Solving Biological Problems that require Math
Lucyle Depoërs, Yannick Ormen, Karim Saied

Molecular constraints of the Major Histocompatibility Complex I (MHC I) protein

Background

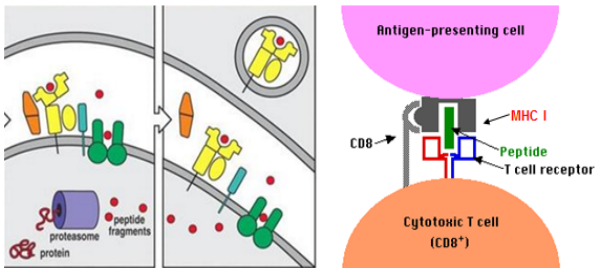

Figure 1: control of protein and the presentation of mutated peptide to Tcell.

CMH I has a key role in the daily control system of folding proteins in one cell. In fact, proteasome degrades the protein in peptides and transmits them via several mechanisms to the CMH I. CMH I bind to non-self and self peptids, mutated or not. If peptide is mutated HLA I go to the surface membrane of the cell and present peptides to Tcell to activate an immune response.

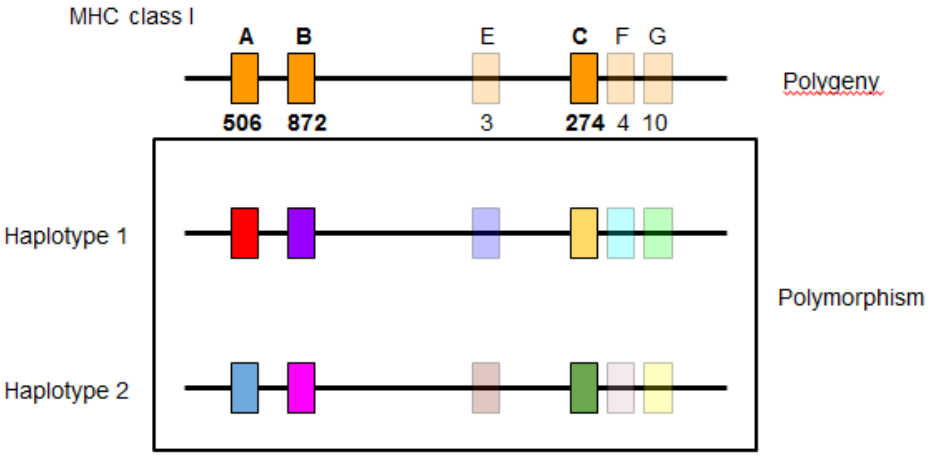

Figure 2. The definition of polymorphic, polygenic and haplotype.

MHCI genes are on the human chromosome 6, and there are 6 genes A B C D E F which encode different part of the MHC I molecule. So we say that MHC I is polygenic. Each gene is polymorphic, that is to say there are many alleles for the gene A , B and C. Genes D E and F are oligomorphic because there are less of ten alleles possible, so there more transparent on the figure. Anybody has the same combination of alleles for the MHC I and we name this combination of allele: the haplotype. This polymorphism is very unusual for the genes of the human genome and we have not enough found all the possible variants for the human.

The peptide binding groove of MHCI is the site which permits the MHC-peptide interaction. We call this binding site a “groove” because the protein forms a “sandwich” which envelop the peptide. We can see that the peptide binding groove is composed of two alpha helix on the sides and beta sheets at the bottom.

Figure 3. the peptide binding groove.

The peptide has only 9 amino acids. Just few amino acids are responsible for the MHC-peptide interaction. We call these residues, the Anchor residues because they anchor the peptide within the peptide binding groove. There are highlighted in green on the figure 4 and we can see that the anchor residues are to the positions 2 and 9.

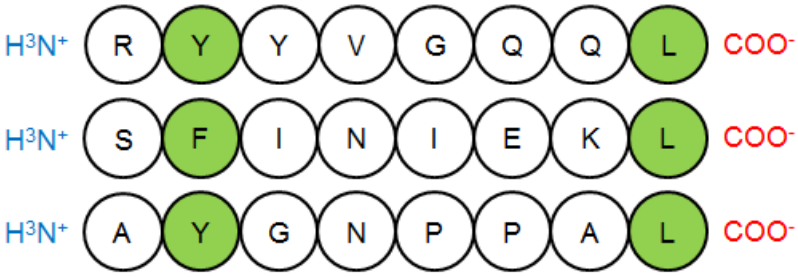

Figure 4. The anchor residues (2 and 9).

Main Goal

Find structural and functional key position in MHC I – peptide interaction.

Methodology

To reach the main goal we looked over sequences of human MHC I and corresponding peptides to find coevolved site between them (MHC I and peptide).

Figure 5. Example of aligned sequences and the coevolution between the amino acids of 2 different proteins and the conservation at a specific position of the protein 1.

We analyzed 95 MHC I human sequences with their peptide from healthy people: the MHC I sequences come from a database called Pfam, that contain 17'000 human sequences of MHC and the data of peptides (self and non-self) from another database called IEDB (Immune Epitope DataBase and analysis resource).

To do so we first aligned our sequences and for that we used a multiple sequence alignment software called MUSCLE.
Then, we investigated coevolved sites between MHC I sequences and corresponding peptide with an algorithm, mutual information, calculated by a software called MISTIC.
Here is the formula of the mutual information (MI) algorithm :

$$MI(i,j) = \sum_{a=1,b=1}^{20} P(a_{i,}\, b_j) \cdot \log\left(\frac{P(a_{i,}b_j)}{P(a_i) \cdot P(b_j)}\right)$$

$P(a_i)$: Probability to have an amino acid "a" at the position "I".
$P(b_j)$: Probability to have an amino acid "b" at the position "j".
$P(a_{i,}b_j)$: Probability to have these 2 amino acids, a and b, at these position, i and j, at the same time

This algorithm calculates a score for each pair of column between the aligned MHC I sequences and the corresponding peptide.
If $P(a_{i,}b_j)$ occurs more often than independently then the log will be greater than one and the MI sore will be positive.
Then, the scores are normalized (z-score transformation) and higher the score is, greater the 2 positions coevolve together.

Results

Positions (peptide – MHC)	Mutual information scores
2 – 44	32.72
2 – 66	27.99
2 – 65	23.97
2 – 69	23.97

Table 1. The four highest scores of mutual information and the relative positions on the peptide and the MHC molecule. The colors highlight the residues on the figures 2 and 3.

The four highest scores show a good coevolution between the second position on the peptide and a position somewhere else on the MHC molecule. For example, the highest score (32.72) highlights a coevolution between a residue at the second position on the peptide and a residue at the position 44 on the MHC molecule (Figure 1). Thus, the position 44 seems to be a key position for the interaction with the peptide antigen. In fact, the two positions involve amino acids that can interact with each other because if we look at the biochemistry properties of the residues, one can note that somehow an interaction can occur between them.

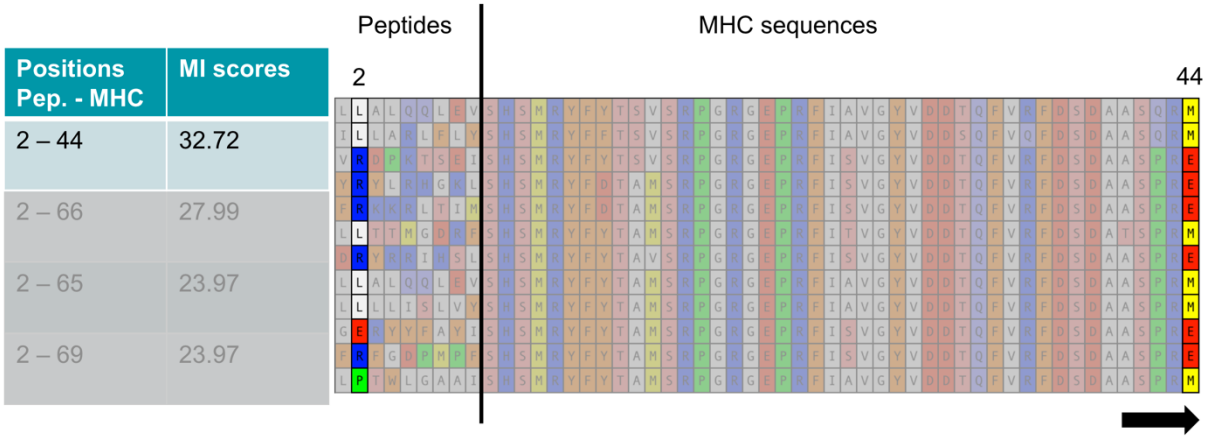

Figure 6. Co-evolution between the second position on the peptide and the position 44 on the MHC sequence relative to the first score. The sequence alignment shown is a small part of the whole alignment. The nine first residue correspond to the peptide.

As shown above (Figure 1), at the second position on the peptide, every time one has a leucine, one finds a methionine at the position 44 on the peptide. Thus one has two non polar amino acids that can interact with each other by non polar interactions. In the same way, every time one has an arginine (basic residue) at the same position on the peptide, one finds a glutamic acid on the MHC molecule. Thus one has two residues that can interact with each other by acidic-basic interactions. The two positions mentioned (2 and 44) evolve together (coevolution) in order to maintain two amino acids that can interact with each other. The example mentioned above concern the first score but it is the same case for the three remaining scores.

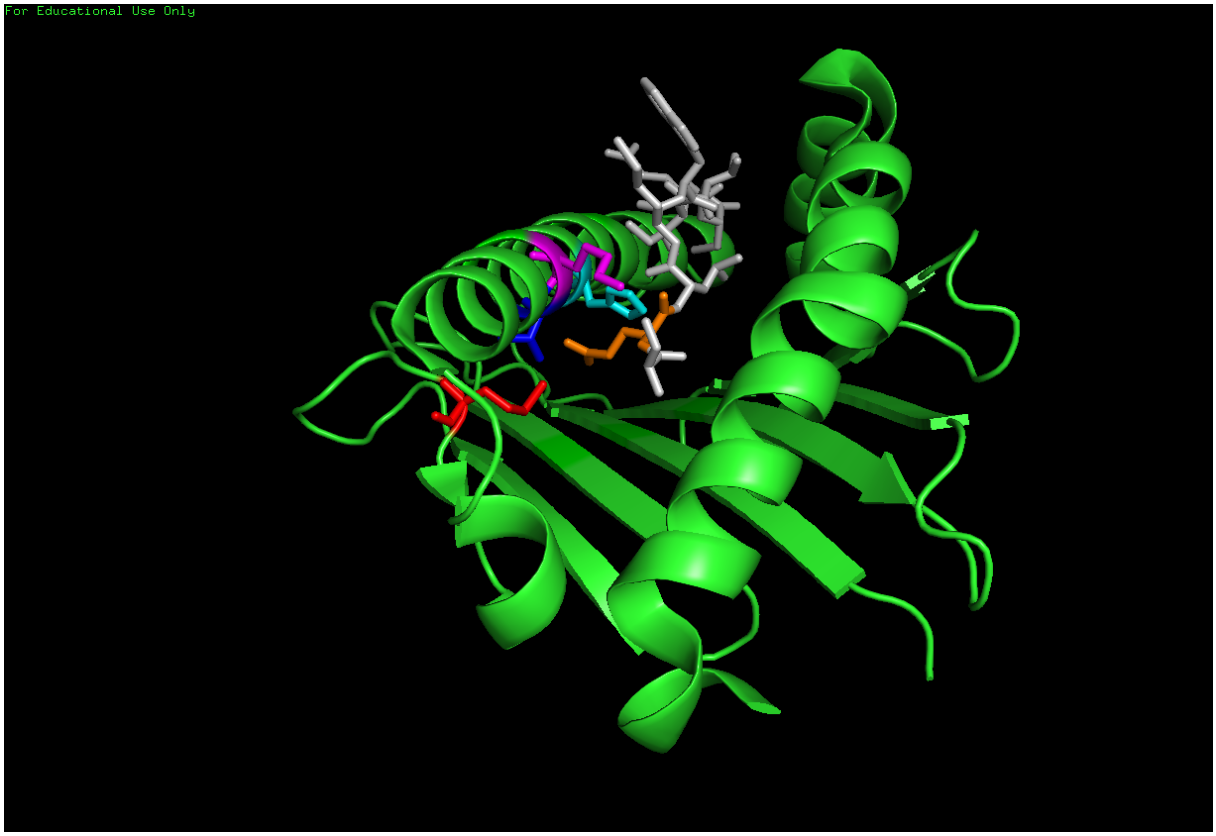

Figure 7. 3D structure of the peptide-binding groove of the MHC molecule. The peptide is in grey while its second position is highlighted in orange. The colors of the residues on the MHC molecule correspond the the colors on table 1.

If one looks at the structure, the residue positions within the peptide-binding groove of the MHC molecule are all located in close proximity with the second position of the peptide (highlighted in orange), as shown on the figure bellow (Figures 2 and 3).

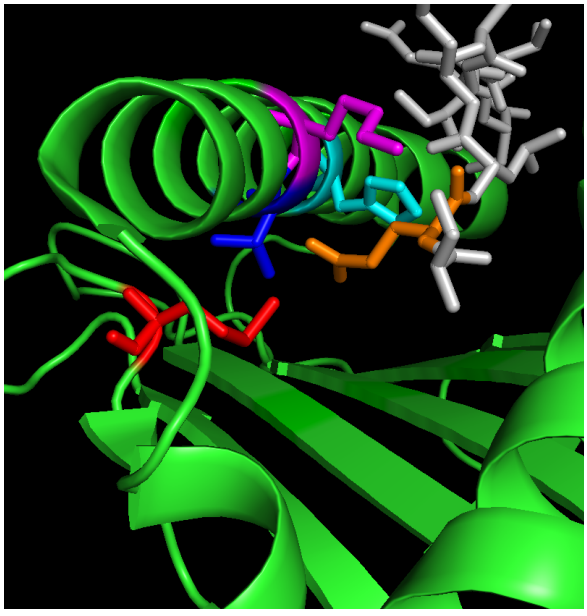

Figure 8. Zoom of the peptide-binding groove shown on figure 2. One can note the close spatial proximity between the residues on both MHC molecule and the peptide.

Conclusion

Taken together, the results show different key positions on the MHC molecule that are important for the interaction with the peptide antigen. Although we have got good results, one has to pay attention to several points.

First of all, upon 17'000 human MHC sequences, only 95 of them have identified peptides. Thus, our study was based only on these 95 sequences; the

number of human MHC sequences considered has to be increased in order to improve the results. However, nowadays we are limited by the number of MHC sequences that have identified peptide.
Another important point is that one MHC molecule can bind more than one peptide. Thus ideally actual methods should be improved in order to study intermolecular co-evolution by considering the fact that we can have several peptides for one MHC molecule.
Finally, a high score is not an absolute proof of co-evolution but it rather suggests it. This kind of correlation can be more complex than just compare and analyze two pairs separately.