PharmaSUG 2016 - Paper DG11

Displaying data from NetMHCIIPan using GMAP: the SAS System as a Bioinformatics Tool

Kevin R. Viel, Ph.D., Histonis, Incorporated; Atlanta, GA

ABSTRACT

The binding of peptides to HLA Class II molecules is a seminal event in adaptive immunology. From vaccine development to immunogenicity, the prediction of binding affinity (strength) is extremely important information. The NetMHCIIpan 3.1 Server provides estimates of the binding affinity for specific peptides and given HLA molecules. The goal of this paper is to describe the basics of immunology with respect to HLA, to describe how to obtain estimated binding affinity interactively using NetMHCIIpan, and how to present these data using the SAS system using the GMAP procedure complimented by the ANNOTATE facility.

INTRODUCTION

To survive the onslaught of microbes, such as bacteria, archaea, viruses, molds, fungi, and protozoans, the body requires effective defenses. The defenses can be simple like barriers, such as the skin, but also include a system of pathways and mechanisms far more complex than that of the coagulation system. Perusing the Human Microbiome Project (http://hmpdacc.org/) created by the National Institutes of Health (NIH) one can begin to appreciate the vastness of one's microbiome. Indeed, some estimates suggest that within one's body, the ratio of bacteria to human cells is 10:1. These microbes can be symbiotic, commensal, or parasitic. The latter class is the proper target of the immune system.

The Human Leukocyte Antigen (HLA) or Major Histocompatibility Complex (MHC) are a cluster of genes that are central to the immune responses and code for a transmembrane proteins that, among the phases of its lifecycle, are ligands for antigen-specific T-cells. They are divided into two classes with a distinguishing feature that Class I has a peptide binding groove that is closed on the ends, limiting the size of the bound peptide to approximately nine amino acids (AAs) in length, also known as a nonamer or 9mer (9 AA polymer, since a peptide is a polymer of amino acids). If a T-cell with an appropriate receptor encounters a peptide-HLA (pHLA) complex, then it may bind it and initiate an immune response of the adaptive immune system. Knowing whether a HLA molecule will bind a given peptide is a major focus in fields such as infectious diseases or the development of vaccine and biologics. The goal of this paper is to briefly describe HLA-II and their role in immune responses, to introduce the NetMHCIIpan 3.1 Server[1], and to detail a program that uses the GMAP procedure and ANNOTATE facility of the SAS System® to create map of estimated binding strengths of peptides that span a protein.

THE HLA (MHC) SYSTEM

Located on Chromosome 6, the HLA genes occupy approximately 4 Mb of DNA. These DNA variants in these genes are among the most abundant in the genomes, especially with regard to functional variants. This density results in the diversity needed to react against the vast microbiome, whose survival depends on overcoming or evading the immune system, resulting in continual evolution.

Macrophages, B-cells, and dendritic cells (DC) are professional antigen presenting cells (APC) that specialize in uptake of proteins, the processing of them to peptides, and presentation of these peptides to other cells, such as T-cells. This paper will focus on DCs, which are quite efficient at activating antigen-specific CD4+ T-cells (the host cell for HIV). The term antigen derives from antibody generating because a T-cell activated by an appropriate encounter with it cognate antigen presented by an APC may activate or "help" a B-cell, which then produces antigen specific antibodies. This class of T-cells is known as Helper T-cells. Cluster of differentiation (CD) are protein receptors in the membranes of cells that were first identified by immunophenotyping protocols. The distribution of such receptors define the cell type, although hundreds of CD have been identified and cells have numerous CD, cells may be named by the "predominate" CD. Other important classes of T-cells are the CD8+ Cytotoxic T-cell (CTL), the CD17+ helper T-cell (Th17), or CD4+CD25+ regulator T-cell (Treg). Helper T-cells are further classified by their predominant cytokine profiles: Th1 and Th2 cells. Th2 cells are responsible for strong antibody responses, hyper-reactive episode of which can result in disorders like asthma.

An immature DC is reminiscent of a snow flake or starfish with is branched dendrites (protrusions) spreading out to its environment. Like T-cells and B-cells, DCs originate from the haematopoietic stem cells of the bone marrow. The immature DCs migrate between the blood, lymph, and peripheral tissues, like the skin. When located where microbes or antigens might enter the body, the DC samples its environment, taking up proteins and other material through endocytosis, pinocytosis, or phagocytosis. The antigen laden DC then migrate to the draining lymph nodes,

where they mature. The DC processes the proteins in proteolytic compartments, endosomes, phagosomes, or lysosomes, which also contain HLA-II molecules, which the immature DC constitutively expresses.

HLA-II molecules are a heterodimer of α and β subunits. The peptide bound in the HLA-II groove may have flanking residues (AAs) that both interact with the HLA-II molecule to stabilize the binding and that may be proteolyzed (cleaved) through the life cycle of the peptide/HLA-II complex (pHLA-II). Considering the HLA-DR isotype (genes in a family) as an example, HLA-DRA codes for α subunit and HLA-DRB1 or its paralogs (distinct genes thought to have arisen by duplication) DRB3, DRB4 or DRB5 code the β subunit of a HLA-DR molecule. Importantly, variants have not been reported in the loci that code for the binding groove of HLA-DRA, unlike those for HLA-DPA and HLA-DQA.

In the endoplasmic reticulum (ER), the newly formed MHC-II associates with the Invariant chain (Ii), which occupies its binding groove with flanks extending out of the groove, and may traffic to the cell membrane or, via the trans-Golgi network, directly to a mature endosome. From the cell membrane, the HLA-II/Invariant chain complex may be endocytosed thus forming an early endosome in the cytoplasm. The internal environment of these compartments becomes increasingly acidic as the endosome matures promoting the processing by endopeptidases of the Invariant chain until it is just the Class II-associated invariant chain peptide (CLIP), an approximately 20 AA peptide (20mer). With its grooved occupied by CLIP, the HLA-II cannot bind peptides also produced in this compartment until HLA-DM interacts with it to induce a conformation change that results in the release of CLIP. When the now free groove of the HLA-II encounters peptides, whether derived from self or foreign proteins, it may bind them with varying affinity (strength).

A peptide/HLA-II complex (pHLA-II) of sufficient stability will then traffic to the cell membrane. Importantly, the pHLA-II generated in the same compartment will traffic together and cluster in lipid rafts or microdomains, 5-100 nanometer regions rich in cholesterol, sphingolipids, and proteins such as pHLA-II that have lifetimes of less than 10 milliseconds[2]. Although a typical DC might display the tens of thousands of HLA-II, no more than 500 appropriate pHLA-II might active antigen-specific T-cells[3]. That is, the local clustering of a very small fraction of the total pool of all HLA-II increases the probability of an immunologic synapse between a DC and naïve T-cells. Disruption in vitro of the lipid rafts in DC loaded with small amounts of antigens suppresses the activation of co-cultured naïve T-cells, but this effect could be overcome by an increased load of antigen.

The process of proteolysis may be stochastic in the very least due to random sample of its environment the DC may obtain, but also because of the (random) variation in the distribution of proteases, such as the cathepsins. Binding competition for the HLA-II grooves may not be limited to peptides derived from other peptides, but, potentially, also those derived from the same protein, which obviously are internalized together if the protein is intact. Whether peptides are liberated (proteolyze) from a given protein is an important question outside the scope of this paper, but the relevant and important question is whether the groove of a HLA-II molecule can bind the peptide. To obtain the empirical data to establish this is prohibitively resource intensive. Our group previously calculated that 951 proteins in what we termed the coagulation biosystem contained 350,570 distinct 10mers[4], far too many to test among the hundreds of HLA-II alleles. To approach this question, several groups have created algorithms to estimate the binding strength of a given pHLA-II. This paper focuses on the NetMHCIIpan server[1].

NETMHCIIPAN 3.0 SERVER

"The NetMHCIIpan method is based on an ensemble of artificial neural networks trained on quantitative peptide binding data covering multiple MHC class II molecules."[1] An interactive version of the NetMHCIIpan 3.0 Server is available to the public through the kindness of the Center for Biological Sequence of Analysis of the Technical University of Denmark.

The user has a choice of the format of the sequence(s) that he or she wishes to submit. This talk will focus on the Fasta format. The Instructions tab includes a description of the Peptide format. Figure 1 illustrates two possible submissions in the Fasta format. Each entry starts with a header on the first line that begins with ">" in the first column. The header text should begin immediately in the second column without a blank space. The widths of the lines are typically 60 or 80 bytes. Those readers who might desire longer lines should verify that this will not cause an error or truncation in the tools that they might use. In Figure 1b, each sequence starts with a header and can span multiple lines.

The sequence in Figure 1a is the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/) reference sequence for full-length (FL) coagulation Factor VIII (FVIII) for release hg19. Deficiency or absent of FVIII activity (FVIII:C) causes Hemophilia A, a bleeding disorder that affects mostly males since the structure gene, *F8*, is on the Chromosome X. The lengths of the lines in Figure 1a are 60 bytes. The sequences in Figure 1b represent AA substitutions caused by single nucleotide variants (SNPs).

1a.

```
>FVIII_FL
MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPPRVPKSFPFN
TSVVYKKTLFVEFTDHLFNIAKPRPPWMGLLGPTIQAEVYDTVVITLKNMASHPVSLHAV
GVSYWKASEGAEYDDQTSQREKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLTYSYLSH
VDLVKDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVFDEGKSWHSETKNSLMQDRD
AASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPEVHSIFLEGHTFLVRNH
RQASLEISPITFLTAQTLLMDLGQFLLFCHISSHQHDGMEAYVKVDSCPEEPQLRMKNNE
EAEDYDDDLTDSEMDVVRFDDDNSPSFIQIRSVAKKHPKTWVHYIAAEEEDWDYAPLVLA
PDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGDTL
LIIFKNQASRPYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGP
TKSDPRCLTRYYSSFVNMERDLASGLIGPLLICYKESVDQRGNQIMSDKRNVILFSVFDE
NRSWYLTENIQRFLPNPAGVQLEDPEFQASNIMHSINGYVFDSLQLSVCLHEVAYWYILS
IGAQTDFLSVFFSGYTFKHKMVYEDTLTLFPFSGETVFMSMENPGLWILGCHNSDFRNRG
MTALLKVSSCDKNTGDYYEDSYEDISAYLLSKNNAIEPRSFSQNSRHPSTRQKQFNATTI
PENDIEKTDPWFAHRTPMPKIQNVSSSDLLMLLRQSPTPHGLSLSDLQEAKYETFSDDPS
PGAIDSNNSLSEMTHFRPQLHHSGDMVFTPESGLQLRLNEKLGTTAATELKKLDFKVSST
SNNLISTIPSDNLAAGTDNTSSLGPPSMPVHYDSQLDTTLFGKKSSPLTESGGPLSLSEE
NNDSKLLESGLMNSQESSWGKNVSSTESGRLFKGKRAHGPALLTKDNALFKVSISLLKTN
KTSNNSATNRKTHIDGPSLLIENSPSVWQNILESDTEFKKVTPLIHDRMLMDKNATALRL
NHMSNKTTSSKNMEMVQQKKEGPIPPDAQNPDMSFFKMLFLPESARWIQRTHGKNSLNSG
QGPSPKQLVSLGPEKSVEGQNFLSEKNKVVVGKGEFTKDVGLKEMVFPSSRNLFLTNLDN
LHENNTHNQEKKIQEEIEKKETLIQENVVLPQIHTVTGTKNFMKNLFLLSTRQNVEGSYD
GAYAPVLQDFRSLNDSTNRTKKHTAHFSKKGEEENLEGLGNQTKQIVEKYACTTRISPNT
SQQNFVTQRSKRALKQFRLPLEETELEKRIIVDDTSTQWSKNMKHLTPSTLTQIDYNEKE
KGAITQSPLSDCLTRSHSIPQANRSPLPIAKVSSFPSIRPIYLTRVLFQDNSSHLPAASY
RKKDSGVQESSHFLQGAKKNNLSLAILTLEMTGDQREVGSLGTSATNSVTYKKVENTVLP
KPDLPKTSGKVELLPKVHIYQKDLFPTETSNGSPGHLDLVEGSLLQGTEGAIKWNEANRP
GKVPFLRVATESSAKTPSKLLDPLAWDNHYGTQIPKEEWKSQEKSPEKTAFKKKDTILSL
NACESNHAIAAINEGQNKPEIEVTWAKQGRTERLCSQNPPVLKRHQREITRTTLQSDQEE
IDYDDTISVEMKKEDFDIYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSSPHVLRNR
AQSGSVPQFKKVVFQEFTDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVTFRNQASR
PYSFYSSLISYEEDQRQGAEPRKNFVKPNETKTYFWKVQHHMAPTKDEFDCKAWAYFSDV
DLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDETKSWYFTENMERNCRA
PCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHSIHFSGH
VFTVRKKEEYKMALYNLYPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKC
QTPLGMASGHIRDFQITASGQYGQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMII
HGIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIFN
PPIIARYIRLHPTHYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFA
TWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLLTSMYVKEFL
ISSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRM
EVLGCEAQDLY
```

1b.

```
>FVIII_0061_0077
TSVVYKKTLFVEFTDHL
>FVIII_0090_0106
LLGPTIQAQVYDTVVIT
>FVIII_0100_0116
YDTVVITLTNMASHPVS
>FVIII_0113_0129
HPVSLHAVDVSYWKASE
>FVIII_0136_0152
QTSQREKEVDKVFPGGS
>FVIII_0149_0165
PGGSHTYVCQVLKENGP
>FVIII_0161_0177
KENGPMASNPLCLTYSY
>FVIII_0166_0182
MASDPLCLNYSYLSHVD
```

Figure 1. Examples of Fasta files submitted to NetMHCIIPan

The header of the first sequence in Figure 1b, states the gene (it could have been the protein product, but they were generated from the DNA sequence), then the start position, and the end position:

> FVIII_0100_0116

Position 100 is the first column of the second row of the sequence (highlighted in pink in the figure). Position 56 is in column 17 of the second row. The sequences highlighted in Figure 1a and 1b mismatched at 9th AA, that is position 108 of FVIII. Conventionally, we denote this as K(0108)T. The user has the choice to copy-and-paste (or type) the sequences or upload a file. The user can specify Peptide lengths of 9-21 (or greater).

The sequences in Figure 1b were designed to be submitted with Peptide length 9. When given a sequence, NetMHCIIPan will generate each possible Xmer with the start AA shifted by 1, where X is Peptide length. For example, for the first peptide listed in Figure 1b, the overlapping peptides would be:

```
YDTVVITLTNMASHPVS
YDTVVITLTNMASHPVS
YDTVVITLTNMASHPVS
YDTVVITLTNMASHPVS
YDTVVITLTNMASHPVS
YDTVVITLTNMASHPVS
YDTVVITLTNMASHPVS
YDTVVITLTNMASHPVS
YDTVVITLTNMASHPVS
```

Figure 2. The nine 9mers that overlap by eight AA in the 17mer FVIII_0100_0116.

The sequences in Figure 1b were generated for each AA variant found in a re-sequencing study of Hemophilia A patients (only a subset is shown). Only nine 9mers contain the given AA variant; starting with the peptide with AA 108 in the ninth (last) position to the peptide with AA 108 in the first position. In these patients, no sub-sequence of

FVIII of length 18 or less had more than one AA variant, simplifying the combinations needed to be generated. For greater Peptide lengths, more of the flanking sequence will have to be included. For instance, for Peptide length 15, the variant would have to be location at position 15 of the sequence and the entire sequence would be length 29 (15 + 15 - 1). Since the variant must occupy each position of peptide, the number of peptides of length 15 is, obviously, 15.

NetMHCIIpan includes the HLA-II isotypes HLA-DR, HLA-DP and HLA-DQ. The user can enter a custom sequence, too, but this paper will not cover that options. As mentioned, the HLA-DRA locus is not known to have AA variants in the region of the binding groove, so choosing alleles for it is not necessary or possible. The typing of HLA is 4-digit typing, reflecting, in part, the order in which they were identified. Up to twenty alleles can be selected for a given run, however, providing long sequences, too many sequences, or too many alleles given the first two issues can result in a the process timing out and not returning results. When the user selects HLA-DP or HLA-DQ, then he or she must choose both the A and B genes. Note that if the user generates the list, NetMHCIIPan may report that a specific combination is not possible (observed), so care must be taken. The author has not had a reason, beyond curiosity of using non-default values for the remaining option, but the reader is welcome to share her or his experience. Upon submitting, the web page automatic changes. For long jobs or to ensure access to the result in case the browser closes, the author recommends that user submits her or his email. The results appear to be available for at least several days after a submission. Figure 3 shows the results for sequence FVIII_0100_0116 in Figure 1b.

NetMHCIIpan Server - prediction results

Technical University of Denmark

```
# NetMHCIIpan version 3.1

# Input is in FASTA format

# Peptide length 9

# Threshold for Strong binding peptides (IC50)  50.000 nM
# Threshold for Weak binding peptides (IC50)    500.000 nM

# Threshold for Strong binding peptides (%Rank) 0.5%
# Threshold for Weak binding peptides (%Rank)   2%

# Allele: DRB1_0101
```

Seq	Allele	Peptide	Identity	Pos	Core	Core_Rel	1-log50k(aff)	Affinity(nM)	%Rank	Exp_Bind	BindingLevel
0	DRB1_0101	YDTVVITLT	FVIII_0100_0116	0	YDTVVITLT	0.700	0.234	3969.06	12.00	NA	
1	DRB1_0101	DTVVITLTN	FVIII_0100_0116	0	DTVVITLTN	0.700	0.115	14431.21	49.00	NA	
2	DRB1_0101	TVVITLTNM	FVIII_0100_0116	0	TVVITLTNM	0.700	0.175	7497.66	25.00	NA	
3	DRB1_0101	VVITLTNMA	FVIII_0100_0116	0	VVITLTNMA	0.700	0.365	958.38	2.00	NA	<=WB
4	DRB1_0101	VITLTNMAS	FVIII_0100_0116	0	VITLTNMAS	0.700	0.406	618.66	1.20	NA	<=WB
5	DRB1_0101	ITLTNMASH	FVIII_0100_0116	0	ITLTNMASH	0.700	0.202	5591.37	18.00	NA	
6	DRB1_0101	TLTNMASHP	FVIII_0100_0116	0	TLTNMASHP	0.700	0.131	12146.68	42.00	NA	
7	DRB1_0101	LTNMASHPV	FVIII_0100_0116	0	LTNMASHPV	0.700	0.433	459.40	0.80	NA	<=WB
8	DRB1_0101	TNMASHPVS	FVIII_0100_0116	0	TNMASHPVS	0.700	0.111	15031.55	55.00	NA	

Core_Histogram

```
Number of strong binders: 0 Number of weak binders: 3
```

Explain the output. Go back.

Figure 3. The results from NetMHCIIPan of running the sequence YDTVVITLTNMASHPVS with Peptide length 9 for DRB1*01:01.

The first column is the position of the sequence peptide within the sequence, with an origin at 0 instead of 1. In bioinformatics tools, this is not uncommon, for instance the UCSC Genome Browser[5] employs the same system. The report will be grouped by the HLA-II allele, but it also appears in column 2, which makes reading the table as raw data easier. The 9mer peptide being evaluated appears in the third column and they match the bold, underlined sequences in Figure 2. The Identity in column 4 matches the Fasta header. POS in column 5 is the position of the binding core register (core), that is the peptide that actually occupies the groove. Since the binding groove holds a 9mer, when the Peptide length is 9, the Peptide and Core are the same. For brevity, the final column of interest is the BindingLevel, which will be a strong (<=SB) or weak (<=WB) binder based on criteria for % Rank (column 9) or binding affinity (IC50, column 10). The former is the rank of predicted affinity of the test peptide compare to the predicted affinity of a set of 200,000 random natural peptides. If one is unfamiliar with chemistry, then the affinity can be thought of a ratio of the product of the concentration of the peptide and the concentration of HLA-II to the

concentration of pHLA-II in a closed system. A strong affinity implies more pHLA-II is observed, whereas when the peptide and HLA-II are found apart, it implies a weaker affinity. As the concentrations have the units Molar (moles per liter), the units describing

$$\text{Affinity} = [\text{ peptide }][\text{ HLA-II }] / [\text{ pHLA-II }]$$

is Molar, with typical magnitude being nanomolar (nM). Figure 4 is an example of a plot produced by the program in this paper plots the BindingLevel against Seq.

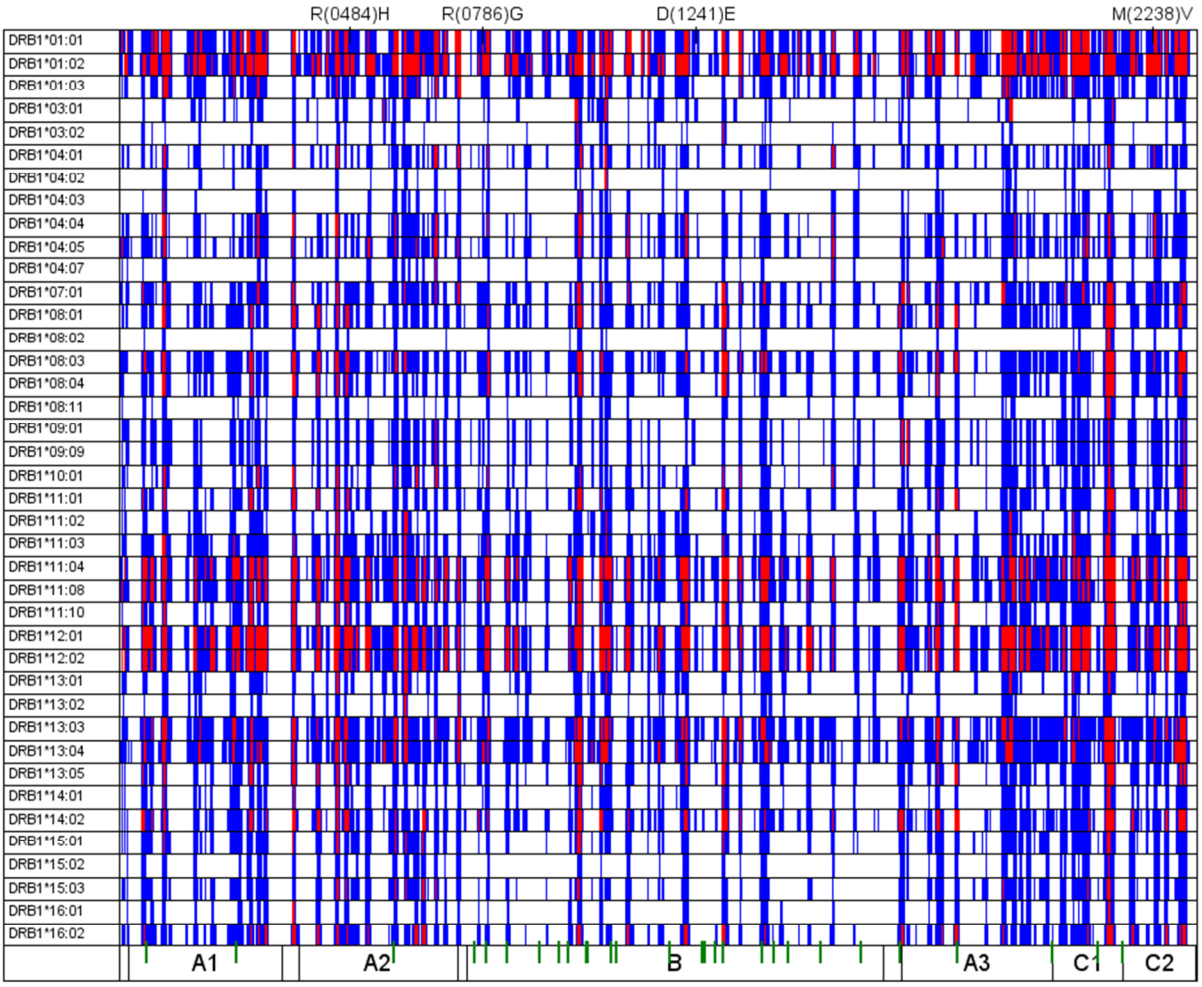

Figure 4. A peptide binding map. A "map" of the binding strength of 18mer derived from coagulation Factor VIII (FVIII) for 40 DRB1 alleles found in patients in a re-sequencing project that investigated the associations with inhibitors, neutralizing alloantibodies against transfused FVIII. **Red** bars indicate Strong binder, whereas **blue** bars indicate weak binding. The bars indicate the position in FVIII of the 18mers, thus contiguous positions have overlapping sequences and, for the eight peptides immediately flanking the position of interest could possibly share the same core. The domains of FVIII are shown at the bottom, with acidic regions a1, a2, and a3 are not labeled. The 19 AA pro-peptide is to the left of A1. The **green** lines are the putative N-Linked glycosylation sites. On the top are the non-synonymous single nucleotide polymorphisms (ns-SNPs) that are found in non-Hemophilia A patients as well[6] and investigated as risk factors for inhibitors[7].

This map of the peptides provides an immense amount of information. A brief background on FVIII will be helpful; Lenting et alia wrote an excellent overview of FVIII[8]. The 2,351 AA protein has a 19 AA leader peptide, which can be seen on the left of the plot before the A1 "domain"; the mature FVIII protein has 2,332 AA. Historically, the regions of FVIII are called domains, but they are from the linear transcript of *F8*. In a more traditional sense, a domain would be from a separate locus (gene). A processing cleavage occurs near the C-terminal region of the B domain, and secreted FVIII circulates and is stored as a heterodimer. Activation cleavages occur in the acidic regions that are not labeled in this figure. One can see that most potential N-linked glycosylation sites occur in the B domain. Upon

activation, for instance, by activated Factor IX (FIXa) or thrombin (activated Factor II), the B domain is freed and FVIII is a heterotrimer of the A1, A2, and A3-C1-C2. The consequence of this is that some AA acids that are adjacent may be in separate peptides when FVIII is internalized by APCs. The exception might possibly be the promiscuous expression[9] by medulla thymic endothelial cells (mTEC), which, combined with full-length FVIII liberated by apoptosis, might be a source of the full-length, linear FVIII. Finally, the amino acid substitutions caused non-synonymous single nucleotide polymorphisms (ns-SNPs) that our group reported[6,7] are listed on the top of the figure.

Figure 4 presents the strength of the estimated binding affinity, the number of peptides expected to bind, the origins of the peptides within FVIII, features such as ns-SNPs that affect the identity of the peptide and glycosylation opens questions as to whether protein-protein interactions or proteolysis might be affect whether peptides might be available to the HLA grooves. The question may not likely be just whether a peptide will bind, but whether it might be liberated in sufficient quantities. To complicate issues, glycosylation is stochastic process. A protein for which the sites of glycosylation and the pattern of glycosylation is known is described as a glycoform; a person has thousands of glycoforms of FVIII, a typical glycoprotein. This figure is a great tool for considering issues like immunogenicity and the potential that protein might stimulate an immune response, a necessity in vaccine development. A few issues remain that are also essential. The first is that the identity of the core peptide not reported in this figure. The second is that peptide length is certainly related to binding affinity. For a small sample of proteins, the figure has been sparse for nonamers. Figure 5 presents the binding map of FVIII for nonamers. Note that nonamers are identical to the core.

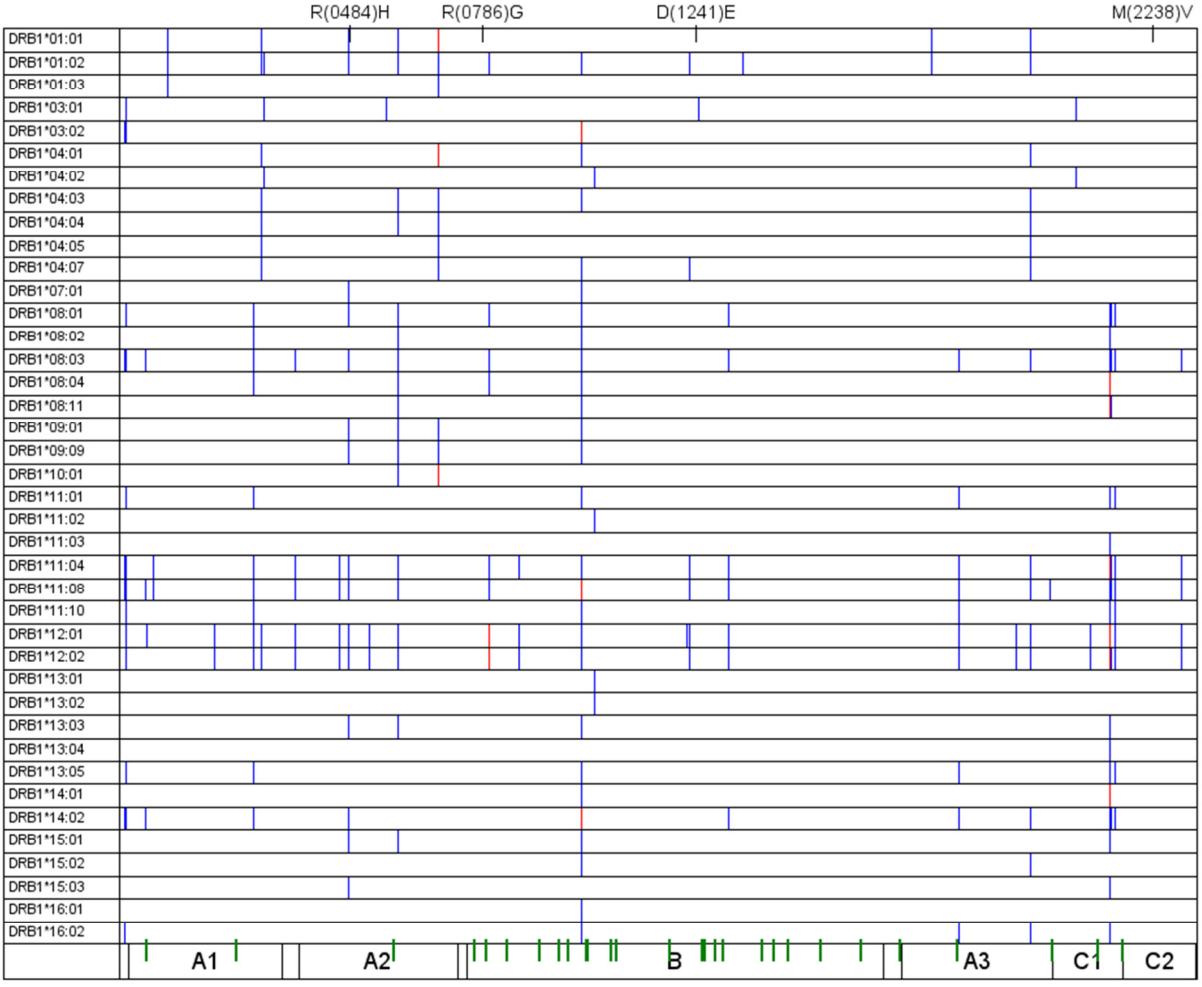

Figure 5. The Peptide Binding Map of FVIII for Nonamers.

The third is that it only illustrates intra-protein competition for binding sites; but many proteins may be internalized with the target protein. For instance, albumin is in great excess over FVIII, but it may not be in the focal volume of active coagulation. Even so, for FVIII, one might expect at least one to one stoichiometry with von Willebrand Factor, which binds FVIII and is essential in platelet interactions. FVIIIa and FIXa, the tenase complex, bind Factor X on the surface of activate platelets. Shepherd et alia[10] expanded this concept to the whole (reference) genome, but

incorporation of the microbiome is an obvious, though technically and logistically challenging, required refinement to such in silico investigations.

How might these maps guide investigations? Consider DRB1*04:01 and compare it to DRB1*01:02 or DRB1*12:01. From Figure 4 one might suspect that a patient with DRB1*04:01 will not bind peptides from an infused product that matches the reference sequence. Thus, even if that patient lacks FVIII sequence in that region, the exposure to this exogenous might not illicit an immune response. The results in Figure 5, however, might suggest that if the length of the peptide that illicit a response is nine, then DRB1*04:01 might be comparable to DRB1*01:02, but DRB1*12:01 still seems to be susceptible. Patients may not match the reference sequence and plasma-derived or recombinant FVIII also may not match the reference sequence. If the sequence of the infused product is known and mismatches the references sequence, then the binding of that peptide could be of interest. The resourced need to thoroughly examine this in investigation might be currently prohibitive, but this map might help guide the threading of the needle.

CODE EXPLAINED: ANNOTATE

nsSNPS AND DOMAINS

Code is included to add detail to the map for nsSNPs, glycosylation, and the domain structure via the ANNOTATE facility. The latter is pertinent only to Factor VIII, but other similar detail could be added, such epitopes or regions that are surface exposed. Lines 1-54 of Figure 6 create the data sets for the nsSNPs and the domain structure, which are hard coded values, and should be understandable to most readers, but two issues merit comment. The first is that, to be general, the program does not consider the leader peptide, so the labels for AA substitutions are numbered with respect to the mature protein (offset by -19, as described above). Secondly, the input data set Seq.F8 is annotated as described by Viel[11]. The SQL procedure in Lines 55-65 demarcates the boundaries of the domains.

```
 1  Data ns_SNP_NNs ;
 2    *allele = "R(0503)H" ;
 3    allele = "R(0484)H" ;
 4    NN =  61620 ;
 5    AA = 503 ;
 6    Output ;
 7    *allele = "R(0795)G" ;
 8    allele = "R(0786)G" ;
 9    NN =  91317 ;
10    AA = 795 ;
11    Output ;
12    *allele = "D(1260)E" ;
13    allele = "D(1241)E" ;
14    NN =  92714 ;
15    AA = 1260 ;
16    Output ;
17    *allele = "M(2257)V" ;
18    allele = "M(2238)V" ;
19    NN = 162161 ;
20    AA = 2257 ;
21    Output ;
22  Run ;
23
24
25  /***********/
26  /* Domains */
27  /***********/
28  Data __Domains
29          ( Keep   = AA
30                     Domain
31          )
32       ;
33    Set Seq.F8
34           ( Keep  = CDS_Nucleotide
35                     Location
36                     Codonic_Nucleotide
37             Where = (     Location           =: "Exon"
38                       and Codonic_Nucleotide = 1
```

```
39                      )
40             )
41          ;
42
43     AA = Ceil(( CDS_Nucleotide / 3 )) ;
44     If 1 <= AA - 19 <= 336 Then Domain = "A1" ;
45      Else If 337 <= AA - 19 <= 372 Then Domain = "a1" ;
46       Else If 373 <= AA - 19 <= 719 Then Domain = "A2" ;
47        Else If 720 <= AA - 19 <= 740 Then Domain = "a2" ;
48         Else If 741 <= AA - 19 <= 1648 Then Domain = "B" ;
49           Else If 1649 <= AA - 19 <= 1689 Then Domain = "a3" ;
50            Else If 1690 <= AA - 19 <= 2019 Then Domain = "A3" ;
51             Else If 2020 <= AA - 19 <= 2172 Then Domain = "C1" ;
52              Else If 2173 <= AA - 19 <= 2332 Then Domain = "C2" ;
53   Run ;
54
55   Proc SQL ;
56     Create Table Domains as
57     Select *
58     From __Domains
59     where Domain ne ""
60     Group By Domain
61     Having     AA = Min( AA )
62            or  AA = Max( AA )
63     Order By AA
64     ;
65   Quit ;
```

Figure 6. Creating Annotate Data Sets for nsSNPs and Domains.

CODE EXPLAINED: ANNOTATE

GYLCOSYLATION

The third ancillary data set is the location of the putative N-lined glycosylation sites within the FVIII. N-linked glycosylation is a covalent bond to asparagine (N) and may occur when the pattern of asparagine (N), not proline (P), and followed by serine (S), threonine (T), or cysteine (C). In terms of regular expressions, this is "N[^P](?:S|T|C)". The ability to use the SAS data step with the PRXNEXT() function will be limited to protein of length 32,767 AA (bytes). Simple alternative exists, some outside of the SAS System, but mapping larger proteins may be questionable with regard to concise figures.

Figure 7 presents the SQL procedure and data step that creates the glycosylation data step. To remain efficient, the SQL procedure in Lines 1-5 determines the length of the protein. Note that the last codon is the termination signal (stop codon) so we subtract one. Using the INTO clause, the SQL procedure creates the macro variable PROTEIN_LENGTH. The SEPARATED BY " " effectively strips the leading and trailing spaces, which can be accomplished by TRIMMED.

The annotated gene data set labels each nucleotide with its codon number. Ending the loop at when LAST.CODON is 1 (Line 21) assures that CODON_SEQ has three nucleotides (Line 31). The format $C2AA. simply maps the three nucleotide codon to the amino acid or termination (stop) signal (Line 31), which is appending via the CATS() function (Line 37) and CODON_SEQ is reset to missing (Line 36). Line 44 creates the Regular Expression ID used in the CALL PRXNEXT() call routine. The non-overlapping three amino acid sequons and the position of the arginine are output for use in the ANNOTATE data set.

```
 1  Proc SQL NoPrint ;
 2    Select Put( Max( Codon ) - 1 , 8. ) Into : Protein_Length Separated By " "
 3    From Seq.F8
 4    ;
 5  Quit ;
 6
 7  Data FVIII_Glyc
 8         ( Keep = Sequon
 9                  Position
10         ) ;
11
```

```
12     Length Sequon       $ 3
13            Position     8
14            FVIII        $ &Protein_Length.
15            Codon        8
16            WT_AA        $ 1
17            Codon_Seq    $ 3
18            ;
19
20     Do Until ( End ) ;
21        Do Until ( Last.Codon ) ;
22
23          Set Seq.F8
24                ( Where = ( Codon ne . )
25                  Keep  = Base Codon
26                )
27                End = End
28              ;
29          By Codon ;
30
31          Codon_Seq = CatS( Codon_Seq , Base ) ;
32        End ;
33
34        WT_AA  = Put( Codon_Seq , $C2AA. ) ;
35
36        Codon_Seq   = "" ;
37        FVIII = CatS( FVIII , WT_AA ) ;
38
39     End ;
40
41     /**********************/
42     /* N-linked Variants  */
43     /**********************/
44     __RC_Parse = PRXParse( CatS( "/N[^P](S|T|C)/o" )) ;
45
46     Start = 1 ;
47     Stop  = Length( FVIII ) ;
48
49     Call PRXNext( __RC_Parse
50                 , Start
51                 , Stop
52                 , FVIII
53                 , Position
54                 , Length
55                 ) ;
56
57     Do While ( Position > 0 ) ;
58
59        Sequon = SubStr( FVIII , Position , Length ) ;
60        Call Missing( Sequon_Position
61                    , WT_AA
62                    , Variant_AA
63                    , Codon_Seq
64                    , Variant_Seq
65                    , Codon_Position
66                    , Substitution
67                    ) ;
68        Output ;
69        Call PRXNext( __RC_Parse
70                    , Start
71                    , Stop
72                    , FVIII
73                    , Position
74                    , Length
```

```
75                    ) ;
76     End ;
77
78   Run ;
```

Figure 7. Creating the Annotate Data Sets for Gycosylation.

THE BINDING_MAP MACRO

The Binding_Map macro is not as flexible as it should be. This will occur in the near future in the effort that substitutes the TEMPLATE and SGENDER procedures for the GMAP approach. Figure 8 presents the Bind_Map macro. Lines 1-5 name the macro and with three macro parameters, one of which has a default value. The HLA distinct names of alleles that will be mapped are obtained in Lines 7-13, with their number saved in the macro variable ALLELES. At this time, the user must insure that a reason number of alleles will be mapped.

The GMAP procedure requires a data set that defines the coordinates of the map and one that contains the response for those coordinates, if any. The layout of this map is simple, a row of even height for each of the alleles. The row has "columns" for certain components. The map starts are -501, 500 units before the start of the protein. Each coordinated is named, the ID variable. The names of the alleles will be linked to a distinct number, which will correspond to their rows in the map (Line 68). The SQL procedure in Lines 71-81 links the name of the allele in the data set obtained from NetMHCIIPan to its row (Y) and orders the data by ALLELE and POS, position within the protein, which is the value of X. The response data set is created in Lines 71-81, adding " NB" as a level and inserting a space before "<=WB" so that the binding level collates in the order of increasing strength.

The data set created in Lines 102-138 reserves space on the map for the components of that will be annotated. Lines 145-204 create the annotation for the allele names. Using the coordinates of the map, the ANNOTATE will be a text label appropriately placed, that is with an X value to the left (negative value) of the first AA position of 1 and the Y coordinate corresponding to its row. Lines 206-271 create the annotation data set for the domains. Instead of simply labeling, the annotation includes drawn lines to demarcate the boundaries, making this data set slightly more complicated than the others, but likely still intuitive enough to follow. Lines 273-301 create the annotation data set for the glycosylation, which simple is the text label "|" with the color green. Lines 303-336 create the annotation data set for nsSNPs. Again, these a text label of "|" and the name of the SNP. The distinguishing feature is the position, "5" versus "2", respectively, placing the name above the bar. The annotation complicates the program, but enhances the information immensely. The remaining code is barely noteworthy; the GMAP procedure uses all three options DATA=, MAP=, and ANNO=, with the CHORO statement, to create a rather structurally bland two-dimensional choropleth map of rows with the shading of bars for three levels of binding strength at the X coordinate corresponding the starting position of the peptide of interest within the protein.

```
 1  %macro binding_map
 2     ( length   =
 3     , bind_ds  =
 4     , y_height = 100
 5     ) ;
 6
 7    proc sql ;
 8      create table MHCII_alleles as
 9      select distinct allele
10      from &bind_ds._&length.
11      order by allele descending
12      ;
13    quit ;
14
15    %let alleles = &sqlobs. ;
16
17    data FVIII_MHCII_map
18            ( drop = __: )
19         ;
20
21      length id $ 9 ;
22
23      __x = -501 ;
24      do __y = 1 to &alleles. ;
25        id = catx( "_"
```

```
26                , put( __x          , z4. )
27                , put( __y          , z4. )
28                ) ;
29        x = __x ;
30        y = __y * &y_height. ;
31        output ;
32      end ;
33
34      __y = -150 ;
35      do __x = -501 , %eval( 2 * &Protein_Length. - 1 ) ;
36        id = catx( "_"
37                , put( __x  , z4. )
38                , put( __y  , z4. )
39                ) ;
40        x = __x ;
41        y = __y ;
42        output ;
43      end ;
44
45      do __x = 1 to &Protein_Length. ;
46        do __y = 1 to &alleles. ;
47          id = catx( "_"
48                   , put( 2 * __x - 1 , z4. )
49                   , put( __y         , z4. )
50                   ) ;
51          x = 2 * __x - 1 ;
52          y = ( __y - 1 ) * &y_height. + 1 ;
53          output ;
54          x = 2 * __x ;
55          output ;
56          x = 2 * __x - 1 ;
57          y = __y * &y_height. ;
58          output ;
59          x = 2 * __x ;
60          output ;
61        end ;
62     end ;
63
64    run ;
65
66    data MHCII_alleles ;
67      set MHCII_alleles ;
68      y = _n_ ;
69    run ;
70
71    proc sql ;
72      create table &bind_ds._&length._alleles as
73      select a.*
74           , b.y
75      from &bind_ds._&length. as a
76         , MHCII_alleles        as b
77      where a.allele = b.allele
78      order by a.allele
79             , a.pos
80      ;
81    quit ;
82
83    data FVIII_MHCII_binding
84           ( keep = id
85                    bindinglevel
86           )
87         ;
88
```

```
 89      length id $ 9 ;
 90
 91      set &bind_ds._&length._alleles ;
 92
 93      id = catx( "_"
 94               , put( 2 * ( pos + 1 ) - 1 , z4. )
 95               , put( y                   , z4. )
 96               ) ;
 97      if bindinglevel = "" then bindinglevel = "  NB" ;
 98      if bindinglevel = "<=WB" then bindinglevel = " <=WB" ;
 99
100    run ;
101
102    data FVIII_terminus
103            ( drop = x
104                     y
105            )
106         ;
107
108      if 0 then set FVIII_MHCII_binding ;
109
110      bindinglevel = "  NB" ;
111
112      do y = 1 to &alleles. ;
113        id = catx( "_"
114                 , "-501"
115                 , put( y            , z4. )
116                 ) ;
117        output ;
118      end ;
119
120      do y = 1 to &alleles. ;
121        id = catx( "_"
122                 , "%sysfunc( putn( 2 * &Protein_Length. - 1 , z4. ))"
123                 , put( y            , z4. )
124                 ) ;
125        output ;
126      end ;
127
128      y = -150 ;
129      do x = -501 , %eval( 2 * &Protein_Length. - 1 ) ;
130        id = catx( "_"
131                 , put( x  , z4. )
132                 , put( y  , z4. )
133                 ) ;
134        output ;
135      end ;
136
137      stop ;
138    run ;
139
140    proc append base = FVIII_MHCII_binding
141                data = FVIII_terminus
142                ;
143    run ;
144
145    data anno_alleles
146            ( drop = __:
147                     allele
148            )
149         ;
150
151      retain xsys "2"
```

```
152             ysys "2"
153             line 1
154             color "black"
155             when "a"
156             ;
157
158       length function $ 8
159              text     $ 100
160              ;
161
162       if _n_ = 1
163       then
164         do ;
165            function = "frame" ;
166            output ;
167
168            y = 1 ;
169            x = 0 ;
170            function = "Move" ;
171            output ;
172            y = &alleles. * &y_height. ;
173            function = "Draw" ;
174            output ;
175            y = 0 ;
176            x = -501 ;
177            function = "Move" ;
178            output ;
179            x = &Protein_Length. * 2 - 1 ;
180            function = "Draw" ;
181            output ;
182       end ;
183
184       set MHCII_alleles
185              ( rename = ( y = __y ))
186           ;
187
188       x = -475 ;
189       y = __y * &y_height. ;
190       function = "label" ;
191       text     = strip( allele ) ;
192       position = "F" ;
193       size     = 0.60 ;
194       output ;
195
196       position = " " ;
197       y = __y * &y_height. ;
198       x = -501 ;
199       function = "Move" ;
200       output ;
201       x = &Protein_Length. * 2 - 1 ;
202       function = "Draw" ;
203       output ;
204     run ;
205
206     data anno_domains
207             ( drop = AA
208                      Domain
209                      lag_AA
210             )
211          ;
212
213       retain xsys "2"
214              ysys "2"
```

```
215             line 1
216             color "black"
217             when "a"
218             ;
219
220       length function $ 8
221              text     $ 100
222              ;
223
224       if _n_ = 1
225       then
226         do ;
227            y = 0 ;
228            x = 1 ;
229            function = "Move" ;
230            output ;
231            y = -150 ;
232            function = "Draw" ;
233            output ;
234            x = &Protein_Length. * 2 - 1  ;
235            output ;
236            y = 0 ;
237            output ;
238         end ;
239
240       set Domains ;
241       by Domain
242             notsorted
243          ;
244
245       y = 0 ;
246       x  = 2 * AA  - 1 ;
247       function = "Move" ;
248       output ;
249       if first.Domain
250       then
251         do ;
252            y = -150 ;
253            function = "Draw" ;
254            output ;
255         end ;
256
257       lag_AA = lag( AA ) ;
258
259       if     last.Domain
260          and Domain ~ =: "a"
261       then
262         do ;
263            y        = 0 ;
264            x        = ceil((( 2 * AA - 1 + 2 * lag_AA - 1 ) / 2 )) ;
265            function = "Label" ;
266            text     = Domain ;
267            position = "E" ;
268            output ;
269            position = "" ;
270         end ;
271     run ;
272
273     data anno_glyco
274             ( drop = Sequon
275                      __Position
276             )
277          ;
```

```
278
279       retain xsys "2"
280              ysys "2"
281              line 1
282              color "black"
283              when "a"
284              ;
285
286       length function $ 8
287              text     $ 100
288              ;
289
290       set FVIII_Glyc
291              ( rename = ( Position = __Position ))
292           ;
293
294       y = 0 ;
295       x = 2 * __Position - 1 ;
296       function = "Label" ;
297       text     = "|" ;
298       position = "5" ;
299       color    = "Green" ;
300       size     = 1 ;
301     run ;
302
303     data anno_SNPs
304             ( drop = AA
305                      NN
306                      Allele
307             )
308          ;
309
310       retain xsys "2"
311              ysys "2"
312              line 1
313              color "black"
314              width 3
315              when "a"
316              ;
317
318       length function $   8
319              text     $ 100
320              ;
321
322       set ns_SNP_NNs ;
323
324       y = &alleles. * &y_height. ;
325       x = 2 * AA - 1 ;
326       text     = strip( allele ) ;
327       position = "2" ;
328       function = "Label" ;
329       size     = 0.75 ;
330       output ;
331
332       text     = "|" ;
333       position = "5" ;
334       output ;
335
336     run ;
337
338     data anno ;
339       set anno_alleles
340           anno_domains
```

```
341          anno_glyco
342          anno_SNPs
343          ;
344    run ;
345
346    GOptions Reset   = All
347             HOrigin =  0  in
348             VOrigin =  0  in
349             VSize   =  8  in
350             HSize   = 10  in
351             Device  = HTML
352             ;
353
354    pattern1
355       value = s
356       color = white
357       ;
358
359    pattern2
360       value = s
361       color = blue
362       ;
363
364    pattern3
365       value = s
366       color = red
367       ;
368
369    ODS Listing Close ;
370    ODS NoResults     ;
371
372    ODS HTML Body  = "C:\Users\Histonis\PharmaSUG\waste.htm"
373             GPath = "C:\Users\Histonis\PharmaSUG\"
374             ;
375
376    proc gmap data = FVIII_MHCII_binding
377              map  = FVIII_MHCII_map
378              anno = anno
379              ;
380     id id ;
381     choro bindinglevel
382         / levels   = 3
383           coutline = same
384           nolegend
385           Name  = "A"
386           ;
387    run ;
388    quit ;
389
390    ODS HTML close ;
391
392    ODS Listing ;
393    ODS Results ;
394
395    Proc Catalog Catalog = WORK.gseg
396                 EntryType = GRSEG
397                 KILL
398                 ;
399    Quit ;
400
401    X rename "C:\Users\Histonis\PharmaSUG\A.gif" "FVIII_&length..gif"  ;
402
403  %mend binding_map ;
```

Figure 8. The Binding Map Macro.

STATISTICAL GRAPHIC PROCEDURES

If the thought of construction a map seems to be a questionable use of time, then the statistical graphics procedures might be worth the investment of effort to learn. Figure 9 presents a binding map produced as stacked needleplots using the TEMPLATE and SGRENDER procedures. The amount of flexibility using these procedures should increase the information, especially with regard to scaling and include tips that display upon mouse-over.

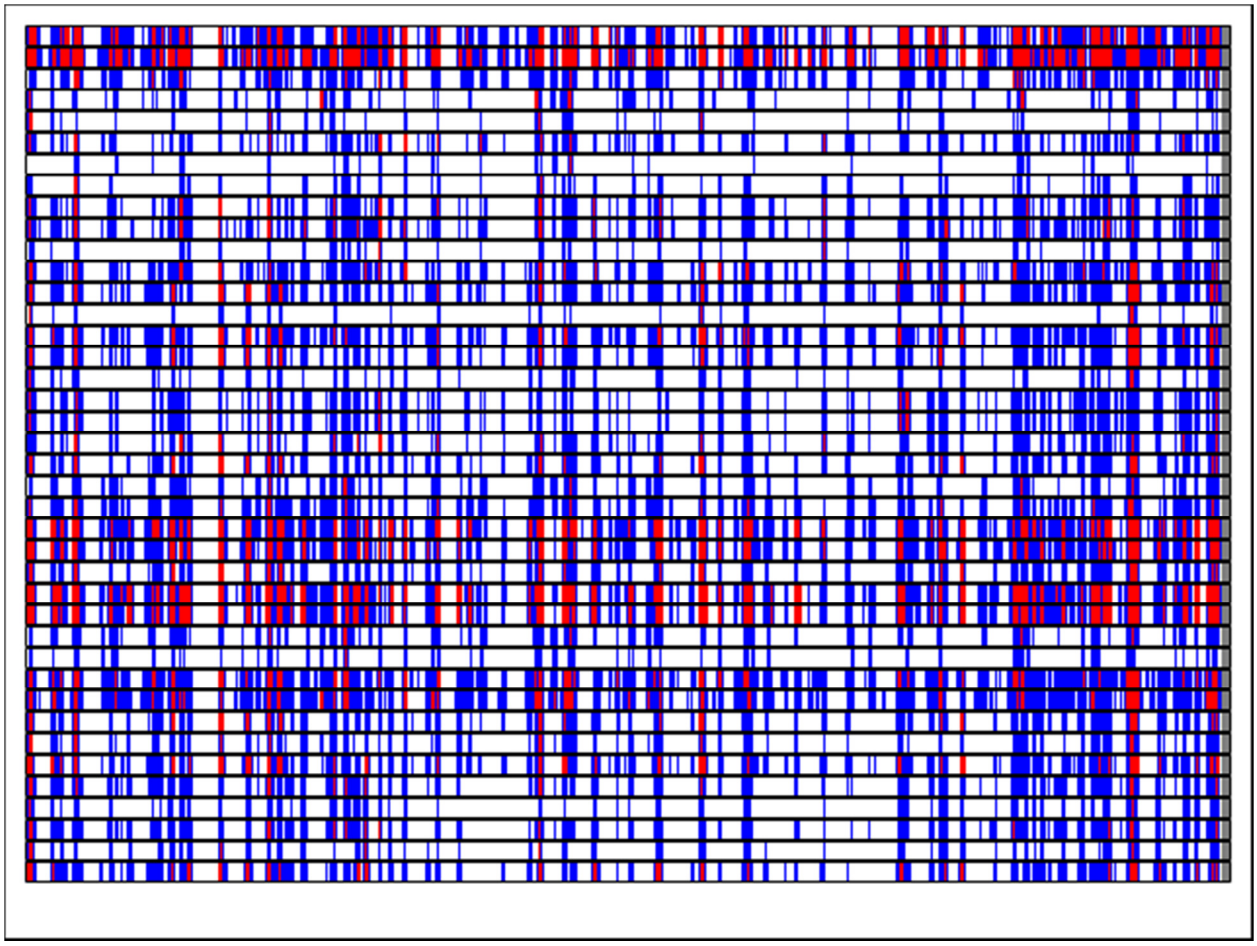

Figure 9. A Binding Map produced with the TEMPLATE and SGRENDER procedure.

Figure 10 presents the code that produced the map in Figure 9. Notably, the data are collected in columns as opposed to rows, but this is minor since each allele has full data. The ANNOTATE facility is available in SAS v9.4 as are additional plots statements like AXISTABLE.

```
data NetMHCIIpan_FVIII_FL_&length. ;
 set MHCIIpan.NetMHCIIpan_FVIII_FL_&length.
        ( Keep   = Allele
                   pos
                   BindingLevel
        )
     ;
 retain y 1 ;
 If BindingLevel = " " then Index = 1 ;
  Else if BindingLevel = "<=WB" then Index = 2 ;
   Else if BindingLevel = "<=SB" then Index = 3 ;
run ;
```

```
proc sql noprint ;
  select count( distinct allele ) into : alleles trimmed
  from NetMHCIIpan_FVIII_FL_&length.
  ;
quit ;

proc transpose data   = NetMHCIIpan_FVIII_FL_&length.
               out    = BindingLevel
                            ( Drop   = _Name_
                                       _Label_
                            )
               prefix = BindingLevel_
               ;
  var BindingLevel
      ;
  by pos ;
run ;

proc transpose data   = NetMHCIIpan_FVIII_FL_&length.
               out    = Index
                            ( Drop   = _Name_
                            )
               prefix = Index_
               ;
  var Index
      ;
  by pos ;
run ;

proc transpose data   = NetMHCIIpan_FVIII_FL_&length.
               out    = HLA
                            ( Drop   = _Name_
                                       _Label_
                            )
               prefix = HLA_
               ;
  var Allele ;
  by pos ;
run ;

data FVIII ;

  retain y 1 ;

  merge BindingLevel
        Index
        HLA
           end = end
        ;
  by pos ;

  pos = pos + 1 ;

  array BindingLevel_( &alleles. ) $ 10 ;
  array Index_       ( &alleles. ) ;
  array y_           ( &alleles. ) ;

  Output ;

  if end = 1
  then
    do ;
       do pos = pos + 1 to ( 2351 ) ;
```

```
         do _n_ = 1 to &alleles. ;
           BindingLevel_( _n_ ) = "<=ZZ" ;
           Index_( _n_ ) = 4 ;
         end ;
         If pos = 2351 Then Domain_y = 1 ;
         output ;
       end ;
    end ;
Run ;

proc template ;
  define style Styles.MyDefault ;
    parent = Styles.Default ;
    style GraphData1 from GraphData1
        / contrastcolor = white
          ;
    style GraphData2 from GraphData2
        / contrastcolor = blue
          ;
    style GraphData3 from GraphData3
        / contrastcolor = red
          ;
    style GraphData4 from GraphData4
        / contrastcolor = grey
          ;
  end ;
run ;

%macro layout_needle
   ( alleles = ) ;

  %do i = 1 %to %eval( &alleles. - 1 ) ;

     layout overlay
          / cycleattrs  = true
            xaxisopts   = ( linearopts = ( viewmin      = 1
                                           viewmax      = &Protein_length.
                                           thresholdmin = 0
                                           thresholdmax = 0
                                         )
                            OffsetMin  = 0
                            OffsetMax  = 0
                            label      = " "
                            display    = none
                          )
            yaxisopts   = ( linearopts = ( viewmin = 0
                                           viewmax = 1
                                         )
                            display = none
                            OffsetMin  = 0
                            OffsetMax  = 0
                          )
            ;
       needleplot x          = pos
                  y          = y
                / group      = BindingLevel_&i.
                  index      = Index_&i.
                  lineattrs  = ( pattern = solid )
                  ;
     endlayout ;
  %end ;

  layout overlay
```

```
       / cycleattrs  = true
         xaxisopts   = ( linearopts = ( viewmin      = 1
                                        viewmax      = &Protein_length.
                                        thresholdmin = 0
                                        thresholdmax = 0
                                      )
                         OffsetMin  = 0
                         OffsetMax  = 0
                         label      = " "
                         display    = none
                       )
         yaxisopts   = ( linearopts = ( viewmin = 0
                                        viewmax = 1
                                      )
                         display = none
                         OffsetMin  = 0
                         OffsetMax  = 0
                       )
         ;
    needleplot x          = pos
               y          = y
             / group      = BindingLevel_&alleles.
               index      = Index_&alleles.
               lineattrs  = ( pattern = solid )
               ;
  endlayout ;

%mend  layout_needle ;

/*************/
proc template ;

  define statgraph needle ;

    begingraph
             / BackGroundColor = white
               ;

      layout lattice
           / columns   = 1
             rowgutter = 0
             ;

      %layout_needle
         ( alleles = &alleles. ) ;

      endlayout ; /* Lattice */

    endgraph ;
  end ;
run ;

GOptions Reset    = All
         HOrigin  =  0 in
         VOrigin  =  0 in
         VSize    =  6 in
         HSize    =  8 in
         ;

ODS Graphics
  / Reset        = All
    ImageFmt     = gif
    AntiAliasMax = 2400
```

```
    ;

ODS Listing Close ;
ODS NoResults     ;

ODS HTML Body  = "C:\Users\Histonis\PharmaSUG\waste.HTML"
         GPath = " C:\Users\Histonis\PharmaSUG\"
         Style = styles.MyDefault
         ;

proc sgrender data     = FVIII
              template = needle
              ;
run ;

ODS Listing ;
ODS Results ;

ODS HTML Close ;
```

Figure 10. The Statistical Graphics Procedure Approach to the Binding Map.

CONCLUSION

Elucidation of the immune response is essential in fields of pharmaceutical research diverse as vaccine development, autoimmunity, and immunogenicity. The binding of peptides in the grooves of HLA and the presentation of pHLA complex on the surface of APCs is required in the pathway. This paper presented a brief overview of the process and peptide binding maps based on the binding strength estimated by the NetMHCIIPan v3.0 server. With these maps, the investigator might gain a better understanding and potentially develop hypotheses to tests that might otherwise have been costly from the perspective of resources and effort. Although the GMAP procedure adequately produced the figures, the flexibility and power of the statistical graphics procedures TEMPLATE and SGENDER, combined with ODS, will further increase the information presented in the figure while decreasing the development time and variability between figures.

REFERENCES

[1] Andreatta, M., E. Karosiene, et al. (2015). Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. Immunogenetics 67(11-12): 641-650.

http://www.cbs.dtu.dk/services/NetMHCIIpan/

[2] Lindner, R. and R. Knorr (2009). "Rafting trips into the cell." Commun Integr Biol 2(5): 420-421.

[3] Bosch, B., E. L. Heipertz, et al. (2013). "Major histocompatibility complex (MHC) class II-peptide complexes arrive at the plasma membrane in cholesterol-rich microclusters." J Biol Chem 288(19): 13236-13242.

[4] Viel, K. R., B. Kim, et al. (2014). "The Spectrum of Amino Acid Substitutions Resulting from Single Nucleotide Substitutions in the Coagulation Biosystem: Impact on Identification By Mass Spectrometry." Blood 124(21): 4221-4221.

[5] Kent, W. J., C. W. Sugnet, et al. (2002). "The human genome browser at UCSC." Genome Res 12(6): 996-1006.

[6] Viel, K. R., D. K. Machiah, et al. (2007). "A sequence variation scan of the coagulation factor VIII (FVIII) structural gene and associations with plasma FVIII activity levels." Blood 109(9): 3713-3724.

[7] Viel, K. R., A. Ameri, et al. (2009). "Inhibitors of factor VIII in black patients with hemophilia." N Engl J Med 360(16): 1618-1627.

[8] Lenting, P. J., J. A. van Mourik, et al. (1998). "The life cycle of coagulation factor VIII in view of its structure and function." Blood 92(11): 3983-3996.

[9] Taubert, R., J. Schwendemann, et al. (2007). "Highly variable expression of tissue-restricted self-antigens in human thymus: implications for self-tolerance and autoimmunity." Eur J Immunol 37(3): 838-848.

[10] Shepherd, A., S. Skelton, et al. (2015). "Modification of Predicted Inhibitor Risk in Non-Severe Hemophilia-a By in silico Analysis of Human Proteome Homology with Wild-Type, FVIII-Derived Peptides." Blood 126(23): 290-290.

[11] Grabich, S. and Viel, K.R. I. Processing the RefSeq and CCDS Annotation Datasets Using the SAS System: Creation of Gene Reference. PharmaSUG 2011 Proceedings.
http://www.pharmasug.org/proceedings/2011/PO/PharmaSUG-2011-PO12.pdf

CONTACT INFORMATION

Your comments, questions, and corrections are valued and encouraged. Contact the author at:
Name: Kevin R. Viel, Ph.D.
Enterprise: Histonis, Incorporated
E-mail: kviel@histonis.org