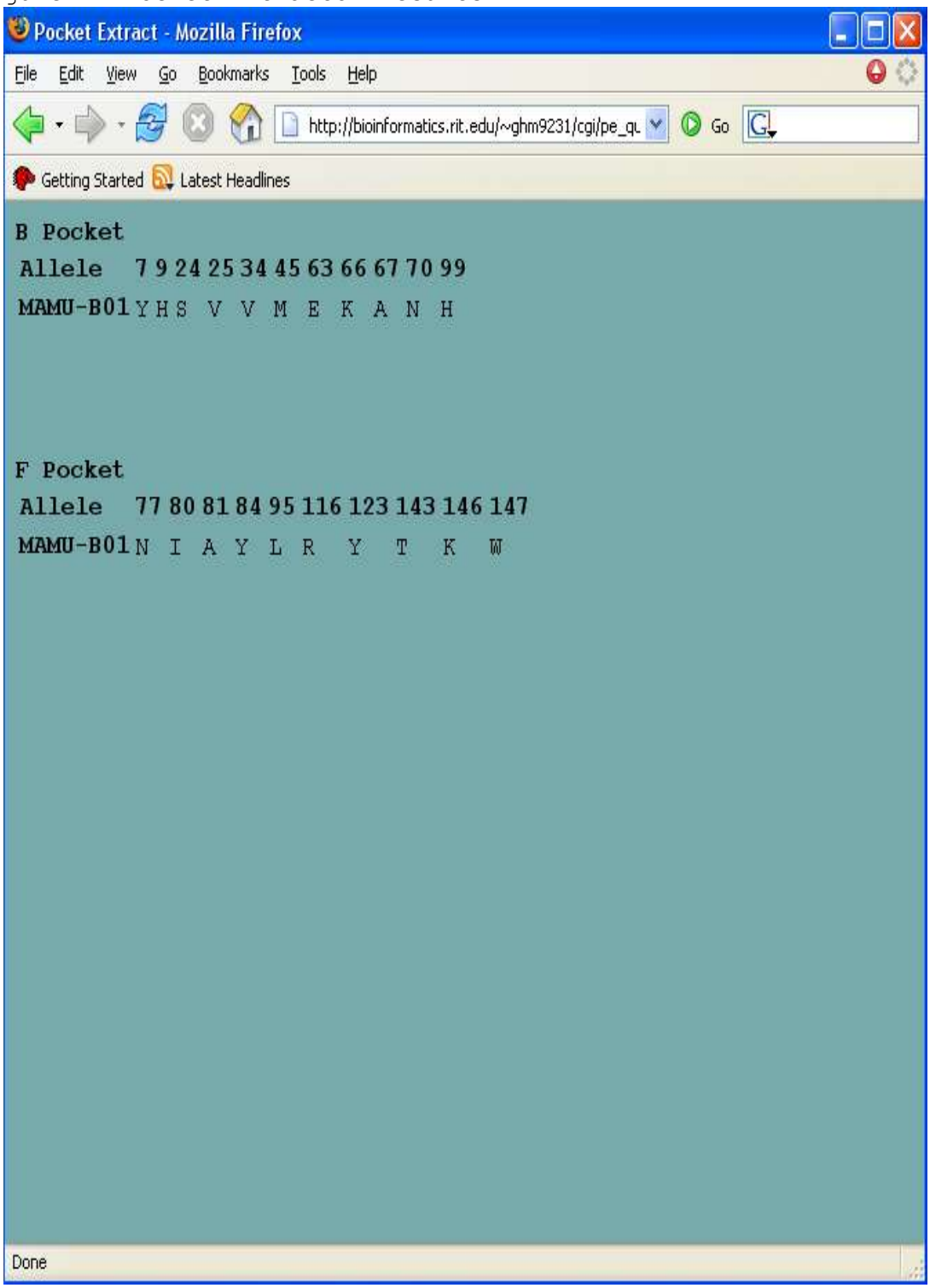


figure 11. Results generated by the webform using the Mamu-B01 protein as an example sequence.

figure 12. MSA of B and F binding pockets for HLA-B*27 alleles and 15-1

B Binding Pocket

	7	9	24	25	34	45	63	66	67	70	99
consensus	Y	H	T	V	V	E	E	I	C	K	Y
hla-b2713	-	-	-	-	-	-	-	-	-	-	-
hla-b2712	-	-	-	-	-	-	-	-	-	N	-
hla-b2709	-	-	-	-	-	-	-	-	-	-	-
hla-b2708	-	-	-	-	-	-	-	-	-	-	-
hla-b270505	-	-	-	-	-	-	-	-	-	-	-
hla-b270502	-	-	-	-	-	-	-	-	-	-	-
hla-b2704	-	-	-	-	-	-	-	-	-	-	-
hla-b2703	-	-	-	-	-	-	-	-	-	-	-
hla-b2702	-	-	-	-	-	-	-	-	-	-	-
hla-b2706	-	-	-	-	-	-	-	-	-	-	-
hla-b2711	-	-	-	-	-	-	-	-	-	-	-
15-1	-	-	A	-	M	-	Q	-	A	N	-

F Binding Pocket

	77	80	81	84	95	116	123	143	146	147
consensus	-	T	L	Y	L	D	Y	T	K	W
hla-b2713	D	-	-	-	-	-	-	-	-	-
hla-b2712	S	N	-	-	-	-	-	-	-	-
hla-b2709	D	-	-	-	-	H	-	-	-	-
hla-b2708	S	N	-	-	-	-	-	-	-	-
hla-b270505	D	-	-	-	-	-	-	-	-	-
hla-b270502	D	-	-	-	-	-	-	-	-	-
hla-b2704	S	-	-	-	-	-	-	-	-	-
hla-b2703	D	-	-	-	-	-	-	-	-	-
hla-b2702	N	I	A	-	-	-	-	-	-	-
hla-b2706	S	-	-	-	-	Y	-	-	-	-
hla-b2711	S	-	-	-	-	Y	-	-	-	-
15-1	N	N	-	-	V	F	-	-	-	-

figure 12. Multiple sequence alignment of the B and F binding pocket positions for HLA-B*27 alleles and sequence 9-6. Dashes (-) indicate agreement with consensus sequence for the given amino acid position.

Because of the desire to use *M. fascicularis* as an animal model for HIV research, the peptide fragment prediction using MAPPP was carried out using the GP160 protein molecule found in SIV. The HLA molecules used by the program for prediction were limited to a subset that were both HLA-B, and contained very similar B and F pocket information with the Macaque B-locus molecules. The resulting output from MAPPP is contained in Appendix 1.

Discussion

The sequences that were realized from *M. fascicularis* correspond to three different genes. The majority of the sequences obtained group very closely with B-locus alleles obtained from closely related species, while the remainder segregate either into the E or F locus.

When comparing the substitution rates observed for the non-classical MHC molecules to those of the classical MHC molecules, there is a drastic difference in substitution rates, especially in the positions that contribute to the antigen binding site. For the *M. fascicularis* B locus alleles, the non-synonymous substitution rate is much greater than the rate for synonymous substitutions. However, when observing the remainder of the α1 and α2 domains, this is not the case. In the remaining positions, the non-synonymous substitution rate is much lower than the synonymous substitution rate. This pattern in *M. fascicularis* is consistent with what is observed for the B locus in both *M. mulatta* and *M. nemestrina.*

When comparing the substitution rates for the Macaque B locus to those observed in the hominoid A locus, there is a marked difference in the percentages for the substitutions in general for the antigen binding site positions. The percent of non-synonymous mutations for the antigen binding sites in Macaques is between 23.8% and 26.6%, while in hominoid A locus sequences, the maximum is 14.8%, almost half the value of that observed in the Macaques. The Hominoid A locus for the class I MHC has been well characterized and has been shown to be quite polymorphic. The pattern observed in the Macaque B locus would seem to indicate that the B

locus for Macaques is experiencing greater diversification as a result of the greater percentage of non-synonymous substitution rates.

While the positions contributing to the antigen binding site in the Macaque B locus molecules exhibits a large amount of nucleotide substitutions in general, the overall rate of substitution for the remainder of the α1 and α2 domains as well as the α3 domain is drastically lower than that observed for the antigen binding site positions. This supports the contention that there is positive selection for non-synonymous substitutions at the antigen binding site positions. Also, the presence of positive selection at the observed sites indicates that they do in fact contribute to the binding pockets in *M. fascicularis* as well as human MHC class I MHC molecules.

When the B-locus portion of the tree was expanded (figure 8) to incorporate a much larger number of sequences from related species, the sequences derived from *M. fascicularis* segregate into 5 different prospective genes. Some of the clades of sequences have slight divergence from another on the order of 0.5 amino acid differences per 100 positions. Because the MHC class I alleles are approximately 350 amino acids in length, this equates to less than two amino acid differences over the entire length of the sequence, including the leader peptide. Thus, it is a fairly safe assertion that the sequences within the clade are all alleles of a gene. The segregation pattern five genes observed in *M. fascicularis* indicates that the duplications observed in the *M. mulatta* exist in the *M. fascicularis* as well. However, verification of this through inheritance patterns within a family of *M. fascicularis* would be necessary for conclusive evidence of gene duplication occurring in *M. fascicularis*.

From the information in table 6, it is apparent that the positions of the binding pockets are subject to positive selection for substitutions. This is most readily apparent in the B and F pockets of the binding groove. The pockets in the B-locus molecules have a large number of amino acid differences between sequences from within a species (Table 1). While a small number of amino acid positions are conserved throughout all of the B-locus molecules, the

majority of the positions are highly variable. This characteristic however is not present when looking at the amino acid positions for the B and F pockets from molecules of the E and F loci. The molecules of both the E and F loci are not only completely conserved within all of the sequences derived from *M. fascicularis*, but are completely conserved when compared to E and F locus molecules from *M. mulatta* and *M. nemestrina*. The marked difference in the variability of amino acids contributing to the B and F binding pockets indicates that the E and F loci bind a much more restricted set of peptide fragments. The variability in the B-locus molecules could have come about as result of two very different paths. As was shown in figure 4, the B locus in *M. mulatta* has had marked expansion and duplication to the point where there are as many as 19 paralogous genes.

Therefore, the variability in the B locus could be a result of the gene duplication, with the gene itself being fairly well conserved, but with differences in the paralogous genes. The alternative is that while there are multiple B-like genes, the variability in the molecules is a result of nucleotide substitution. While both of these are possible explanations, given that many of the 19 molecules are truncated or non-functional and the high rate of non-synonymous substitutions, it is much more probable that the variability is a result of both the duplication, and positive selection for substitutions.

After verification of the positions that contribute to the binding groove, analysis was carried out on the B and F pockets of the binding groove. The B and F pockets were chosen for analysis because they are responsible for anchoring the peptide fragment in the groove so that it can be presented to TCR. In the phylogenetic tree (figure 9) generated using the amino acids contributing to the pockets, one very diverse group was found. The upper threshold in percent divergence when restricted to the amino acid positions of the B and F peptide-binding pockets is ~57%.

Based on the results shown in figure 12, the 15-1 clone is very similar to the HLA-B*2702 molecule. While there are several amino acid positions that differ between the two sequences,

almost all of these differences are conservative amino acid substitutions. Based on the similarities of between these two molecules, these molecules are likely to bind similar sets of peptide fragments.

Taking into consideration the goal of making *M. fascicularis* a viable animal model for HIV and other infectious disease research, the gp160 molecule from the HIV virus was used in an attempt to predict possible peptide fragments that would be bound by the MHC molecules in *M. fascicularis.* Therefore, using the HLA-BB*2702 molecules that are available in MAPPP, a fairly relevant set of peptide fragments have been returned by the MAPPP program (Appendix 1). Each of the predicted sequences have a score based on how well they bind to by the given molecule. With the results generated by the MAPPP program using the GP160 viral peptide sequence and the HLA-B*2702 molecule, the list of peptide fragments that would be likely to be bound by the molecule will help guide further analysis and provide a list of prospective peptide fragments that could be used in a vaccine for animals with the 9-6 allele.

1. Lawlor, D. A., J. Zemmour, P. D. Ennis, and P. Parham. 1990. Evolution of Class-I MHC Genes and Proteins: From Natural Selection to Thymic Selection. *Annu. Rev. Immunol.* 8:23-63.
2. Vogel, T. U., D. T. Evans, J. A. Urvater, D. H. O'Connor, A. L. Hughes, and D. I Watkins. 1999. Major histocompatibility complex class I genes in primates: co-evolution with pathogens. *Immunol. Rev.* 167:327-337.
3. Daza-Vamenta, R., G. Glusman, L. Rowen, B. Guthrie, and D. E. Geraghty. 2004. Genetic Divergence of the Rhesus Macaque Major Histocompatibility Complex. *Genome Research.* 1501-1515.
4. Holzhutter, H. G., C. Frommel, and P. M. Kloetzel. 1999. A Theoretical Approach Towards the Identification of Cleavage-determining Amino Acid Motifs of the 20 S Proteasome. *J. Mol. Biol.* 286:1251-1265.
5. Kuttler, C., A. K. Nussbaum, T. P. Dick, H. G. Rammensee, H. Schild, and K. P. Hedeler. 2000. An Algorithm for the Prediction of Proteasomal Cleavages. *J. Mol. Biol.* 298:417-429.
6. Boyson, J. E., C. Shufflebotham, L. F. Cadavid, J. A. Urvater, L. A. Knapp, A. L. Hughes, D. I. Watkins. 1996. The MHC Class I Genes of the Rhesus Monkey. *The Journal of Immunology.* 4656-4665.
7. Mothe, B. E., J. Sidney, J. L. Dzuris, M. E. Liebl, S. Fuenger, D. I. Watkins, and A. Sette. 2002. Characterization of the Peptide-Binding Specificity of Mamu-B*17 and Identification of Mamu-B*17-Restricted epitopes derived from Simian Immunodeficiency Virus Proteins. *Journal of Immunology*. 210-219.
8. Lafont, B. A. P., A. Buckler-White, R. Plishka, C. Buckler, and M. A. Martin. 2003. Characterization of Pig-Tailed Macaque Classical MHC Class I Genes: Implications for MHC Evolution and Antigen Presentation in Macaques. Journal of Immunology. 171: 875-885.
9. Fernandez N. and G. Butcher. 1998. MHC Volume 2: A Practical Approach. 153-179.
10. Hughes, A. L., M. Nei. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*. 167-170.
11. Kumar S, Tamura K, Jakobsen IB, & Nei M (2001)
MEGA2: Molecular Evolutionary Genetics Analysis Software.
Bioinformatics Vol. 17, 12:1244-1245
12. Jörg Hakenberg, Alexander Nussbaum, Hansjörg Schild, Hans-Georg Rammensee, Christina Kuttler, Hermann-Georg Holzhütter, Peter-M. Kloetzel, Stefan H.E. Kaufmann, and Hans-Joachim Mollenkopf. 2003. MHC-I Antigenic Peptide Processing Prediction. *Applied Bioinformatics*. 2(3): 155-158.
13. Kuttler, Christina, A. K. Nussbaum, T. P. Dick, H.G. Rammensee, H. Schild, and K.P. Hedeler. 2000. An Algorithm for Prediction of Proteasomal Cleavages. *Journal of Molecular Biology.* 298: 417-429.
14. Holzhutter, H.G., C. Frommel, P.M. Kloetzel. 1999. A Theoretical Approach Towards the Identification of Cleavage-determining Amino Acid Motifs of the 20S Proteasome. *Journal of Molecular Biology.* 286: 1251-1265.
15. Rammensee, H.G., J. Buchmann, N.P.N. Emmerich, O.A. Bachor, S. Stevanovic. 1999. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*. 50:213-219.
16. Adams, E.J., P. Parham. 2001. Species-specific evolution of MHC class I genes in higher primates. *Immunological Reviews*. 183:41-64.
17. Saitou, M. and M. Nei. 1987. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. *Molecular Biology Evolution*. 4:406-425
18. Kimura, M. 1980. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16: 111-120.