Journal of Immunological Methods xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Immunological Methods

journal homepage: www.elsevier.com/locate/jim

Research paper

Viral peptides-MHC interaction: Binding probability and distance from human peptides

Daniele Santoni*

Institute for System Analysis and Computer Science "Antonio Ruberti", National Research Council of Italy, Via dei Taurini 19, Rome 00185, Italy

ARTICLE INFO

Keywords:
Nullomers
Peptide-HLA
Immunoinformatics
Viral genomes
Self/non-self

ABSTRACT

Identification of peptides binding to MHC class I complex can play a crucial role in retrieving potential targets able to trigger an immune response. Affinity binding of viral peptides can be estimated through effective computational methods that in the most of cases are based on machine learning approach. Achieving a better insight into peptide features that impact on the affinity binding rate is a challenging issue.

In the present work we focused on 9-mer peptides of Human immunodeficiency virus type 1 and Human herpes simplex virus 1, studying their binding to MHC class I. Viral 9-mers were partitioned into different classes, where each class is characterized by how far (in terms of mutation steps) the peptides belonging to that class are from human 9-mers. Viral 9-mers were partitioned in different classes, based on the number of mutation steps they are far from human 9-mers. We showed that the overall binding probability significantly differs among classes, and it typically increases as the distance, computed in terms of number of mutation steps from the human set of 9-mers, increases. The binding probability is particularly high when considering viral 9-mers that are far from all human 9-mers more than three mutation steps. A further evidence, providing significance to those special viral peptides and suggesting a potential role they can play, comes from the analysis of their distribution along viral genomes, as it revealed they are not randomly located, but they preferentially occur in specific genes.

1. Introduction

Human Leukocyte Antigens (HLA) are genes coding for the Major Histocompatibility Complex (MHC) proteins in humans. HLAs belonging to MHC class I are able to bind and present peptides from inside the cell (Yewdell et al., 2003).

When a cell is infected by a virus some fragments of viral proteins bind to MHC complex that allows the presentation of those peptides to the surface of the cell, so that the infected cell can be recognized by the immune system (Neefjes et al., 2011). Those fragments are the product of proteasome-mediated degradation of viral proteins; peptides able to bind to MHC class I are typically made of 9 aminoacids (ranging from 8 till 11) (Goldberg et al., 2002).

An immune response is triggered when T-cell receptors recognize peptides displayed on the cell surface as foreign molecules and consequently potentially harmful (Jiang and Chess, 2009). The distinction between endogenous (self) and exogenous (non-self) peptides is fundamental for a functional and effective immune response; many factors can affect the recognition process as highlighted by Grignolio and colleagues (Grignolio et al., 2014). Coherently with the self/non-self paradigm the overwhelming majority of viral peptides is not shared with human and only few peptides are in common (as also shown in the Results section).

Our results are consistent with other works that studied the relationships between self and non self and the degree of overlap between microbial and human proteomes. Both Burroughs and colleagues (Burroughs et al., 2004) and Calis and colleagues (Calis et al., 2012) studied the information contained in presented peptides (peptides binding to MHC class I complex) to investigate how the immune system achieve a proper self/non-self assessment. Other recent works studied virus-human proteome overlap looking for small DNA epitopes in the human genome originated from ancient viral sequences (Davis, 2016).

Several algorithms were designed to predict the interaction between peptides and MHC class I (see the review (Schirle et al., 2001) for more details). For our analysis we used NetMHC, a method based on artificial neural networks trained on a large dataset, that has been showed to generate high-accuracy predictions (Nielsen et al., 2003; Lundegaard et al., 2008).

In recent years many studies, based on those predicting algorithms, focused on the identification of peptides strongly binding to MHC-class

* Corresponding author.
E-mail address: daniele.santoni@iasi.cnr.it.

https://doi.org/10.1016/j.jim.2018.05.009
Received 21 November 2017; Received in revised form 26 March 2018; Accepted 9 May 2018

I complex to map targets of immune response and aid in developing new vaccine targets (Patronov and Doytchinova, 2013). He and colleagues developed a Web-Based Vaccine Design tool called Vaxign, it was applied to > 70 viral genomes, it can also perform dynamic vaccine target prediction based on input sequences (He et al., 2010). Several other studies focused on epitopes prediction for specific viruses. Zheng and colleagues proposed a computational method to predict suitable Hepatitis B (HBV) targeting peptides (Zheng et al., 2017); Cunha-Neto and colleagues applied their method to Zika Flavivirus (Cunha-Neto et al., 2017), Sànchez-Burgos and colleagues to Dengue virus (Sànchez-Burgos et al., 2010) and Srivastava and colleagues to Ebola virus (Srivastava et al., 2016).

Similar computational approaches were also applied in studies on cancer therapeutics, Dhanik and colleagues developed a computational framework to identify potential peptide-HLA cancer targets (Dhanik et al., 2016).

In the present work we extended, to some extent, the concept of self/non-self. Non-self peptides were partitioned into different classes depending on their shortest distance from the whole collection of human peptides. The higher the distance from the set of human peptides the higher the non-self *order* and the stronger the non-self *status* of a given peptide will be; we referred to those classes of peptides as nullomers or high order nullomers. We then investigated the affinity of different peptide classes to MHC class I.

We focused on viral 9-mers of Human Immunodeficiency Virus type 1 (HIV1) and Human herpes simplex virus 1 (HHV1), two of the most studied human-host viruses, predicting their interaction with MHC class I by NetMHC, considering 78 different HLAs. We computed the interaction probability distribution of all possible viral 9-mers for the 78 different HLAs, selecting for each given peptide the highest score out of the 78 HLAs. Then we focused on nullomers and high order nullomer classes, showing that the overall binding probability significantly differs among classes, and it typically increases as the number of mutation steps they are far from human 9-mers increases. In particular we showed that the binding probability is significantly higher when considering a particular class, made of viral 9-mers that are far from all human 9-mers more than three mutation steps. We finally studied the distribution of 9-mers belonging to this special class in viral proteins.

2. Materials and methods

2.1. Proteomes

The proteome of *Homo sapiens* sapiens (HSA) GRCh38, was downloaded from Ensembl site (http://ftp.ensembl.org/pub/current_fasta/homo_sapiens/pep/). Available strain protein sequences of Human Immunodeficiency Virus type 1 (HIV1) - 770,532 sequences - and Human herpes simplex virus 1 (HHV1) - 4546 sequences - were downloaded from UNIPROT site (http://www.uniprot.org/uniprot/). Ad hoc scripts were designed in order to extract all possible 9-mers from the sequences of the organisms mentioned above. The whole set of 9-mers of

- HSA is indicated with $HSA_P{}^9$
- HIV1 is indicated with $HIV1_P{}^9$
- HHV1 is indicated with $HHV1_P{}^9$

where the superscript 9 is related to the size of k-mers and the subscript *P* stands for'Present'. For the sake of simplicity we will omit the superscript 9 in the following definitions since we will always refer to 9-mers.

2.2. Nullomers and high order nullomers

A nullomer is a word that does not occur as a subword in a given reference sequence (or in a collection of reference sequences), i.e. it is an absent word of that sequence (or those sequences). According to (Vergni and Santoni, 2016) we consider an extension of the notion of nullomers, namely high order nullomers, which are nullomers whose mutated sequences are still nullomers. For instance, given a reference sequence *s*, a 1-order nullomer w_1 is an absent sequence of *s* such that all the sequences obtained by any single mutations of w_1 are still absent in *s*. In the same way we define a 2-order nullomers w_2 as an absent sequence such that all the sequences obtained by any two mutations of w_2 are still absent in *s*. Accordingly we define a *t*-order nullomer w_t as an absent sequence such that all the sequences obtained by any *t* mutations of w_t are still absent. We can say that a k-mer w_t is a *t*-order nullomer with respect to a reference *s* if w_t is far from every k-mer of *s* more than *t* mutation steps (see Materials and Methods - Nullomers and high order nullomers (Vergni and Santoni, 2016) for details).

Taking as a reference the set of human 9-mers HSA_P we define HSA_{H0} as the set of simple nullomers, HSA_{H1} as the set of 1-order nullomers, HSA_{H2} as the set of 2-order nullomers and HSA_{H3} as the set of 3-order nullomers.

We firstly define the sets of 9-mers that viruses share with human:

$$HIV1_C = HSA_P \cap HIV1_P$$
$$HHV1_C = HSA_P \cap HHV1_P \quad (1)$$

We then define, for each considered virus, the sets of viral 9-mers that are nullomers or high order nullomers with respect to human:

$$HIV1_{Hi} = HSA_{Hi} \cap HIV1_P$$
$$HHV1_{Hi} = HSA_{Hi} \cap HHV1_P \quad (2)$$

for $i = 0, 1, 2, 3\ldots$ It is worth noting that nullomers and high order nullomers are always computed with respect to human present peptides HSA_P. As a consequence of the above definitions we will have for each considered virus $V \in \{HIV\ 1, HHV\ 1\}$ a collection of sets such that:

$$V_P \supseteq V_{H0} \supseteq V_{H1} \supseteq \ldots \supseteq V_{Hi}\ldots \quad (3)$$

for $i = 0, 1, 2, 3\ldots$
and

$$V_P = V_C \cup V_{H0} \quad (4)$$

Ad hoc scripts were designed in order to compute the above defined classes of 9-mers till 3-order.

2.3. Peptide-MHC class I interaction score

The software NetMHC (version 3.4) (Nielsen et al., 2003; Lundegaard et al., 2008) was used to predict the interaction of peptides (9-mers) with the MHC class I complex in terms of binding probability, taking into account 78 different HLAs (34 class A, 33 class B 10 class C and 1 class E). NetMHC provides for every peptide and for any given HLA an interaction probability score. We associated to every considered peptide the highest score out of all the 78 scores computed on the corresponding HLAs, since we want to identify those peptides that are able to bind at least one HLA.

According to NetMHC software three possible interval scores are defined:

- NB - No Bind: score smaller than 0.4
- WB - Weak Bind: score equal or higher than 0.4 and smaller than 0.65
- SB - Strong Bind score equal or higher than 0.65

3. Results

As reported in the previous section, taking as a reference the human proteome, we naturally define a collection of 9-mer classes. In the next subsections we will firstly consider human 9-mers and we will compare the binding affinity of peptides, belonging to different classes, to MHC class I. We will then consider viral peptides and analyze their affinity to