Motifscan: A Web-based Tool to Find HLA Anchor Residues in Proteins or Peptides

Rama Thakallapally[1], Warren Kibbe[2], Dorothy Lang[1], Bette Korber[1]

[1] *Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM*

[2] *Northwestern University, Chicago, IL*

Abstract

Motifscan is a new Web-based tool that summarizes anchor residue locations for specified HLA class I and HLA class II serotypes or genotypes in an input protein sequence (http://hiv-web.lanl.gov/immunology/). A searchable database of associated HLA serotypes and genotypes and anchor residues has been compiled from the literature; this dictionary will be updated annually. Motifscan is useful as a reference for HLA genotype/serotype nomenclature, as a resource for interpretation of mutations within epitopes, and as a tool for narrowing the search for optimal epitopes in a reactive region.

What Does MotifScan Do?

Motifscan is a simple tool that allows users to identify known anchor residue motifs for epitopes presented by class I proteins, and then use these motifs to highlight potential epitopes within a protein sequence. The listing of anchor residues (Rammensee, Hidehiro, Schreuder) and genotype nomenclature relationships (Schreuder), have been entered into searchable tables including information based primarily on data summarized in the first three references. There are many fewer defined anchor residues for HLA class II molecules at the present. Sometimes the anchor residues for a particular HLA genotype are undefined, but are defined for a closely related HLA genotype. Literature based updates of this table will be conducted annually. Once anchor motifs of interest have been identified from the table, these can be used to scan protein sequences. The tool can also be used as a quick Web-based reference to look up associated HLA genotype/serotype nomenclature, and to look up anchor residue motifs to help interpret mutations within epitopes.

We envision MotifScan being a particularly useful tool if a CTL response is characterized from an individual with a known HLA type, and already localized to a protein or protein region. Frequently, specified anchor residues will

be present in a region with no reactive epitope (false positives), and frequently one finds true epitopes that contain exceptions to anchor residue motifs (false negatives). The presence of HLA appropriate anchor residues could, however, help focus the search for potential epitopes in known reactive protein regions. An alternative for screening whole proteins for likely epitopes *de novo* is the program Epimatrix (De Groot), designed by Anne De Groot and colleagues at Brown University (http://tbhiv.biomed.brown.edu/).

A user can select either an HLA serotype or genotype from a listing. For example, if a user selects A2, anchor residues that have been defined for the HLA class I serotype A2 and for all related HLA A*02 genotypes will be returned. If a user knew only that the individual was A2, it might be useful to have the anchor residues for all related genotypes displayed; if the specific HLA genotype was defined, then the specific anchor residues would be of greater interest.

A search on HLA genotypes A*0201 and A*0207 yields two sets of slightly different anchor motifs. HLA A*0201 and A*0207 would both be classified serologically as A2:

```
A2   A*0207   x[L][D]xxxx[L],  x[L][D]xxxxx[L],  x[L][D]xxxxx[L]
A2   A*0201   x[LM]xxxxx[LV],  x[LM]xxxxx[LV],  x[LM]xxxxxx[LV]
```

In this output for A*0207, X can be any amino acid, L is favored in the second position, D in the third position, and L again in the C-terminal position. For A*0201, the second positions favors either L or M, and the C-terminal position either L or V, and the third position is undefined (an X). Optimal epitopes tend to be nine amino acids long, but this varies. Anchor motifs for epitopes between 8 and 10 amino acids are listed, however on rare occasion optimal epitopes longer than 10 amino acids have been described. Motifscan should be used cautiously; not all HLA genotypes or serotypes have defined anchor residues, and the tool will not be as up to date as the primary literature.

Once a motif has been identified, whether from the literature or from the HLA dictionary, it can be used to scan a protein sequence, like this A-subtype Rev protein sequence:

```
>A.SE.SE8538
AGRSGNSDEELLRAIRIIKILYQSNPHPKPRGSRQARKNRRRRWRARQRQ        50
IDSISERILSTCLGRSAEPVPLQLPPLEGLHLDCCEDCGTSGTEGVGRPQ       100
         **********   *********
         *********   ***********
                    **********
```

Motifscan

A list of potential epitopes is provided, and on the Web site these positions are highlighted in red (here they are indicated by asterisks). In this case HLA A*0201 anchor residue motifs were identified in the protein considering spacing appropriate for epitopes 8, 9, and 10 amino acids long:

Sequence I.D.	Location in Protein	Potential Epitope
A.SE.SE8538	(73–80)	QLPPLEGL (8mer)
A.SE.SE8538	(62–70)	CLGRSAEPV (9mer)
A.SE.SE8538	(71–80)	PLQLPPLEGL (10mer)
A.SE.SE8538	(73–82)	QLPPLEGLHL

This analysis could provide a reasonable starting place to narrow down the search for optimal epitopes in a reactive region.

References

[1] H. G. Rammensee, J. Bachmann and S. S. Stevanovic, *MHC Ligands and Peptide Motifs*, Landes Bioscience, Georgetown, TX (1997)

[2] Hidehiro, A compilation of anchor residue motifs available at the Graduate School of Genetic Resources Technology, Kyushu University
http://www.grt.kyushu-u.ac.jp/~hidehiro
/public_old/motifs.html

[3] G. M. Schreuder, C. K. Hurley, S. G. E. Marsh, M. Lau, M. Maiers, C. Kollman, H. Noreen, The HLA dictionary 1999: a summary of HLA-A, -B, -C, -DRB1/3/4/5, -DQB1 alleles and their association with serologically defined HLA-A, -B, -C, -DR and -DQ antigens. *Tissue Antigens* **54**:409–437 (1999)

[4] A. De Groot, B. Jesdale, E. Szy, J. Schafer, R. Chicz, and G. Deocampo, An Interactive Web Site Providing Major Histocompatibility Ligan Prediction: Application to HIV Research *AIDS Res Hum Retroviruses*, **13**:7, 529–531.
http://tbhiv.biomed.brown.edu/