Self-Nonself Discrimination Based on Incompatibility of Amino Acid Sequences of Human and Viruses

Wataru Honda honda@kuicr.kyoto-u.ac.jp
Shuichi Kawashima shuichi@kuicr.kyoto-u.ac.jp
Minoru Kanehisa kanehisa@kuicr.kyoto-u.ac.jp
Bioinformatics Center, Institute for Chemical Research, Kyoto University, Uji, Kyoto, Japan

Keywords: MHC, peptide, virus, epitope, MHC class I binding motif, KEGG/GENES, VGENES

1 Introduction

The immune system is known to have a greatly important function for human. It must be not only exclusive against pathogens or foreign molecules but also tolerant to self antigens. That is, the system distinguishes self components from non self substances. Major histocompatibility complex (MHC) class I molecules play a pivotal role in a immune response. Viral proteins translated in a host cell, so called endogenous antigens, are degraded into peptides composed of 8-10 amino acids [2] via the ubiquitin-proteasome pathway. Many of these peptides are loaded and presented by MHC class I molecules on the host cell surface. On the cell surface, MHC class I molecules with loaded viral derived peptides are recognized by Killer T lymphocyte as a token of the viral infected cell [4]. Additionally, these peptides share certain patterns of sequences called MHC class I binding motifs [1]. This well-known fact explains that merely 8-10 residues of amino acid sequences are enough in length to be discriminated from host protein fragments. Here we analyze all 8-10mer peptide patterns in all virus proteins and survey the frequency to match known motifs. Our results clarify that many of peptides observed only in viruses have a feature to be bound on MHC class I molecules.

2 Datasets and Methods

We selected 659 virus genomes from KEGG/VGENOME database. All human protein sequences and all protein sequences from viruses known to infect human were obtained from the KEGG/GENES and VGENES databases. We extracted all patterns of 8-10mer peptides in the virus and human proteins and listed patterns observed only in viruses. Known MHC class I binding motifs were represented in regular expressions based on anchor and auxiliary anchor residues data from SYFPEITHI [3]. Anchor and auxiliary anchor residues are defined by the score which is calculated by the frequency of amino acids in the respective position in aligned peptides.

Table 1: Proteins of human and viruses.

	Number of amino acids	Number of proteins
H. sapiens	8359195	16468
viruses	3842080	24745

Table 2: Example of MHC binding motifs.

Type of MHC class I	Motif
HLA-A*31012	x(1)-[LVYF]-[FLYW]-x(2)-[LFVI]-x(2)-R
HLA-B*39011	x(1)-[RH]-x(3)-[IVL]-x(2)-L
HLA-Cw*0401	x(1)-[YPF]-x(3)-[VIL]-x(2)-[LFM]

3 Results and Discussion

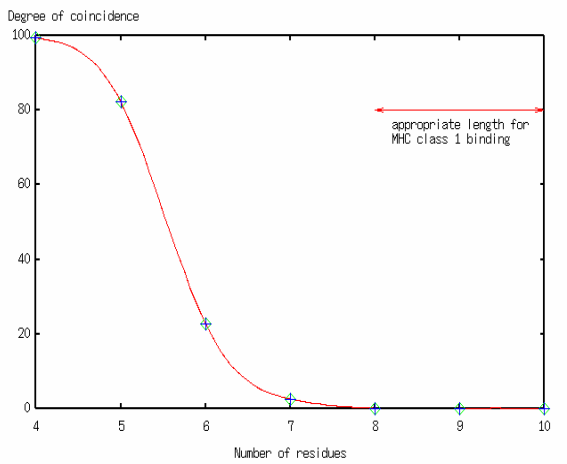

Figure 1:

As shown in the figure, the degree of coincidence in oligopeptide sequence patterns between *H.sapiens* and viruses decline markedly around 5∼7mers. Between 8-10mers, known as suitable lengths for MHC class I groove, almost all the oligopeptides derived from virus protein sequences are unique for viruses.

Next, we examined matching of oligopeptides unique for virus and MHC class I binding motifs. In 120 motifs we used, there are some clear features. For instance, hydrophobic residues are preferred in the carboxy terminus, and aromatic residues are often seen nearby the amino terminus. These allocations of amino acids in the motifs are suitable for binding to MHC hydrophobic groove. In the case of 10mers, we found almost half of the peptide sequences observed only in virus proteins shared common features with MHC class I binding motifs. Although several motif patterns have a possibility to overestimate the number of peptide matching motifs because of the limitation of our regular expressions, this result shows us that viruses are encoded by totally different amino acid sequences compared with human. This difference is remarkably interesting immunologically.

Table 3:

	9residues	10residues
All oligopeptides in virus	3075199	3102211
The number of	3071667	3100029
Oligopeptides unique in virus	99.9%	99.9%
The number of	887029	1532018
Motif sharing	28.9%	49.4%

In this report, matching of MHC class I binding motifs are examined. Analysis of other antigen presenting molecules, for example MHC class II, are ongoing. To integrate pattern recognition of these molecules will lead us to thorough comprehension of immune system strategy of exclusion, that is, sef-nonself discrimination.

4 Acknowledgments

This work was supported by the grants from the Ministry of Education, Science, Sports and Culture, the Science and Technology Agency, and the Japan Society for the Promotion of Science. The Computational resource was provided by the Bioinformatics center, Institute for Chemical Research, Kyoto University.

References

[1] Bouvier, M. and Wiley, D.C., Importance of peptide amino and carboxyl termini to the stability of MHC class I molecules, *Science*, 265(5170):398–402, 1994.

[2] Lauvau, G., Kakimi, K., Niedermann, G., Ostankovitch, M., Yotnda, P., Firat, H., Chisari, F.V., and van Endert, P.M., Human transporters associated with antigen processing (TAPs) select epitope precursor peptides for processing in the endoplasmic reticulum and presentation to T cells, *J. Exp. Med.*, 190(9):1227–1240, 1999.

[3] Rammensee, H., Bachmann, J., Emmerich, N.P., Bachor, O.A., and Stevanovic, S., SYFPEITHI: database for MHC ligands and peptide motifs, *Immunogenetics*, 50(3-4):213–219, 1999.

[4] Rock, K.L., Gramm, C., Rothstein, L., Clark, K., Stein, R., Dick, L., Hwang, D., and Goldberg, A.L., Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules, *Cell*, 78(5):761–71, 1994.