Biophysics 205
Project proposal
Xinli Hu
April 29, 2014

Differential peptide binding with a single amino acid residue change in an HLA class II molecule

Background and Significance

Genetic variants in the human leukocyte antigen (HLA) genes, which encode major histocompatibility molecules (MHC) in humans, have been known for decades to confer strong risk or protection toward dozens of autoimmune diseases. HLA-DR is class II molecule, a heterodimer formed by an α chain (encoded by HLA-DRA, invariant) and a β chain (encoded by HLA-DRB1, -DRB3, DRB4, or DRB5). Through decades of linkage and recent genome-wide association studies, we have confirmed the contribution of HLA-DRB1*0401, among other alleles, to the risk of rheumatoid arthritis (RA). Recently, we have made important progress by pinpointing five amino acid residues in three HLA molecules that confer independent risk (1). These amino acid variants nearly completely explain the HLA association to RA risk. The single strongest amino acid risk variant is at DRβ1 position 11, where the presence of valine at position 11 of DRβ1 confers risk, whereas the presence of ser-11 is protective (**Figure 1**). While DRβ1-11 does not explain all of the RA risk conferred by the HLA region, DRβ1 is by far the strongest gene association, and position 11 is the strongest of the three associated amino acid associations in the gene. **I now hypothesize that this specific amino acid significantly alters the molecule's binding affinities toward antigenic peptides**.

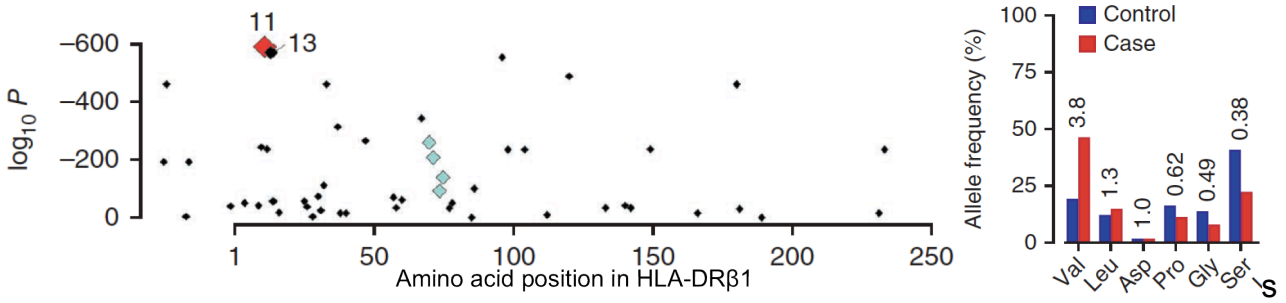

Figure 1. Amino acid position 11 of HLA-DRβ1 showed the strongest association with rheumatoid arthritis ($P < 10^{-581}$). Cases show an excess of valine at position 11, while controls show an excess of serine at the same position.

Each HLA class II molecule is capable of binding many peptides of varying lengths, and binding affinity to each HLA molecule is determined by the presence of "anchor residues" (as shown in **Figure 2** in green) (2). The position and the specific amino acid residues may change at an anchor position; however, physiochemical properties (*e.g.* hydrophobicity, charge, size) are highly conserved. The goal of the proposed project is therefore to compare the peptides that bind HLA-DR molecules containing valine or serine at position 11 of DRβ1, by screening against a library of potential autoantigenic peptides.

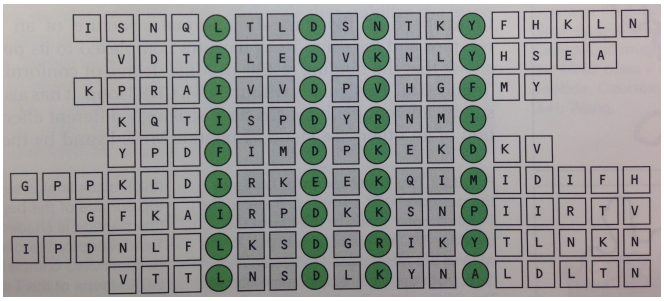

Figure 2. Example of peptides that bind to an HLA class II molecule. Anchor residues are shown in green circles (2).

Biophysics 205
Project proposal
Xinli Hu
April 29, 2014

Overall Strategy

The goal of the proposed work is to identify individual or families of human peptides that differentially bind to HLA-DR molecules with amino acid residues that either confers strong risk (11V) or relative protection (11S) toward rheumatoid arthritis. The workflow is illustrated in **Figure 3** below.

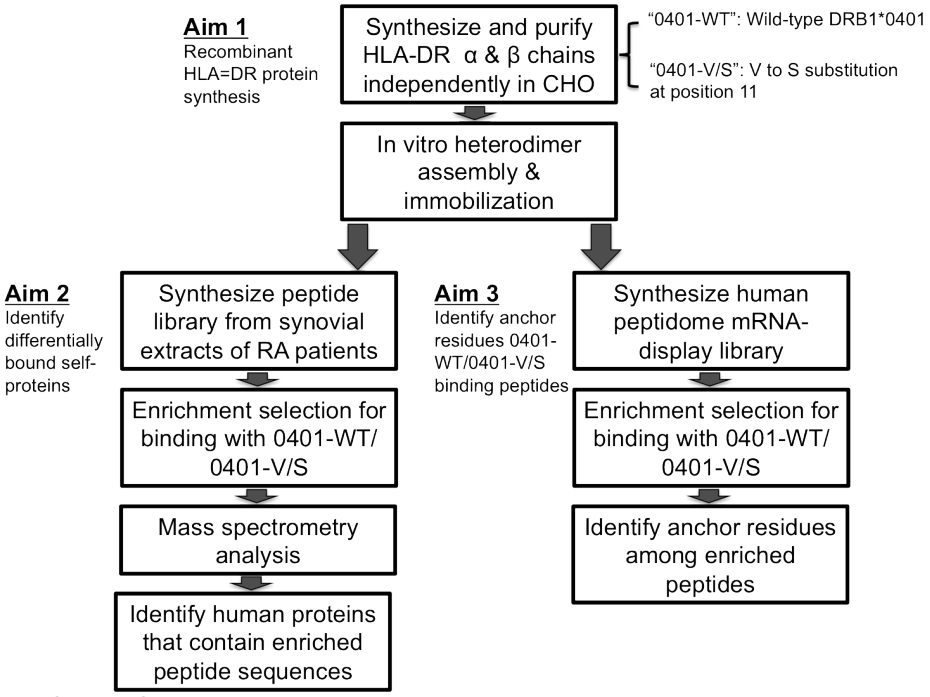

Figure 3. Schematic of workflow.

Aim 1. Synthesis of recombinant HLA-DR proteins.

The goal of Aim 1 is to synthesize and immobilize soluble and functional recombinant HLA-DR proteins with empty peptide grooves. I will synthesize two forms of the DR protein, from now on denoted "0401-WT" and "0401-V/S". Both forms have the same invariant α chain, while differing only at position 11 of the β chain. The 0401-WT β chain will have identical amino acid sequence to gene products of the HLA-DRB1*0401 alleles, having valine at position 11; 0401-V/S β will instead have serine at position 11 while keeping all other positions identical.

During *in vivo* class II HLA synthesis in human cells, the α and β chains spontaneously complex as a heterodimer, and the peptide binding groove is promptly occupied by the invariant chain (Ii). Ii is the default peptide that prevents the binding of cellular peptides and also facilitates the transport of the newly formed complex from the rough ER, and is eventually displaced by peptides with higher binding affinities to the binding groove. In synthesis of recombinant HLA-DR molecules in host cell systems, especially those in E. coli, it is difficult to insure the correct and timely incorporation of Ii. Two solutions have been demonstrated by previous works (3, 4) to circumvent the need for incorporating Ii. Justesen et al. synthesized the α and β chains in E. coli independently, followed by denaturing, refolding, and then assembly in a protein-free environment. Zhu et al. created a DR-peptide complex as a covalently linked single-chain in mouse L cells. In this proposed work, I will independently synthesize α and β chains in the Chinese hamster ovarian (CHO) cell line as the host system. The use of CHO is based on its mammalian machinery. Though CHO produces lower expression of recombinant proteins than E. coli, it also avoids the formation of inclusion bodies and the need to denature and refold.

To facilitate *in vitro* dimerization, and remove the unneeded transmembrane (TM) domains of both α and β chains chains, the TM domain on each chain will be replaced by a dimerization motif. This motif consists of a pentaglycine linker followed by the fos and jun leucine zipper, added to the C-termini (5). Justesen et al. used an E. coli nucleotide exchange factor, GrpE, which form symmetric dimers in solution, as a chaperone to direct dimerization. Here we use the human mitochondrial homologue, GrpE-mt (6). All chains will be tagged by N-terminus poly-histidine tags. To allow simultaneous purification and immobilization of assembled DR proteins, all β chains will also be fused at the C terminus to the AviTag, then biotinylated *in vitro* by incubation with BirA (7). Schematics of the α, β_{0401}, and $\beta_{0401\text{-}V/S}$ vectors are provided in **Figure 4**.

After cell lysis and purification of the α and β (0401 or 0401-V/S) chains, they will be combined to assemble each form of the DR heterodimer, with the facilitation of the dimerization motif and GrpE-mt chaperone. The peptide binding grooves will remain unoccupied.Assembled biotinylated DR complexes will be immobilized onto streptavidin argarose column.

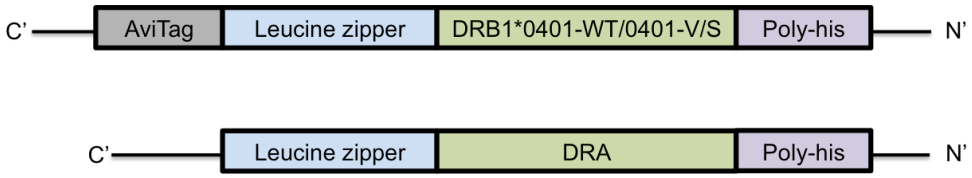

Figure 4. Vector design for recombinant DR α (bottom) and β (top) chains. AviTag is the biotin acceptor peptide. Leucine zipper is the dimerization motif that replaces the transmembrane portion of the proteins.

Validation of final DR protein product

1. Size validation by western blot. The expected size of the α/β heterodimer is around 65kD (36kD + 27kD)
2. Validation by Ii binding. Correctly folded and assembled DR proteins should be able to bind Ii (~30kD), and a size shift can be detected.
3. Binding by fluorescent-tagged commercial DR antibody.

Caveats and Alternative Approach

Due to the multi-step expression and purification of α/β chains followed by *in vitro* assembly of the complex, it is possible that the yield of functional protein product would be too low. Indeed, the yield achieved by Justesen et al. was around 20% (3). I hope the proposed system will achieve a higher yield by avoiding denaturation and refolding of α/β chains in vitro. However, if the proposed synthesis fails, both systems demonstrated by Justesen et al. and Zhu et al. may be modified to replace the current design. In addition, though Justesen et al. obtained free-form α and β chains and heterodimers with empty peptide binding grooves, it is nevertheless possible for the independent chains to nonspecifically bind to intracellular peptides and proteins (either in or outside the binding groove) in the host cell. DR proteins with nonspecifically bound peptides can be partly filtered out by size selection.

Aim 2. Selection for differential binders using peptides from RA patient synovial fluids.

The goal of Aim 2 is to identify relevant human synovial peptides that bind differentially to 0401-WT and 0401-V/S, and to identify the self-proteins that contain these peptides.

Synthesis of peptides from RA patients. Synovial fluids will be extracted from RA patients' affected joints and combined. The proteins will be digested using proteases to create proteolytic peptides following standard protocol. After digestion, the peptide sample will be randomly split equally into two fractions, to be used for selection against 0401-WT and 0401-V/S, respectively.

Enrichment selection. The peptide library will be poured through the 0401-WT column, then washed. During each exposure-wash cycle, the high-affinity binders remaining bound, while the weaker binders are progressively displaced by competitors. After a final wash and releasing the DR-bound peptides, peptides are labeled with an iTraq regent (8). Using the second fraction of the synovial peptides, the same enrichment selection procedures will be followed for binding to 0401-V/S. Released peptides will be labeled with a second iTraq reagent.

Protein identification. To identify the peptides and proteins that bound preferentially to 0401-WT or 0401-V/S, I will analyze the iTraq reagent-labeled samples from the two together. Obtained MS/MS spectra will be searched against all human proteins. Since modifications of peptides have been known to be relevant to RA pathology, the search for proteins will include citrullination, deamidation, and oxidation as variable modifications (9).

Aim 3. Identify preferential binding signatures to 0401-WT vs 0401-V/S using mRNA-display of synthetic human peptidome

The goal of Aim 3 is to use a larger and comprehensive peptide library to identify the anchor residues of peptide families that differentially bind to 0401-WT and 0401-V/S.

Synthesis of display library

In order to comprehensively assay the binding of relevant self-peptides, I choose to construct and utilize a synthetic human peptidome display library. The concept of this library is modified from the phage-display library constructed by Larman et al., and the final peptide repertoire of this library will be similar (10). However, I make two major modifications to better suit the proposed work. First, I will use mRNA instead of phage-display, as mRNA tags are smaller in size than phage, and potentially allow more flexible and natural peptide-DR binding conformation. In addition to size consideration, although the human peptidome used in this proposed work does not push the limit of phage library size bound by transfection efficiency (11), the mRNA system allows for expansion of tested peptidome diversity in the future. Second, Larman et al. divided the human open reading frame into overlapping segments of 36-amino acid peptides. Although peptides that bind class II HLA are not constrainted by size, most high-affinity binding peptides are shorter, at 18-20 amino acids (12). This modification roughly doubles the size of the display library from about 400,000 peptides to about 800,000 (10).

Construction of the mRNA will follow protocols described by Cotten et al. (13), while the peptide-coding sequences will be overlapping segments of all open reading frames from the human genome, as described by Larman et al (10). Each cDNA in the library will contain a T7 RNA polymerase promoter, a TMV translation enhancer sequence, an N-terminal FLAG tag coding sequence, a sequence encoding an 18-amino acid peptide, and a poly-histidine tag coding sequence. After generation of the library, all mRNA-tagged peptides are purified using anti-FLAG affinity column and used for selection (13).

Enrichment selection

Biophysics 205
Project proposal
Xinli Hu
April 29, 2014

The mRNA-display peptide library will be poured through the column, should saturate the DR proteins on chip, and multiple clones of each peptide should be present in the library as input. During each exposure-wash cycle, the copy numbers of high-affinity binders remaining on the chip accumulate, while weak binders are progressively displaced. After selection, DR-peptide complexes that remain bound are PCR amplified, which further enriches the species with higher binding affinities. PCR products are then sequenced to identify DR-bound peptides.

Data analysis.

Statistical evaluation of enrichment. Larman et al. devised a maximum likelihood estimator to assess the deviation of enrichment from a null distribution described by a two-parameter generalized Poisson (GP) distribution

$$pmf(x) = \theta(\theta + x\lambda)^{x-1} e^{-\theta - x\lambda} / x! \quad (10)$$

Given an input abundance of a species, the two parameters of GP, λ and θ, can be estimated.

$$\sum_{i=1}^{n} \frac{x_i(1-x_i)}{X+(x_i-X)\lambda} - nX = 0 \text{ where } X = \sum_{i=1}^{n} \frac{x_i}{n} \text{ and } \theta = X(1-\lambda) \quad (10)$$

Based on the estimated parameters, the p-value of obtaining *n* reads can then be estimated.

Identification of anchor residues. Here I will identify a set of amino acid positions in the peptides that resemble an “anchor residues”. Let positively-selected peptides that bound 0401-WT be a set {A}, and those that bound 0401-V/S be another set {B}. I hypothesize that {A} share a set of anchor residues A’, while {B} share another set of anchor residues B’. Because 0401-WT and 0401-V/S only differ by one amino acid, I expect A’ and B’ to differ at only one or a few positions. For each set A’ and B’, I will first use a peptide alignment and clustering algorithms based on Gibbs sampling approach, such as one devised by Andreatta et al. (14) to identify clusters of sequences that show enrichment and homology. However, because anchor residues are often separated by a certain distance and each position can be polymorphic, I will further conduct a pattern search to identify putative anchor residues, using a method which considers the physiochemical similarities of amino acids, such as one devised by Hobohm *et al.* (15). Once A’ and B’ are identified, they will be compared. In addition, I will use the conserved sequences to search against human proteins, and identify any other potential self-antigenic proteins not found in **Aim 2**.

Caveats

The proposed work cannot test the hypotheses that MHC-peptide binding influences autoimmune risk by modulating binding to T cell receptors. In addition, although the human peptidome library used in Aim 3 covers genome-wide peptide species, it does not include many isoforms of known genes, which may contribute to peptide binding. Finally, by focusing on one amino acid change at position 11, this work will have no ability to examine interactive effects between multiple amino acid positions, which could be the focus of follow-up work.

Acknowledgment

Development of this proposal benefited from helpful discussions with Nikola Teslovich, Yin-Yin Wang, Drs. Letian Kuai and Tobias Lenz.

Biophysics 205
Project proposal
Xinli Hu
April 29, 2014

References

1. Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nature genetics. 2012;44(3):291-6. doi: 10.1038/ng.1076. PubMed PMID: 22286218; PubMed Central PMCID: PMC3288335.
2. Kenneth Murphy PT, Mark Walport. Janeway's Immunology. 7th ed2008. p. 130-1.
3. Justesen S, Harndahl M, Lamberth K, Nielsen LL, Buus S. Functional recombinant MHC class II molecules and high-throughput peptide-binding assays. Immunome research. 2009;5:2. doi: 10.1186/1745-7580-5-2. PubMed PMID: 19416502; PubMed Central PMCID: PMC2690590.
4. Zhu X, Bavari S, Ulrich R, Sadegh-Nasseri S, Ferrone S, McHugh L, et al. A recombinant single-chain human class II MHC molecule (HLA-DR1) as a covalently linked heterotrimer of alpha chain, beta chain, and antigenic peptide, with immunogenicity in vitro and reduced affinity for bacterial superantigens. European journal of immunology. 1997;27(8):1933-41. doi: 10.1002/eji.1830270817. PubMed PMID: 9295029.
5. Kalandadze A, Galleno M, Foncerrada L, Strominger JL, Wucherpfennig KW. Expression of recombinant HLA-DR2 molecules. Replacement of the hydrophobic transmembrane region by a leucine zipper dimerization motif allows the assembly and secretion of soluble DR alpha beta heterodimers. The Journal of biological chemistry. 1996;271(33):20156-62. PubMed PMID: 8702739.
6. Borges JC, Fischer H, Craievich AF, Hansen LD, Ramos CH. Free human mitochondrial GrpE is a symmetric dimer in solution. The Journal of biological chemistry. 2003;278(37):35337-44. doi: 10.1074/jbc.M305083200. PubMed PMID: 12840016.
7. Kay BK, Thai S, Volgina VV. High-throughput biotinylation of proteins. Methods in molecular biology. 2009;498:185-96. doi: 10.1007/978-1-59745-196-3_13. PubMed PMID: 18988027; PubMed Central PMCID: PMC3223083.
8. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, et al. Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Molecular & cellular proteomics : MCP. 2004;3(12):1154-69. doi: 10.1074/mcp.M400129-MCP200. PubMed PMID: 15385600.
9. Raijmakers R, van Beers JJ, El-Azzouny M, Visser NF, Bozic B, Pruijn GJ, et al. Elevated levels of fibrinogen-derived endogenous citrullinated peptides in synovial fluid of rheumatoid arthritis patients. Arthritis research & therapy. 2012;14(3):R114. doi: 10.1186/ar3840. PubMed PMID: 22584083; PubMed Central PMCID: PMC3446491.
10. Larman HB, Zhao Z, Laserson U, Li MZ, Ciccia A, Gakidis MA, et al. Autoantigen discovery with a synthetic human peptidome. Nature biotechnology. 2011;29(6):535-41. doi: 10.1038/nbt.1856. PubMed PMID: 21602805.
11. Roberts RW, Szostak JW. RNA-peptide fusions for the in vitro selection of peptides and proteins. Proceedings of the National Academy of Sciences of the United States of America. 1997;94(23):12297-302. PubMed PMID: 9356443; PubMed Central PMCID: PMC24913.
12. O'Brien C, Flower DR, Feighery C. Peptide length significantly influences in vitro affinity for MHC class II molecules. Immunome research. 2008;4:6. doi: 10.1186/1745-7580-4-6. PubMed PMID: 19036163; PubMed Central PMCID: PMC2640366.
13. Cotten SW, Zou J, Wang R, Huang BC, Liu R. mRNA display-based selections using synthetic peptide and natural protein libraries. Methods in molecular biology. 2012;805:287-97. doi: 10.1007/978-1-61779-379-0_16. PubMed PMID: 22094812.
14. Andreatta M, Lund O, Nielsen M. Simultaneous alignment and clustering of peptide data using a Gibbs sampling approach. Bioinformatics. 2013;29(1):8-14. doi: 10.1093/bioinformatics/bts621. PubMed PMID: 23097419.

15. Hobohm U, Meyerhans A. A pattern search method for putative anchor residues in T cell epitopes. European journal of immunology. 1993;23(6):1271-6. doi: 10.1002/eji.1830230612. PubMed PMID: 7684684.