Transmembrane helices are also an overlooked source of major histocompatibility complex class II epitopes and evolutionary more conserved than expected by chance

Richel J.C. Bilderbeek[1] and Frans Bianchi[2]

[1]Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands
[2]Frans' Institute, University of Groningen, Groningen, The Netherlands

April 12,2019

Abstract

Transmembrane helices (TMHs) are an overrepresented source of epitopes on major histocompatibility complex (MHC) class I due to a preference for hydrophobic for the majority of HLA-I haplotypes. It is unknown why there is such a preference and whether this is the case for MHC-II haplotypes. This study investigates if there is an evolutionary conservation of a pathogens TMHs and if this correlates with the affinity for binding to MHC-II. TMHs more/less/equal as expected by chance. We nd that …

Keywords: antigen presentation, membrane proteins, bioinformatics, adaptive immunity, transmembrane domain, epitopes, T lymphocyte, MHC-2

1 Introduction

For MHC-1, it was found that epitopes derived from transmembrane helices (TMHs) are over-presented by all human leukocyte antigen (HLA)-A and most HLA-B super types [Bianchi et al. 2017]. One explanation is that the presentation of especially TMHs may have an evolutionary advantage for the (human) host. The argumentation for this is that the TMHs are more conserved due to the functional requirement of being able to span a lipid bilayer. Additionally, there are limited escape mutations possible for the pathogen, as many will result in a non-functional TMH. However, that argumentation can also be countered as there are limited unique polypeptide fragments and TMHs are a conserved throughout all species, it is likelier that host and pathogen share the same unique epitope sequences, resulting in the pathogen avoiding detection due to the negative T-cell selection process in the thymus.

If presentation of TMHs on MHC-I would bring an evolutionary advantage in the recognition of pathogens by the immune, it would follow that this is equally important for MHC-II. MHC-II, which is necessary for the activation of B-cells after the detection of bacterial pathogens, e.g. tuberculosis, by dendritic cells or [...]. From the earlier research done on MHC-I, it naturally follows to check if also MHC-II is more likely to present TMHs than expected by chance. Extending the previous study on MHC-I, we also measure the evolutionary conservation of TMH epitopes in Mycoplasma pathogens to address if presentation of TMH derived epitopes has an evolutionary advantage.

2 Methods

Transmembrane helices and strong MHC-II-binding peptides were predicted for a tuberculosis reference proteome from https://www.ebi.ac.uk/reference_proteomes using the epitopeome [RB: I'd enjoy a better name] R package [Bilderbeek 2019a]. We picked the 117 MHC-II alleles that are most abundant in the current human population [Greenbaum et al., 2011]. The 5% peptides with the lowest IC50 values were defined as binders. We then simply counted the number of amino acids that were present inside the cell, within the membrane or outside of the cell, as well as if it was part of a strong MHC-II binding site, as shown in figure 1.

The epitopeome R package Bilderbeek 2019a binds [RB: pun intended] together the tmhmm [Bilderbeek 2019d] and netmhc2pan [Bilderbeek 2019c] R packages. tmhmm provides an R interface to TMHMM [Krogh et al. 2001, Sonnhammer et al. 1998], a tool to predict where membrane proteins' amino acids are located within the membrance. netmhc2pan provides an R interface to NetMHC2pan [Jensen et al. 2018], a tool to predict MHC-II binding to proteins.

The evolutionary conservation of TMHs was measured from a DNA alignment of multiple transmembrane proteins in multiple species of the mycoplasma bacterial family [RB: how obtained exactly? how alignment done?]. Using the tmhprot R package [Bilderbeek 2019e], the DNA alignment is split into two alignments, one for the TMH parts, another for the non-TMH

parts. Each alignment was tested by the mcbette R package [Bilderbeek 2019b] to select the Bayesian inference model with the highest evidence (a.k.a. the marginal likelihood) using the nested sampling approach as described in Maturana et al. 2017, using the popular Bayesian phylogenetic tool BEAST2 [Bouckaert et al. 2014] in the back-end. mcbette used a set of 40 candidate inference models, consisting of all combinations of 4 site models (JC, HKY, TN, GTR), 2 clock models (strict, RLN) and 5 tree priors (Yule, BD, CBS, CCP, CEP).

For each alignment, the inference model with the highest evidence is used in a Bayesian inference. From each Bayesian posterior, the parameter estimates regarding mutation rates (including clock rate) were obtained and compared to determine which realized mutation rate is lower. The Markov chain Monte Carlo was set up in such a way that the effective sample size for the likelihood of the inference model is above the recommended value of 200 [Bouckaert et al. 2014].

3 Results

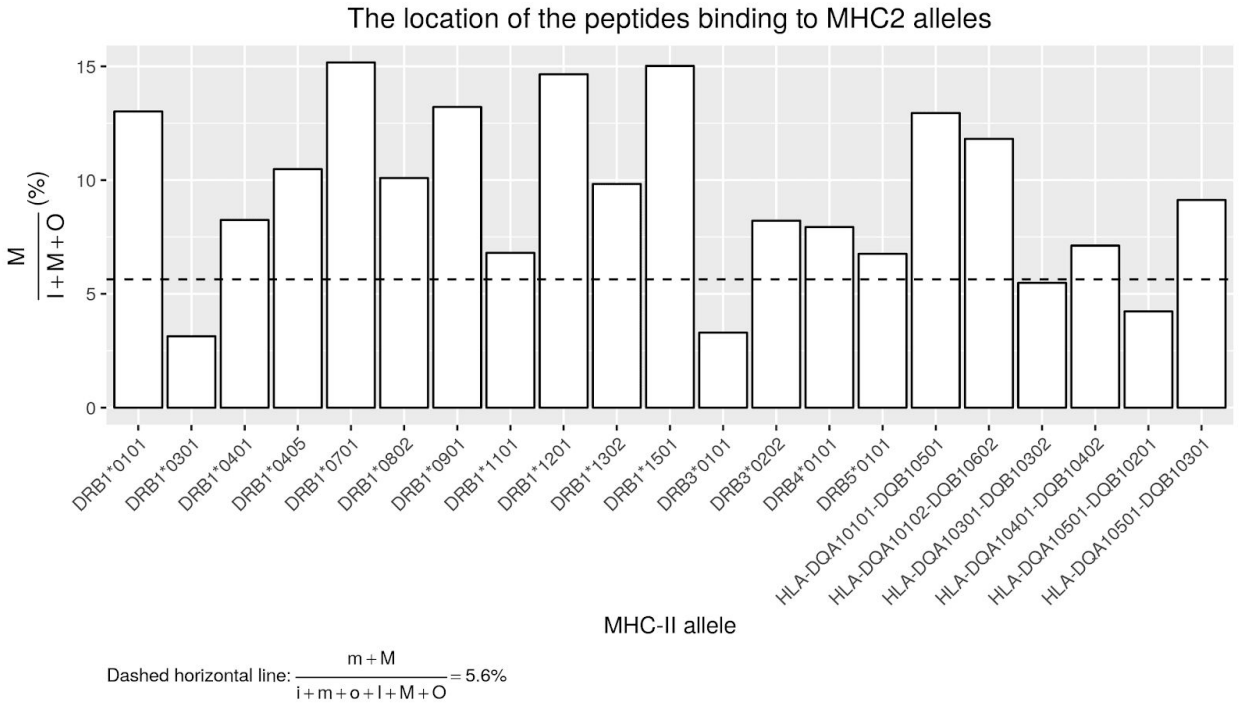

Figure 1: Location and binding of amino acids for the alleles. I: non-binding and inside, m: non-binding and transmembrane, o: non-binding and outside, I: binding and inside, M: binding and transmembrane, O: binding and outside

Figure 1 shows the location of TMH peptides binding to MHC2 alleles [RB: elaborate].

The inference model with the highest evidence in the TMH-only alignment was [yet unknown] and [also yet unknown] for the non-TMH alignment. Individual model weights are shown in tables [absent ones]. The Bayesian inference resulted in [a distribution of mutation rates], as shown in [absent figure]. The ESSes of the Bayesian parameter estimates was above 200, exact values are shown in tables 4 and 5.

4 Conclusion

We conclude that MHC-II binds to TMH peptides with a higher/lower/equal probability than expected by chance. We conclude that the evolutionary conservation if the TMH parts of membrane proteins is higher/less/equal compare to its non-TMH counterparts.

5 Discussion

We compared the mutation rates between the TMH and non-TMH part of multiple mycoplasma species. Where we expect no variation in mutation rate for every TMH amino acid, [RB: we can test this, but unsure if that would make sense] we know that non-TMH part will have regions of different evolutionary conservation: functional domains, especially in protein-protein interactions will be strongly conserved, due to an even more constrained set of peptides that enable a certain function. [RB: Note that most bacteria are opportunistic pathogens. Note that most bacteria are generalists. Note that most bacteria have different cell membranes (and walls), that may have different functional constraints than a human cell membrane]

6 Acknowledgements

Thanks to Geert van den Bogaart for his help [RB: reword]. Thanks to Maxim Boranov for his help [RB: reword].

7 Data Accessibility

All code is archived at `http://github.com/richelbilderbeek/someplace`, with DOI `https://doi.org/12.3456/zenodo.1234567`.

8 Authors' contributions

RJCB and FB conceived the idea for this research. RJCB wrote the code. FB supplied the DNA alignment. RJCB and FB contributed to writing the article.

References

Bianchi, F., Textor, J. & van den Bogaart, G. (2017) transmembrane helices are an overlooked source of major histocompatibility complex class i epitopes. Frontiers in

immunology, 8, 1118.

Bilderbeek, R.J. (2019a) epitopeome. https://github.com/richelbilderbeek/epitopeome [Accessed: 2019-03-08].

Bilderbeek, R.J. (2019b) mcbette. https://github.com/richelbilderbeek/mcbette [Accessed: 2019-01-21].

Bilderbeek, R.J. (2019c) netmhc2pan. https://github.com/richelbilderbeek/netmhc2pan [Accessed: 2019-03-08].

Bilderbeek, R.J. (2019d) tmhmm. https://github.com/richelbilderbeek/tmhmm [Accessed: 2019-03-08].

Bilderbeek, R.J. (2019e) tmhprot. https://github.com/richelbilderbeek/tmhprot [Accessed: 2019-03-08].

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A. & Drummond, A.J. (2014) Beast 2: a software platform for bayesian evolutionary analysis. PLoS Comput Biol, 10, e1003537.

Jensen, K.K., Andreatta, M., Marcatili, P., Buus, S., Greenbaum, J.A., Yan, Z., Sette, A., Peters, B. & Nielsen, M. (2018) Improved methods for predicting peptide binding affinity to mhc class ii molecules. Immunology, 154, 394–406.

Krogh, A., Larsson, B., Von Heijne, G. & Sonnhammer, E.L. (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. Journal of molecular biology, 305, 567–580.

Maturana, P., Brewer, B.J., Klaere, S. & Bouckaert, R. (2017) Model selection and parameter inference in phylogenetics using nested sampling. arXiv preprint arXiv:170305471.

Sonnhammer, E.L., Von Heijne, G., Krogh, A. et al. (1998) A hidden markov model for predicting transmembrane helices in protein sequences. Ismb, volume 6, pp. 175–182.