

Molecular Mimicry as a Mechanism of Viral Immune Evasion and Autoimmunity

Corresponding Author: Dr Esther Melamed

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

In this manuscript, Melamed and colleagues analyze the protein sequences of 134 human infecting viruses for evidence of molecular mimicry of cellular host proteins. The authors focused their analysis on short linear sequences (8-, 12- or 18-amino acids) in the viruses and human genome. A subset of viruses, primarily chronically infecting viruses, especially from the herpesvirus and poxvirus families, were found to contain linear epitopes similar to human protein sequences (3 or less amino acid mismatches), which was taken as a sign of molecular mimicry. Mimicry was confirmed using sequence reversal and scrambling controls. Poxvirus mimic sequences were on average significantly longer than herpesvirus mimic sequences. The length distribution of linear mimic peptides was quantified (with some >30 aa identified), and instances of viral peptides mimicking sequences found in multiple human proteins were also found. The authors used the KEGG database to identify pathways that the mimicked human proteins contributed to, primarily cellular replication and inflammation. The Human Protein Atlas was also interrogated to ask whether mimicked proteins were expressed in specific tissues or cell types, but this was not found to be the case. However, the Y chromosome had fewer than average mimicked sequences, potentially reflecting a reduced benefit of mimicking proteins expressed only in males. Interestingly, latent stage viral proteins displayed more mimicry than lytic stage proteins in EBV, and sequences mimicked by herpesviruses were enriched in thymic cells. Lastly, multiple sclerosis (MS) auto-antibody sequences were found to be more likely to contain EBV mimicry, consistent with a previous study identifying an MS patient subset with autoantibodies bearing a motif homologous to sequences found in the envelope glycoprotein and BRRF2 protein from EBV.

This well written article provides clear evidence of molecular mimicry in linear epitopes in herpesviruses and poxviruses. The discussion of the potential evolutionary and disease implications is thoughtful and insightful. The key limitation of this study is that it is purely *in silico*, so the hypotheses proposed herein therefore remain untested experimentally. However, the datasets and code provided by the author on GitHub should constitute a useful resource for future experimental studies in the wider virology and immunology community.

Main Comments

A DOI should be generated for the Github repository to facilitate access and distribution. This can be done by releasing the repository on GitHub as a "Release" and selecting the option to generate a DOI via Zenodo.

The impact of the study would be increased if the authors could extend their analysis to ask whether endogenous virus antigens (or other neoantigens) that are more highly expressed in chronic disease, including multiple sclerosis and cancer – both of which have been associated with neoantigen or retroelement expression – also mimic host proteins.

Minor comments

1. Lines 182-3. While reversing protein sequences are probably more likely to maintain predicted secondary structures than scrambling of the sequence, there is no evidence that protein folds and tertiary structures, which will ultimately determine evolutionary pressures, would be maintained in a meaningful way following sequence reversal. Hence this statement should be removed.

2. It would be appropriate to cite some additional previously documented examples of host mimicry in poxviruses from the

literature. For example, Boys et al (10.1016/j.celrep.2023.112878) showing that a domain with homology to gasdermins is present in poxviruses.

3. Line 413. Since the mimicry signal detected in this study involve mainly only two virus families, the authors should temper the phrase "extremely viable".

4. Limitations section. This section should be shorted to half its current length and integrated into the Discussion. Currently, only half of the Limitations section is about limitations of the study.

5. The Conclusion section is repetitive and dispensable given the long Discussion. Also papers in Nat Commun. rarely have a Conclusions section. This section should be deleted.

6. Fig. 1B. The color scale could be modified to emphasize how much herpes- and poxviruses stand out relative to the other families (eg make 5 on mimicry scale more red).

7. Fig. 3D. The number of human genes mimicked per viral mimic sequence is actually highest in Picornaviridae. Notably, there is also an autoimmune disease link for this virus family: type 1 diabetes onset is more likely after a picornavirus (specifically enterovirus) infection. This could be added to the relevant part of the Discussion.

8. Fig. 3E. How was the function for the curve extrapolations chosen? The fit looks unconvincing.

Reviewer #2

(Remarks to the Author)

This study examines viral mimicry of host proteins in the form of short linear subsequences of viral proteins with high-identity matches to human proteins. The suggestion is that these could be evolutionarily favored, for example to suppress immune surveillance of the virus by exploiting negative selection of T cells reactive to shared (self-like) epitopes. To assess statistical significance, three different shuffled versions of the viral proteomes are also searched for examples of mimicry, with viruses in the herpes and pox virus families showing elevated levels of mimicry relative to these shuffled controls. Mimicked proteins show biases with respect to functional pathway and chromosomal location, and EBV-mimicked motifs are enriched in the linear targets of auto-antibodies of MS patients.

This is an interesting study, but I think additional controls are needed to rule out alternative explanations for mimicry. The big one, in my mind, is overall sequence homology. It will be important to assess the degree to which mimicry could be explained by conservation of shared structural or functional features, with overall sequence similarity being the simplest example. I guess one might find elevated "mimicry" when comparing many other species' proteomes to the human proteome, but this would likely be explained by sequence similarity. Even for distantly related organisms (e.g., non-pathogenic viruses or bacteria), strong functional constraints can induce apparent mimicry. The example that the authors themselves highlight in Figure 3 is a great example: this is a "Walker A" or "P loop" phosphate binding motif (https://en.wikipedia.org/wiki/Walker_motifs). One doesn't need to postulate mimicry to explain its conservation in viral sequences (assuming they also have a need to bind to phosphates). I think the authors need to stratify their examples of mimicry on the basis of (1) global sequence similarity and (2) functional site conservation (using for example the ELM database <http://elm.eu.org/>; I guess this would catch examples like the Walker A motif). It could also be helpful to look at mimicry in non-pathogenic organisms with large genomes like the herpes viruses. It could be that in these large viral genomes, there is greater sharing of homologous proteins and functional sites than for smaller, more constrained viruses. Or the reverse: look at whether these human viruses appear to be "mimicking" the mouse genome, for example.

I don't understand the justification for the Figure 4 legend heading: "Viral mimicry targets ****vulnerable**** biological pathways." I also couldn't understand the measure being plotted Figure 5C; perhaps the description could be expanded a bit.

I found a number of typos:

L160 "separetly for each mismatch"

L173 "reveresed viral "

L258 "chromosome"

L323 "motiff found in EBV's Envelope Glyocprotein" (2 typos)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors's responses and their edits and clarifications added to the manuscript fully address my questions and concerns.

Reviewer #2

(Remarks to the Author)

My concerns have been addressed. I appreciate the authors' attention to the reviewer comments.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

We thank the reviewers for their time and constructive comments on our manuscript, entitled “Molecular Mimicry as a Mechanism of Viral Immune Evasion and Autoimmunity.” We feel that the manuscript has significantly improved as a result of the comments provided by the reviewers, and we added an acknowledgement to the reviewers in the manuscript.

We have carried out additional analyses and have revised our manuscript to address the reviewers' specific recommendations, as detailed in the point-by-point response to reviewers' comments below. Corresponding modifications in the manuscript are denoted in **red text** below, and tracked in the manuscript.

Reviewer #2 (Remarks to the Author)

In this manuscript, Melamed and colleagues analyze the protein sequences of 134 human infecting viruses for evidence of molecular mimicry of cellular host proteins. The authors focused their analysis on short linear sequences (8-, 12- or 18-amino acids) in the viruses and human genome. A subset of viruses, primarily chronically infecting viruses, especially from the herpesvirus and poxvirus families, were found to contain linear epitopes similar to human protein sequences (3 or less amino acid mismatches), which was taken as a sign of molecular mimicry. Mimicry was confirmed using sequence reversal and scrambling controls. Poxvirus mimic sequences were on average significantly longer than herpesvirus mimic sequences. The length distribution of linear mimic peptides was quantified (with some >30 aa identified), and instances of viral peptides mimicking sequences found in multiple human proteins were also found. The authors used the KEGG database to identify pathways that the mimicked human proteins contributed to, primarily cellular replication and inflammation. The Human Protein Atlas was also interrogated to ask whether mimicked proteins were expressed in specific tissues or cell types, but this was not found to be the case. However, the Y chromosome had fewer than average mimicked sequences, potentially reflecting a reduced benefit of mimicking proteins expressed only in males. Interestingly, latent stage viral proteins displayed more mimicry than lytic stage proteins in EBV, and sequences mimicked by herpesviruses were enriched in thymic cells. Lastly, multiple sclerosis (MS) auto-antibody sequences were found to be more likely to contain EBV mimicry, consistent with a previous study identifying an MS patient subset with autoantibodies bearing a motif homologous to sequences found in the envelope glycoprotein and BRRF2 protein from EBV.

This well written article provides clear evidence of molecular mimicry in linear epitopes in herpesviruses and poxviruses. The discussion of the potential evolutionary and disease implications is thoughtful and insightful. The key limitation of this study is that it is purely in silico, so the hypotheses proposed herein therefore remain untested experimentally. However, the datasets and code provided by the author on GitHub should constitute a useful resource for future experimental studies in the wider virology and immunology community.

Main Comments

A DOI should be generated for the Github repository to facilitate access and distribution. This can be done by releasing the repository on GitHub as a “Release” and selecting the option to generate a DOI via Zenodo.

We thank the reviewer for informing us of this process, which also tracks version releases, which we agree is important. We have made a doi for the github which is now included both below and in the manuscript.

DOI: [10.5281/zenodo.11411891](https://doi.org/10.5281/zenodo.11411891)

The impact of the study would be increased if the authors could extend their analysis to ask whether endogenous virus antigens (or other neoantigens) that are more highly expressed in chronic disease, including multiple sclerosis and cancer – both of which have been associated with neoantigen or retroelement expression – also mimic host proteins.

We thank the reviewer for this excellent suggestion. We have revised the discussion section to expand upon the possible relevance of our findings in relation to human endogenous retroviruses and neoantigens.

We are limited in our ability to run meaningful *in silico* analyses due an incomplete understanding of the biology on these topics. For example, in the case of human endogenous retroviruses (HERVs), they are known to comprise ~8% of the human genome and some HERVs are known to be expressed in the human thymus⁵²⁻⁵⁴. Thus, it has been suggested that there is incomplete central tolerance for HERVs, with only the HERVs that continually have basal levels of expression being capable of developing immune tolerance⁵⁵. In line with this hypothesis, in autoimmunity and cancer, it is the non-basally expressed HERVs that aggravate and generate adaptive immune responses in disease⁵²⁻⁵⁴. Thus, we would hypothesize that mimicry may only be relevant for the basally expressed HERVs. Of note however, a total of 79 HERVs are included in the *homo sapiens* proteome by Uniprot (as despite their naming, they are generally classified as human proteins). However, when filtering out Eukaryotic Linear Motifs (ELMs, suggested by reviewer 3), there are only 4 instances of mimicry (12mers with 3 or less mismatches) in all of the 79 HERV proteins included in human proteome from UniProt. Thus, while we also believe that HERVs provide an interesting avenue for future directions, an *in silico* study would require a more robust understanding of the HERVs, particularly which ones are basally expressed, and a more complete database of HERV protein sequences.

Similarly, we attempted an analysis to evaluate whether there were elevated rates of mimicry in specific isoforms of proteins, and particularly whether it could link to known autoimmune disease. However, in the process of this analysis, we quickly realized the limitations of databases such as UniProt which do not have a complete record of all protein isoforms, particularly those only found in specific diseases.

“Molecular mimicry may also play a role in human endogenous retroviruses (HERVs) which comprise ~8% of the human genomes and are frequently expressed in many chronic diseases, such as autoimmunity and cancer^{52,53}. However, due to HERVs being a part of the human genome and being expressed in the thymus⁵⁴, there is some central tolerance, although incomplete, against HERVs⁵⁵. Thus, in future studies, it will be critical to elucidate how adaptive tolerance is maintained against different HERVs.”

52. Grabski, D. F., Hu, Y., Sharma, M. & Rasmussen, S. K. Close to the Bedside: A Systematic Review of Endogenous Retrovirus and Their Impact in Oncology. *J Surg Res* 240, 145–155 (2019).

53. Balada, E., Vilardell-Tarrés, M. & Ordi-Ros, J. Implication of Human Endogenous Retroviruses in the Development of Autoimmune Diseases. *International Reviews of Immunology* 29, 351–370 (2010).

54. Passos, V., Pires, A. R., Foxall, R. B., Nunes-Cabaço, H. & Sousa, A. E. Expression of Human Endogenous Retroviruses in the Human Thymus Along T Cell Development. *Front. Virol.* 2, (2022).

55. Alcazer, V., Bonaventura, P. & Depil, S. Human Endogenous Retroviruses (HERVs): Shaping the Innate Immune Response in Cancers. *Cancers (Basel)* 12, 610 (2020).

Molecular mimicry may also play a critical role in cancer neoantigens, defined new sequences in human proteins due to the high rate of mutation in cancer cells. Recently, CAR-T therapies combined with immune checkpoint inhibitors have shown promising outcomes in targeting personalized neoantigens from sequencing individual tumors. However, auto-immunity is a common side effect, largely attributed to the immune checkpoint inhibitors. Yet, depending on the neo-antigen, potentially a similar cross-reactivity to the viruses may be observed and explain portions of the increase in autoimmunity following this therapy.

We agree with the insightful comment by the reviewer regarding the hypothesis of autoimmunity post CAR-T /immune checkpoint inhibitor therapy. We have added text in the Discussion section to highlight the relevance of our findings to this emerging and important topic.

“The relevance of our findings may also extend to explain the emergence of autoimmunity post CAR-T and immune checkpoint inhibitor therapies^{56–61}. Recently advances have enabled more personalized treatments, with CAR-T therapy capable of targeting personalized “neo-antigens” (novel protein sequences that occur as a result of genetic mutations or incorrect protein splicing) from sequenced tumor biopsies^{62–64}. Due to the founding sequence being human, neo-antigens may pose a similar threat to molecular mimicry, and result in cross-reactive immune responses. Furthermore, the concomitant use of checkpoint inhibitors with CAR-T may exacerbate cross-reactivity due to the weakened lymphocyte activation thresholds^{65,66}. Similarly, neo-antigens may exist in unique disease-specific isoforms of proteins^{67,68}, often occurring at novel splice junctions, and thereby contribute to molecular mimicry in different autoimmune syndromes.”

56. Khalid, F. et al. Neurological Adverse Effects of Immune Checkpoint Inhibitors and Chimeric Antigen Receptor T-Cell Therapy. *World J Oncol* 14, 109–118 (2023).

57. Marini, A. et al. Neurologic Adverse Events of Immune Checkpoint Inhibitors: A Systematic Review. *Neurology* 96, 754–766 (2021).
58. Gritsch, D. & Valencia-Sanchez, C. Drug-related immune-mediated myelopathies. *Front Neurol* 13, 1003270 (2022).
59. Hottinger, A. F. Neurologic complications of immune checkpoint inhibitors. *Curr Opin Neurol* 29, 806–812 (2016).
60. Oliveira, M. C. B., de Brito, M. H. & Simabukuro, M. M. Central Nervous System Demyelination Associated With Immune Checkpoint Inhibitors: Review of the Literature. *Front Neurol* 11, 538695 (2020).
61. Ramos-Casals, M. et al. Immune-related adverse events of checkpoint inhibitors. *Nat Rev Dis Primers* 6, 38 (2020).
62. Leko, V. et al. Identification of neoantigen-reactive T lymphocytes in the peripheral blood of a patient with glioblastoma. *J Immunother Cancer* 9, e002882 (2021).
63. Zhou, W., Yu, J., Li, Y. & Wang, K. Neoantigen-specific TCR-T cell-based immunotherapy for acute myeloid leukemia. *Experimental Hematology & Oncology* 11, 100 (2022).
64. Zhu, Y., Qian, Y., Li, Z., Li, Y. & Li, B. Neoantigen-reactive T cell: An emerging role in adoptive cellular immunotherapy. *MedComm* (2020) 2, 207–220 (2021).
65. Grosser, R., Cherkassky, L., Chintala, N. & Adusumilli, P. S. Combination Immunotherapy with CAR T Cells and Checkpoint Blockade for the Treatment of Solid Tumors. *Cancer Cell* 36, 471–482 (2019).
66. Yoon, D. H., Osborn, M. J., Tolar, J. & Kim, C. J. Incorporation of Immune Checkpoint Blockade into Chimeric Antigen Receptor T Cells (CAR-Ts): Combination or Built-In CAR-T. *International Journal of Molecular Sciences* 19, 340 (2018).
67. Park, J. & Chung, Y.-J. Identification of neoantigens derived from alternative splicing and RNA modification. *Genomics Inform* 17, e23 (2019).
68. Huang, P., Wen, F., Tuerhong, N., Yang, Y. & Li, Q. Neoantigens in cancer immunotherapy: focusing on alternative splicing. *Front. Immunol.* 15, (2024).

Minor comments

1. Lines 182-3. While reversing protein sequences are probably more likely to maintain predicted secondary structures than scrambling of the sequence, there is no evidence that protein folds and tertiary structures, which will ultimately determine evolutionary pressures, would be maintained in a meaningful way following sequence reversal. Hence this statement should be removed.

We agree with the reviewer's suggestion and have revised our statement, as per below.

“However, this permutation strategy does not fully capture the fact that protein sequences are not truly random. Thus, to evaluate the robustness of the permuted results, as an alternative strategy, we also evaluated the frequency of mimicry in viral proteomes with reversed protein sequences.”

2. It would be appropriate to cite some additional previously documented examples of host mimicry in poxviruses from the literature. For example, Boys et al

(10.1016/j.celrep.2023.112878) showing that a domain with homology to gasdermins is present in poxviruses.

We thank the reviewer for the citation and we have added it to the manuscript along with additional citations to other examples of host mimicry in poxviruses (list of citations below).

41. Boys, I. N., Johnson, A. G., Quinlan, M. R., Kranzusch, P. J. & Elde, N. C. Structural homology screens reveal host-derived poxvirus protein families impacting inflammasome activity. *Cell Rep* **42**, 112878 (2023).

42. Albarnaz, J. D. *et al.* Molecular mimicry of NF- κ B by vaccinia virus protein enables selective inhibition of antiviral responses. *Nat Microbiol* **7**, 154–168 (2022).

43. Howard, J., Justus, D. E., Totmenin, A. V., Shchelkunov, S. & Kotwal, G. J. Molecular mimicry of the inflammation modulatory proteins (IMPs) of poxviruses: evasion of the inflammatory response to preserve viral habitat. *Journal of Leukocyte Biology* **64**, 68–71 (1998).

44. Gubser, C. *et al.* A New Inhibitor of Apoptosis from Vaccinia Virus and Eukaryotes. *PLOS Pathogens* **3**, e17 (2007).

45. Mansur, D. S. *et al.* Poxvirus Targeting of E3 Ligase β -TrCP by Molecular Mimicry: A Mechanism to Inhibit NF- κ B Activation and Promote Immune Evasion and Virulence. *PLOS Pathogens* **9**, e1003183 (2013).

3. Line 413. Since the mimicry signal detected in this study involve mainly only two virus families, the authors should temper the phrase "extremely viable".

We agree and have adjusted the language.

"Thus, despite the potential drawbacks of mimicry, its usage on a micro scale may prove viable."

4. Limitations section. This section should be shorted to half its current length and integrated into the Discussion. Currently, only half of the Limitations section is about limitations of the study.

We thank the reviewer for their comment and have revised the limitations paragraph and integrated some of the comments into the discussion.

5. The Conclusion section is repetitive and dispensable given the long Discussion. Also papers in Nat Commun. rarely have a Conclusions section. This section should be deleted.

We appreciate the reviewer pointing out the formatting changes required for Nature Communications. We have removed the conclusion section.

6. Fig. 1B. The color scale could be modified to emphasize how much herpes- and poxviruses stand out relative to the other families (eg make 5 on mimicry scale more red).

We agree with the reviewer and have adjusted the color scale (in Figure 1B, Figure 2B, and Figure S5) per the reviewer's suggestion (while also ensuring we maintain red/green color blindness readability).

7. Fig. 3D. The number of human genes mimicked per viral mimic sequence is actually highest in Picornaviridae. Notably, there is also an autoimmune disease link for this virus family: type 1 diabetes onset is more likely after a picornavirus (specifically enterovirus) infection. This could be added to the relevant part of the Discussion.

We thank the reviewer for recommending addition of the connection between Picornaviridae and type 1 diabetes, which we agree is important and interesting. We have added this point to the discussion.

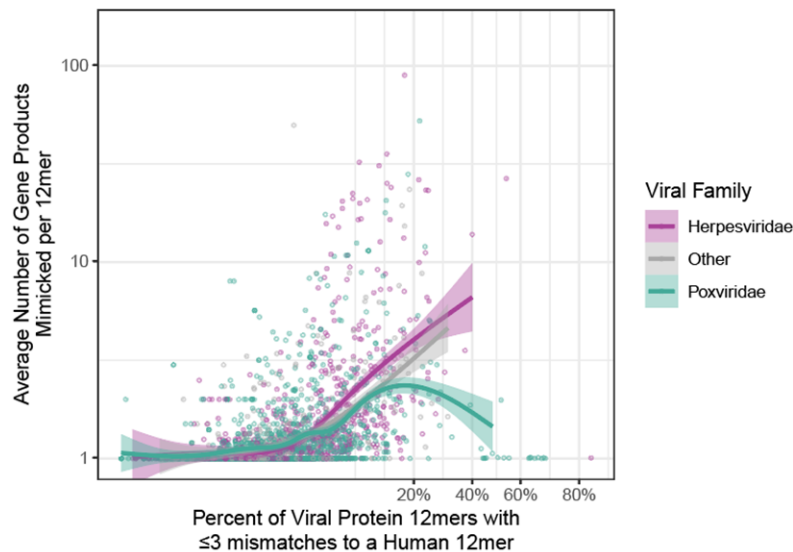
“Similarly to Herpesviridae, members of the Picornaviridae family also have a greater percentage of mimicry that targets human protein motifs compared to other viral families. As Picornaviridae infections often precede initiation of auto-immune diseases, such as type-I diabetes³⁷, mimicry by these viruses to human protein motifs may carry a greater autoimmune risk and explain why both Herpesviridae and Picornaviridae are epidemiologically more associated with autoimmune diseases.”

37. Oikarinen, M. et al. Enterovirus Infections Are Associated With the Development of Celiac Disease in a Birth Cohort Study. *Front Immunol* 11, 604529 (2020).

8. Fig. 3E. How was the function for the curve extrapolations chosen? The fit looks unconvincing.

We appreciate the reviewer’s comment regarding the fit of Fig. 3E. The fit is from a generalized additive model, with the trendlines representing the fit of the model and the shaded region is the 95% confidence interval. We agree with the reviewer and have changed the scaling of the axes as well as limited the range of the model to better represent the fit of this model.

We also note that the fit of the model is not a critical message of the graph, but rather used to help visualize the relationship between the two variables (i.e. average number of gene products mimicked per 12mer and percent of viral proteins). Furthermore, the graph reveals that our technique captures the previously known full protein mimicry (the data points near the x axis approaching 50-80%). Interestingly, we see that as viral proteins dedicate slightly more of their protein to mimicry (in the range of 10-20% of the protein), they begin to target human motifs more commonly, which is true for Herpesviruses, Poxviruses, and other viral families.



Revised Figure 3E - Average number of host gene products mimicked, plotted against the percent of the viral protein that participates in mimicry (defined as a 12mer with 3 or less mismatches). Trendlines represent fit and 95% confidence interval from a generalized additive model.

Reviewer #3 (Remarks to the Author)

This study examines viral mimicry of host proteins in the form of short linear subsequences of viral proteins with high-identity matches to human proteins. The suggestion is that these could be evolutionarily favored, for example to suppress immune surveillance of the virus by exploiting negative selection of T cells reactive to shared (self-like) epitopes. To assess statistical significance, three different shuffled versions of the viral proteomes are also searched for examples of mimicry, with viruses in the herpes and pox virus families showing elevated levels of mimicry relative to these shuffled controls. Mimicked proteins show biases with respect to functional pathway and chromosomal location, and EBV-mimicked motifs are enriched in the linear targets of auto-antibodies of MS patients.

This is an interesting study, but I think additional controls are needed to rule out alternative explanations for mimicry. The big one, in my mind, is overall sequence homology. It will be important to assess the degree to which mimicry could be explained by conservation of shared structural or functional features, with overall sequence similarity being the simplest example. I guess one might find elevated "mimicry" when comparing many other species' proteomes to the human proteome, but this would likely be explained by sequence similarity. Even for distantly related organisms (e.g., non-pathogenic viruses or bacteria), strong functional constraints can induce apparent mimicry. The example that the authors themselves highlight in Figure 3 is a great example: this is a "Walker A" or "P loop" phosphate binding motif (https://en.wikipedia.org/wiki/Walker_motifs) . One doesn't need to postulate mimicry to

explain its conservation in viral sequences (assuming they also have a need to bind to phosphates). I think the authors need to stratify their examples of mimicry on the basis of (1) global sequence similarity and (2) functional site conservation (using for example the ELM database <http://elm.eu.org/>; I guess this would catch examples like the Walker A motif). It could also be helpful to look at mimicry in non-pathogenic organisms with large genomes like the herpes viruses. It could be that in these large viral genomes, there is greater sharing of homologous proteins and functional sites than for smaller, more constrained viruses. Or the reverse: look at whether these human viruses appear to be "mimicking" the mouse genome, for example.

We greatly appreciate the reviewer's insightful comments about the evolutionary angle of molecular mimicry, and are currently working on a new manuscript to address this topic in greater detail that this topic deserves.

Our current paper focuses on the potential immunological consequences of mimicry, and thus evaluates this phenomenon from the perspective of the immune system regardless of the specific evolution of the similar sequences (i.e. as long as the sequences look similar, there is possibility for immune cross-reactivity). The main evidence we originally presented, in addressing whether mimicry is truly an evolutionary feature independent of functional sites, was the permutation experiments and a supplemental analysis demonstrating similar distribution of amino acid frequency between all the viruses in our viral cohort (Supplemental Figure 2).

Even though our paper's main focus is on immunological consequences of mimicry with application to autoimmune diseases, we agree with the reviewer that the evolutionary angle is quite interesting and thus have added several further analyses to expand upon this topic.

- 1) To address the reviewer's question about the eukaryotic linear motifs (ELM) driving the significances of our findings, we have repeated the analysis of Figure 1, this time filtering out any k-mer that contained an ELM reported in the ELM database suggested by the reviewer. We took a stringent approach and did not count any k-mer as a mimic if it contained an ELM (even if only a single amino acid in the k-mer was a part of the ELM). This analysis supported our previous findings, and in fact revealed a stronger association with Poxviruses and Herpesviruses, with Herpesviruses overtaking Poxviruses.

We have added this analysis to Supplemental Figure 1 with accompanying text in the methods and results. We choose not to replace Figure 1 with this analysis due to the potential immune cross-reactivity of ELMs still being relevant to autoimmunity. Furthermore, as outlined in the ELM database, some of these motifs are known functional mimics that interfere with host cell signaling (e.g. caspase cleavage sites) which is important for the functional enrichment analyses conducted in Figure 4.

"To confirm that these results were not driven by embedded eukaryotic linear motifs (e.g. ATP docking motif), we leveraged the Eukaryotic Linear Motif (ELM) database²⁹ to repeat the analysis but exclude any k-mer that contained even a single amino acid of an ELM (Figure S1D-I). This analysis replicated our findings and demonstrated Herpesviruses and

Poxviruses were significantly more likely to contain mimicry under 8, 12, and 18 length k-mers, even when filtering for functional protein motifs (Table S3)."

Eukaryotic Linear Motif Filtered Results

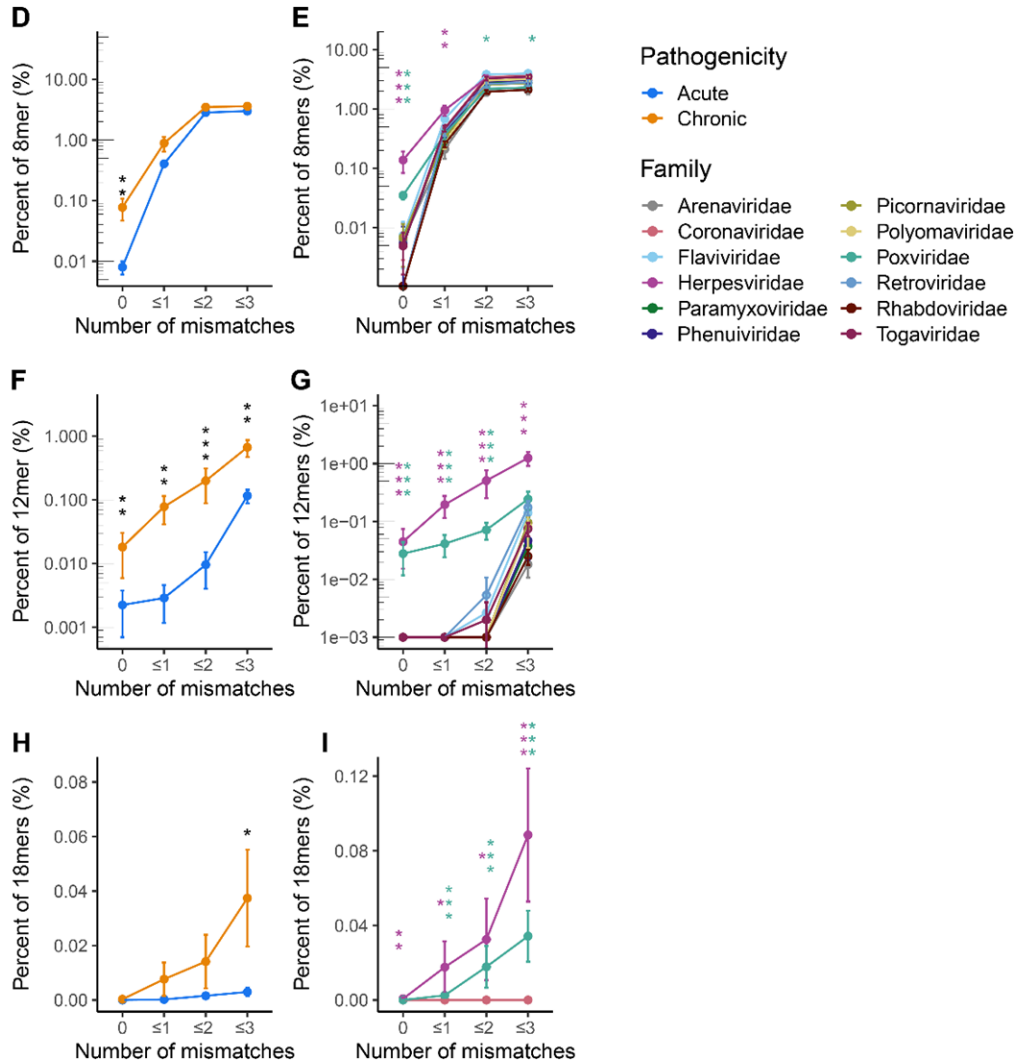


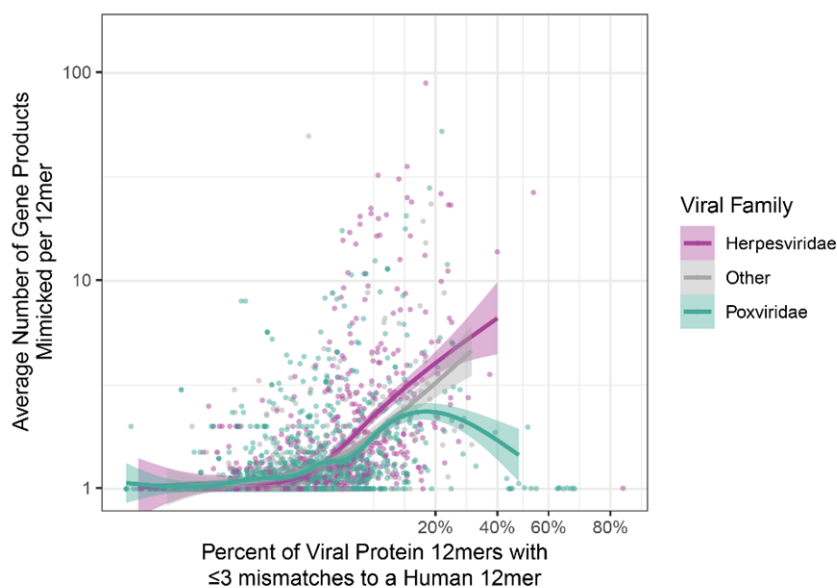
Figure S1 (panels D-I). Percent of viral 8mers with 3 or less mismatches to a human 8mer grouped by **D)** pathogenicity (chronically vs acutely infecting) and **E)** viral family while not counting 8mers that meet mimicry criteria if they contain a portion of a eukaryotic linear motif (though the ELM containing 8mers still contribute to the denominator). Percent of viral 12mers with 3 or less mismatches to a human 12mer grouped by **F)** pathogenicity (chronically vs acutely infecting) and **G)** viral family while not counting 12mers that meet mimicry criteria if they contain a portion of a eukaryotic linear motif (though the ELM containing 12mers still contribute to the denominator). Percent of viral 18mers with 3 or less mismatches to a human 18mer grouped by **H)** pathogenicity (chronically vs acutely infecting) and **I)** viral family while not counting 18mers that meet mimicry criteria if they contain a portion of a eukaryotic linear motif (though the ELM containing 18mers still contribute to the denominator). Viral families only shown in E, G, and I if containing at least 5 species. (Black stars indicate comparisons between

chronic and acute viruses, magenta and green stars indicate comparisons of the Herpesviruses and Poxviruses against all other viruses. Kruskal-Wallis test was used for multigroup comparison, and Wilcoxon rank-sum test for pairwise comparison. Error bars denote mean $\pm$ standard error of the mean. All p values adjusted for multiple hypothesis testing using Benjamini-Hochberg corrections (* p.adj $\leq$ 0.05, ** p.adj $\leq$ 0.01, *** p.adj $\leq$ 0.001)).

- 2) We interpreted the reviewer's comment about global sequence similarity to be a question about overall protein homology, due to similar functions between a viral protein and its human counterpart.

To address the reviewer's concern about global sequence similarity, we have revised text in the manuscript to address this critical topic, better highlighting the analyses in both Figure 3E and Supplemental Figure 2. In particular, the revised Figure 3E demonstrates that for most viral proteins, a vast majority have less than 20% of their 12mers with ≤ 3 mismatches to any human 12mer. This figure also shows the known instances of proteins with near complete sequence homology (the data points near the x axis approaching 50-80%). If global sequence similarity were extremely prevalent, we might expect a higher percentage of viral proteins to have detected mimicry (i.e. more points shifted towards the right-hand side of the x axis). Additionally, current Figure S2 expands on this topic and also demonstrates that the viruses in our cohort have very similar amino acid composition between themselves.

"Furthermore, the observed elevation in mimicry across diverse viruses was not limited to specific viral proteins (e.g. known virokines) but rather was more broadly distributed throughout the proteomes (Figure S2A-B). Additionally, relative amino acid usage did not greatly differ between viruses and was unlikely to explain the significant differences in mimicry (Figure S2C). Thus, most of the mimicry observed was not due to global sequence similarity, rather spread throughout proteins at a low frequency, except in the case of a few known proteins which have been adopted and modified by the virus from the host."



Revised Figure 3E - Average number of host gene products mimicked, plotted against the percent of the viral protein that participates in mimicry (defined as a 12mer with 3 or less mismatches). Trendlines represent fit and 95% confidence interval from a generalized additive model.

- 3) We agree with the reviewer that the specificity of viral mimicry to the human host is a very important and exciting evolutionary question. We are currently in the process of preparing a second manuscript on the evolution of mimicry in Herpesviruses, as this topic requires more substantial space and analysis to adequately answer.

Nevertheless, per the reviewer's suggestion, we wanted to share some of our analyses in the response here (Figure R1). Specifically, in response to the question regarding molecular mimicry to the mouse, we repeated our analysis against the mouse proteome instead of the human proteome. Interestingly, this analysis (Figure below) demonstrates that herpesviruses, in particular, have a preference to human specific mimicry (when evaluating the ratio of the rate of mimicry towards the human vs the mouse proteome, while correcting for the difference in length between each proteome).

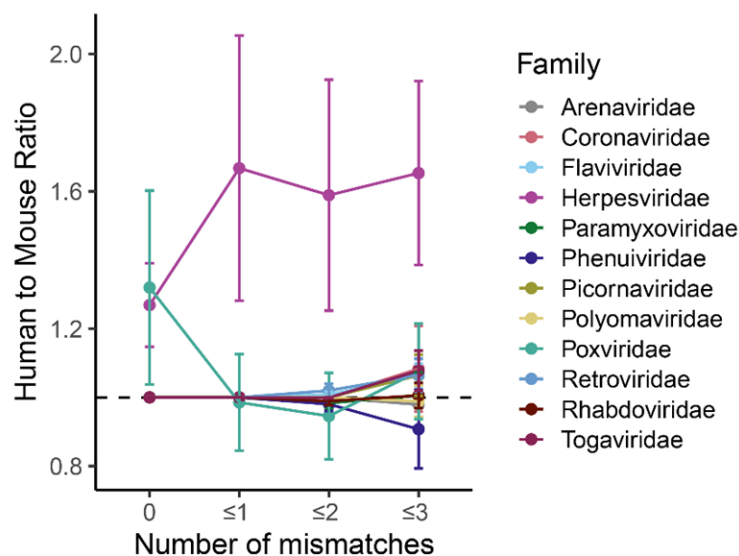


Figure R1 – Human to mouse ratio of rate of 12mer mimicry for 0, ≤1, ≤2, and ≤3 mismatches for human-infecting viruses reveals elevation of human specific mimicry in Herpesviruses. Ratio calculated as the percentage of viral 12mers meeting mimicry criteria for human proteins over the rate to the mouse proteome, while correcting for the difference in the length of the proteome between mouse and human.

I don't understand the justification for the Figure 4 legend heading: "Viral mimicry targets ****vulnerable**** biological pathways." I also couldn't understand the measure being plotted Figure 5C; perhaps the description could be expanded a bit.

We thank the reviewer for this suggestion and have revised the heading on Figure 4 to be

clearer. Furthermore, we have also revised other manuscript figure descriptions to increase clarity (particularly Figure 5C).

“Figure 4 - Viral mimicry targets critical cellular pathways of cellular replication and inflammation and avoids proteins from the Y chromosome.”

Percent difference of mimics expressed in mTEC, CD19+ B-cells, CD141+ dendritic cells, CD123+ dendritic cells, or “any” of these cells in the human thymus (all antigen presenting cells), compared to the percentage of all human genes expressed by the cell type (as determined by RNA expression in Gabrielsen et al.²⁵). Adjusted p-value calculated using Fisher exact test.

I found a number of typos:

L160 "separetly for each mismatch"

L173 "reveresed viral "

L258 "chromosome"

L323 "motiff found in EBV's Envelope Glyocprotein" (2 typos)

We thank the reviewer for pointing out these typos, which we have corrected.

Response to Reviewers

1. Reviewer #1 (Remarks to the Author): The author's responses and their edits and clarifications added to the manuscript fully address my questions and concerns. (Remarks on code availability): A DOI was generated and version tracking implemented, as requested.

We appreciate the Reviewer's comments that we have addressed their questions and concerns

Reviewer #2 (Remarks to the Author): My concerns have been addressed. I appreciate the authors' attention to the reviewer comments. (Remarks on code availability): I looked briefly at the code, but unfortunately I am not an R user so I can't really comment on the usability and I did not try installing it. I did see some hard-coded file paths, for example https://github.com/melamedlab/MolecularMimicry/blob/main/Analysis/Scripts/human_proteome_distinct_12mer_count.R that would need to be modified by a user to match their own file system.

We appreciate Reviewer 2 identifying the hard-coded file paths and confirm that hard-coded file paths have now been corrected