Revision

Editor

Thank you for the submission. Your manuscript has been examined by three expert reviewers. Two of the reviewers raised some significant concerns regarding the manuscript that can only be addressed by additional analyses and work. In particular, please pay special attention to the comment by reviewer #3, as follows: "In addition to the two peptidome studies used by the authors, many comprehensive immunopeptidomes are available for analysis. For example, the NIH Immune Epitope Database and Analysis Resource lists several hundred studies with many hundreds of thousands of epitopes. A detailed analysis of the frequency of TMH in experimentally determined MHC-I and MHC-II peptidomes (rather than predicted binders) could provide the information sought by the authors." Should you decide to revise the manuscript and resubmit, please make sure that you address all of the reviewer's concerns in the interactive forum. If you need additional time for the revision, please feel free to let me know.

We thank you for your help and the reviewers' feedback. Below we address all reviewers' comments.

First and foremost, it is clear something had gone wrong with the figures and their corresponding legends in our original submission. We apologize for this, and have double-checked the revised version of the manuscript.

As requested, we have added another source of data for naturally occurring epitopes, which is the suggested IEDB (the Immune Epitope Database from NIH), resulting in a rewrite of paragraph 2.2 ('TMH epitopes obtained from experimental data') and a new figure 2. The collection of this new data was the main reason for needing the extended amount of time. The added IEDB data strengthened the message of this paper, as we found an over-presentation of epitopes derived from tranmembrane helices for both MHC class I and II.

Reviewer 1

Interactive review activated: 26 Sep 2021 This Reviewer withdrew from the peer-review of this manuscript and independent review report is not submitted.

Reviewer 2

- Independent review report submitted: 04 Sep 2021
- Recommendation for the Editor: The manuscript can be accepted

The manuscript entitled "Transmembrane helices are an overlooked

and evolutionarily conserved source of major histocompatibility complex class I and II epitopes" by R.J.C. Bilderbeek et al. reports a rigorous bioinformatics analysis with the aim to address two key questions in the field of antigen presentation. This work is a continuation of their previous work on the over-presentation of transmembrane helices (TMH)-derived epitopes by MHC class I molecules (Fornt. Immunol. 20017), which is now extended to the MHC-II class. Their key finding is that TMH-derived epitopes irrespective of the origin of the protein, are presented more often for almost all human MHC-I and MHC-II haplotypes. An additional analysis of single nucleotide polymorphisms (SNPs) in the human proteome supports previous observations that TMHs are evolutionary more conserved than soluble protein domains, an observations that underpins the evolutionary advantage of their key finding.

The bioinformatics analysis has been performed using suitable methodology as clearly described in the text and the supporting information, and the statistical methods are correct. In fact, I find that the statistical analysis and cutoffs used in this study give rise to more realistic observations about the conservation score of TMHs than previous reports. Thus the observations drawn are supported by their analysis to the extent displayed by the statistics. The manuscript is very well written, in a concise but comprehensive manner, and the quality of the figures is fine. It should be noted that Fig. 2 was not included in the manuscript, probably due to a mislabeling of Fig. 3-4. This should be corrected, although the exact numbers can be found in Table S6.

Overall, the work is of high quality and has the merit of a very interesting notion that the field might have overlooked. Thus, albeit the work is not characterized by high level of novelty, I endorse its publication in Frontiers in Immunology with the hope that it will set the basis for further research in the field of antigen presentation.

Thanks for the appreciation of our work!

Minor comments:

The authors might want to comment on their observation that a lower percent of variability is explained by the hydrophobicity score of TMH-derived epitopes in MHC-II than in MHC-I, with regard to the discussion in section 4.2

Section 4.2 discusses why TMHs are evolutionarily more conververved and the role of the immune system in this. Figures 1b and 1c show that there is a correlation between MHC alleles presenting TMH-derived epitopes and preferring hydrophobic elements. Figure 1c shows that this correlation is weaker in MHC-II.

We fail to see a relation between these two sections strong enough to describe it. Apparently, we overlook something and would enjoy to hear it.

> L. 64: "[6] found that" can be written as "S. Tang et al. [6] found that"

Thanks, we changed the LaTeX to:

> For example, S. Tang et al. [**?**] found that, In addition, we have amended similar instances throughout the manuscript.

Check List

> a. Is the quality of the figures and tables satisfactory?
>
> • Yes

Great!

> b. Does the reference list cover the relevant literature adequately and in an unbiased manner?
>
> • No

We hope that our revised manuscript covers the literature in a more adequate fashion.

> c. Are the statistical methods valid and correctly applied? (e.g. sample size, choice of test)
>
> • Yes

Great!

> d. Are the methods sufficiently documented to allow replication studies?
>
> • Yes

Great!

Reviewer 3

- Independent review report submitted: 23 Sep 2021
- Interactive review activated: 26 Sep 2021
- Recommendation for the Editor: Substantial revision is required

> Bilderbeek et al investigate the role of transmembrane helices as sources for class I and class II MHC epitopes. This is an interesting question in MHC protein biology, with important implications for epitope prediction, vaccine design, and understanding of intracellular antigen processing and presentation pathways.

First and foremost, something must have gone wrong with the figures and their corresponding legends in the original submission. We apologize for this and have corrected this in our revised manuscript.

1) The authors claim to have shown that transmembrane helices (TMH) are important sources of MHC class I and class II epitopes. But they have not analyzed actual MHC epitopes, only predicted epitopes generated by MHC-peptide binding prediction algorithms. As pointed out repeatedly by the authors, both MHC-I and MHC-II proteins are well-known to bind hydrophobic peptides, and this is reflected in the prediction algorithms. What the authors have shown is that TMH are important sources of predicted epitopes. Whether or not such peptides actually are processed, presented at the cell surface, and recognized by T cells, has not been investigated. TMH might or might be processed into peptide efficiently, they might or might not be loaded well by the various MHC-I and MHC-II machinery, etc. The analyses of how epitope prediction favors TMHs is interesting but only part of the question.

As suggested, we have performed an analysis of epitopes taken from the IEDB (the Immune Epitope Database from NIH), confirming that peptides derived from transmembrane helices (TMHs) are more likely to be presented on MHC class I and II (new Figure 2). Furthermore, we mined the IEDB database for epitopes that evoke T cell responses (as shown in new Figure S4), which are also likelier to be derived from TMHs.

We also correlated the percentage of TMH-derived epitopes taken from 9-mers (for MHC-I) and 14-mers (for MHC-II) from a human reference proteome, with these percentages as obtained from the IEDB (as shown in new Figure S3), indicating that our findings are robust.

The authors do present an example of TMH that are known to be presented by MHC-I and MHC-II, from two peptide elution studies. These data indeed appear to show that TMH are found among the eluted peptides ("appears" because the figure is missing). However, the authors do not provide any analysis of whether the frequency of such epitopes is more or less than would be expected by chance, as they do in detail for the predicted epitopes, just the presence of TMH-derived epitopes in the peptidome is noted.

As discussed in the response above, we replaced the epitopes obtained from two elustion studies by epitope sequences from the IEDB.

In addition to the two peptidome studies used by the authors, many comprehensive immunopeptidomes are available for analysis. For example, the NIH Immune Epitope Database and Analysis Resource lists several hundred studies with many hundreds of thousands of epitopes. A detailed analysis of the frequency of TMH in experimentally

determined MHC-I and MHC-II peptidomes (rather than predicted binders) could provide the information sought by the authors.

Please see our comments above. As requested, we have performed the analysis with epitopes from the IEDB.

There are other major issues:

2) The figures are numbered incorrectly and don't correspond to the text. Figure 1 matches its legend, but Figure 2 appears to be missing. Figure 3 legend appears to correspond to Figure 2.TIF but with the panels scrambled. Figure 4 legend appears to correspond to Figure 3.TIFF, but the bold red line on each plot is not described, and indicated confidence intervals are missing.

We are sorry for these mistakes and have corrected them.

3) A key message from this would appear to be that TMH sequences are overlooked as a source of MHC-I and MHC-II epitopes, but this statement is not supported by any data. There are many published instances of TMH sequences as MHC-I epitopes, as noted by the authors themselves in their peptidome analysis. HLA-E is well-known to bind signal sequence peptides, which are a type of TM sequences.

The key message of the paper that TMHs are an overlooked source MHC-I and MHC-II epitopes comes from the fact that we predict that, due to the hydrophobic nature of TMHs, peptides derived from TMHs make good binders for the majority of MHC alleles. As discussed above, we now provide a new analysis of peptides from IEDB confirming our predictions. In addition, we now include a description of the presentation of signal peptides by HLA-E in our revised Discussion section. Please note that HLA-E is a non-canonical MHC-I form, whereas our study considers the canonical HLA-A, -B, and -C forms.

4) It is claimed that both that data suggest that both cytolytic and helper T cells respond to more to membrane proteins, but no T cell data at all are considered or analyzed. Presentation of MHC-I and MHC-II bound peptides is related to T cell responses, presentation does not imply T cell response because T cell selection has not been considered. If TMH sequences are highly conserved across organisms, and such TM-peptides are highly represented in thymic epithelium, one might expect that TM-specific T cells would be deleted or tolerized. So TMH-specific responses might be promoted or suppressed by increased TM peptide presentation.

We agree with the reviewer. Although T cell responses against TMH derived epitopes have been reported, we have no indication if more T cells respond to membrane proteins than expected by chance. We adapted the text to the

following: "Our bioinformatics predictions indicate that the TMH-derived epitope repertoire is larger than expected by chance for both MHC-I and MHC-II, regardless of the organism. Moreover, reanalysis of MHC-ligands from the IEDB database confirmed the presentation of TMH-derived epitopes. Therefor, it seems likely that TMH-derived epitopes would also result in enhanced T cell responses, although the conservation of TMHs might promote the deletion of T cells responsive to TMH-derived epitopes by central tolerance mechanisms."

Also some parts of manuscript are confusing:

5) The use of MHC haplotype to refer to a particular MHC allele is incorrect and confusing. An MHC haplotype is a set of HLA genes present on a single chromosome and therefore genetically linked. The peptide binding molecules are called MHC proteins or HLA gene products or (less frequently) MHC allotypes.

We agree that there was a mix up of nomenclature and have modified the manuscript accordingly.

6) The authors search for the "physiological reason" why peptides derived from TMHs are frequently presented is puzzling, as they state that it is already known that MHC protein prefer hydrophobic peptide and TMHs are hydrophobic. This would appear to be the reason. They seem to be asking how the preference of MHCs to bind TMHs impacts MHC cellular and immunological function and evolution, but those are consequences, not reasons.

We politely disagree. First, that a peptide is hydrophobic does not mean it stems from a TMH. Second, as the affinity of a peptide for MHC is generated by a few residues, their preference is not a prediction for the overall hydrophobic preference of the allele. Although it was previously reported that MHC-I in general seems to have a slight preference for hydrophobic residues, the correlation with TMHs was not yet addressed nor the consequences for the cellular and immunological function and evolution.

5) The section on T cells recognizing different parts of an immunogenic protein than B cells is confusing and unnecessary since this is already well appreciated.

We agree with the reviewer that this is a well known phenomena and have removed it from our revised manuscript.

6) Finally, the analyses of chance versus observed epitopes do not include any uncertainly estimates.

We have added the uncertainty estimates as requested.

Check List

a. Is the quality of the figures and tables satisfactory?

- No

We hope this is fixed now.

      b. Does the reference list cover the relevant literature adequately and in an unbiased manner?

- Yes

      c. Are the statistical methods valid and correctly applied? (e.g. sample size, choice of test)

- No

We hope this is fixed now.

      d. Are the methods sufficiently documented to allow replication studies?

- Yes

Reviewer 4

- Independent review report submitted: 26 Sep 2021
- Interactive review activated: 26 Sep 2021
- Recommendation for the Editor: The manuscript should be rejected

This reviewer recommended rejection of the manuscript on 26 Sep 2021. Discussions for this review are closed. Reason:

The manuscript contains fundamental errors that cannot be rectified through author revisions. Note to the authors:

The authors have shown that TMH/TMP should be overrepresented in MHC-associated peptidomes. However, they are overestimating the contribution of these structures in the ligand generation considering that, in some cases, the contribution of one residue of the TMH to the ligand is almost anecdotic even more if this contribution mainly affects N or C-terminal positions and not in anchor positions. To state that TMH are overrepresented in the MHC peptidome, the authors should have considered at least a minimum length where part of the TMH sequence binds to the main or secondary positions of MHC allele.

EVALUATION

1. There is a confusion, authors are working with allotypes instead haplotypes. They are considering the protein expressed by one HLA allele, so it is allotype or allomorph. For instance: HLA-A*02:01, HLA-B*27:05. The only If they were considering haplotypes, they must notice that an MHC haplotype considers

> a combination of more than one gene; i.e the DQ2.5, DQ2.2 and DQ7.5, in this case each haplotype is a combination of two specific alleles derived from DQA and/or DQB.

We agree that there has been a misconception in nomenclature and corrected the manuscript accordingly.

> 2. Why the authors stated that MHC alleles are slightly hydrophobic? Some of the allotypes they used are highly hydrophilic because they have highly polarized anchor positions.

If we interpret the question correctly, we believe our statement that most MHC alleles prefer hydrophobic epitopes is warranted by figures 1A and 1B, since these figures demonstrate that most MHC alleles have a positive hydrophobicity score.

> 3. It is not clear how the authors chose the MHC-I allotypes? It is explained only for MHC-II alleles. I assume for their representation in the population.

As requested, we have and revised the text to the following: Both MCH-I and MHC-II alleles were selected to have a high occurrence, where the alleles of MHC-I are the alleles representing the 13 supertypes with over 99.6% coverage of the populations MHC-I repertoire as defined by [1, 2] and the 21 MHC-II alleles, have a phenotypic frequency of 14% or more in the human population [3].

- [1] Lund, O., Nielsen, M., Kesmir, C. et al. Definition of supertypes for HLA molecules using clustering of specificity matrices. Immunogenetics 55, 797–810 (2004). https://doi.org/10.1007/s00251-004-0647-4
- [2] Sette, A., Sidney, J. Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. Immunogenetics 50, 201–212 (1999). https://doi.org/10.1007/s002510050594
- [3] Greenbaum, Jason, et al. "Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes." Immunogenetics 63.6 (2011): 325-335.

> 4. Is it significant the overlapping of one residue for MHC binding? Why not consider only those peptides that are really predicted and analyze them if they are really found in experimental data? Contribution of one residue of the TMH to the ligand is almost anecdotic even more if this contribution mainly affects N or C-terminal positions. It is logically to see these residues located near outside the lipid bilayer because they are more "available" to be processed independently of what the mechanistic process is. It would be nice to saw how the TMH contributes to every position of the ligand predicted compared to those experimentally confirmed.

We defined a TMH-derived epitope as any epitope that overlaps with at least one

residue with a TMH, because the underlying cellular mechanisms of presentation must differ from the canonical mechanisms for presentation of soluble proteins. As we explain in the Discussion section of our revised manuscript, the presentation of a TMH requires its extraction from the membrane even if just a single amino acid contributes to the epitope.

In addition, please note that this definition of a TMH-derived epitope does not alter the conclusion, as increasing the overlap equally reduces the chance that an epitope stems from a TMH. To prove this point, we did exactly the same analysis as shown in Figure 2a, yet with defining a TMH-derived epitope as an epitope that overlaps with a TMH for at least 2 residues, as shown in Figure S14. These two figures are highly similar. See Table S14 and S5, and S15 and S7 for the corresponding values.

> 5. The latter point is also more relevant to the case of MHC-II, where the anchor positions are "internal" and the extremes are more diverse and less conserved. Are the authors considering significant to state that TMH are overrepresented if an MHC-II-derived peptide has only one residue coming from a TMP/TMH which is almost "floating" in extreme of the binding groove?

Please see above. Increasing the overlap does not alter the conclusion.

> 6. To state that TMH are overrepresented in the MHC peptidome, the authors should have considered at least a minimum length where part of the TMH binds to the main or secondary positions of MHC allele.

Please see above. Increasing the overlap does not alter the conclusion.

> 7. Considering that the predicted peptides of the TMH are hydrophobic, are these values comparable to those obtained from peptide elution studies? Is it well correlated of what we know about their peptide preferences of each MHC in this study? There is no explanation of how MHC alleles with polar binding preferences bind part of the TMH.

As suggested, we have performed an analysis of epitopes taken from the IEDB (the Immune Epitope Database from NIH), confirming that peptides derived from TMHs are more likely to be presented on MHC class I and II as predicted (new Figure 2). Furthermore, we mined the IEDB database for epitopes that evoke T cell responses (as shown in new Figure S4), which are also likelier to be derived from TMHs.

We also correlated the percentage of TMH-derived epitopes taken from 9-mers (for MHC-I) and 14-mers (for MHC-II) from a human reference proteome, with these percentages as obtained from the IEDB (as shown in new Figure S2), indicating that our findings are robust.

In Bianchi. et. al. 2017 we mapped the distance of the predicted binders to

the TMH regions and showed that for especially MHC alleles that bind polar regions a peak was found at the interface of TMH, showing that these structural requirements make preferred antigens.

8. Conservation of the TMH, or it is lack of variation, has been documented many years ago by different biochemical experiments and it is mainly explained by the nature of the few amino acids that can compose the hydrophobic helix. However, the authors could try to explain if the SNPs identified by them could affect the presentation of some predicted or experimental epitopes.

We agree with the reviewer that it is an interesting idea to determine if SNPs that disrupt a TMH also prevent MHC presentation, and we actually explored this [1]. However, we abandoned this approach, because it became computationally intractable.

- [1] https://github.com/richelbilderbeek/bbbq_article/blob/v1.1/p_transition/create_dataset.R

9. Figure 2 has not explanation of what it is included in each graph. Which one is from MHC-I? MHC-II? Is all together?

We apologize if this is unclear and now state this in the legend.

10. There is no Figure 4 in the manuscript.

We apologize for the mistake and have corrected this.

11. This study would give more information if it was based on all the experimental data available in public databases, as IEDB, instead the proteome and predictive algorithms. I think authors must analyze the data using a more immunologically point of view.

Please see above, we have performed the requested analysis of the IEDB.

Check List

a. Is the quality of the figures and tables satisfactory?

- No

We hope this is now fixed.

b. Does the reference list cover the relevant literature adequately and in an unbiased manner?

- Yes

c. Are the statistical methods valid and correctly applied? (e.g. sample size, choice of test)

- Yes

d. Are the methods sufficiently documented to allow replication studies?

- Yes