Trends in Parasitology

CellPress

Opinion

Does peptidome mimicry shape host–parasite coevolution?

Jaroslav Flegr [1,*]

The peptidome mimicry hypothesis (PMH) builds on the principle that vertebrate immunity recognizes peptides absent from the host proteome. It extends this idea to predict host–parasite coevolution outcomes, systematic 'missing peptides', the narrow host specificity of many parasites, and the higher susceptibility of some interspecies hybrids to infection. PMH proposes that long-term coevolution reduces parasite peptide vocabularies and drives convergence toward host repertoires – a pattern that can help to infer a parasite's original host. For example, analyses of SARS-CoV-2 peptide vocabularies have been used to reconstruct the virus's likely host-switching history. Beyond theory, PMH provides an independent and effective way to nominate immunogenic peptide targets for vaccine design, complementary to existing prediction methods.

Highlights

The peptidome mimicry hypothesis (PMH) reframes self–non-self discrimination by focusing on gaps between host and parasite peptidomes, the sets of peptides encoded in their proteins.

PMH explains missing peptides in peptidomes, strict host specificity of parasites, and reduced resistance of hybrids.

Empirical evidence supports parasite peptidome impoverishment, host–parasite convergence, and T cell focus on host-absent peptides.

Peptidome analysis across hosts and parasites reveals patterns of host specificity and host-switching in coronaviruses.

Peptidome-based filtering of host-absent peptides offers a new strategy for vaccine target prediction.

Peptide recognition in immunity and host–parasite coevolution

In the mid-1980s, immunology was transformed by two linked advances: it was shown that (i) **MHC class I molecules** (see Glossary) display endogenous protein fragments on the cell surface, and (ii) **T cells** undergo selection in the thymus [1,2]. Until then, immunology had accumulated a wealth of detailed knowledge, yet a central question remained: how can the immune system recognize antigens it has never encountered? The molecular basis and function of MHC molecules provided the answer (Box 1). These findings showed that self–non-self discrimination rests on detecting **peptides** absent from the host's **proteome**. Such proteins usually originate from parasites. They can also arise from mutations in self-proteins (tumor immunology) or from allogeneic tissues (transplantation immunology).

The coevolutionary struggle between hosts and parasites is widely recognized as a major driving force in evolution, shaping both ecological performance and macroevolutionary trajectories. Hosts succeed by reliably detecting parasites; parasites succeed by evading detection. A key parasite adaptation is to remove peptides absent from the host, thereby making its **peptide vocabulary**, and thus its **peptidome**, less visible to the host's immune system. Parasites may achieve this by deleting non-essential peptides or by replacing parasite-specific peptides with host-present ones. This convergence of parasite and host peptide vocabularies is described by the **peptidome mimicry** hypothesis (PMH). Box 2 contrasts peptidome mimicry with classical **molecular mimicry** in a side-by-side format.

The aim of this article is to introduce PMH and its mechanisms, outline host counterstrategies, review existing evidence, identify gaps where further testing is feasible, and suggest practical applications, particularly for vaccine design. Although the conceptual basis rests on established immunological principles, PMH itself is a recent hypothesis whose empirical testing has only become feasible with the advent of large-scale proteome databases and advanced computational tools.

[1]Laboratory of Evolutionary Biology, Department of Philosophy and History of Sciences, Faculty of Science, Charles University, Viničná 7, 128 00, Prague, Czechia

*Correspondence: flegr@cesnet.cz (J. Flegr).

Box 1. How vertebrate immunity distinguishes self from non-self

In nucleated cells, proteasomes continuously degrade portions of endogenously synthesized proteins. Most fragments are recycled, but some are transported into the endoplasmic reticulum, typically trimmed to 8–10 residues, loaded onto MHC class I molecules, and displayed for T cell inspection. **Antigen-presenting cells (APCs)**, such as macrophages and B cells, process proteins taken up from the extracellular space, including proteins from pathogens or tissue debris, and display their peptides on MHC class II molecules. Importantly, both MHC classes present self- and non-self peptides indiscriminately.

T cells cannot distinguish whether the peptide presented by an MHC molecule is self or foreign. Their receptors for peptide–MHC complexes achieve enormous diversity primarily through somatic recombination of several gene segments, each present in the genome in multiple variants. The specificity of the T cell population toward foreign peptides emerges only during thymic education, when developing T cells encounter MHC molecules loaded with self-peptides. Clones that fail to recognize these complexes die by neglect, whereas those that bind too strongly are eliminated or diverted into regulatory lineages. This dual selection process ensures that only clones with a low-to-intermediate affinity for self peptide-MHC complexes survive, producing a mature T cell repertoire that is tolerant to self yet poised to mount a strong response against novel foreign peptides.

Mature T cells are of two main types, each with distinct roles. Cytotoxic T cells recognize peptides presented by MHC class I molecules and trigger the death of the presenting cell. They are pivotal in combating intracellular parasites. By contrast, helper T cells identify peptides bound to MHC class II molecules and supply the presenting cells with growth factors. Here, the presence of a non-self peptide on a class II molecule signals that the APC has internalized a foreign protein and should be supported to mount a broader immune response. Helper T cells are vital in combating extracellular parasites and refining B cell antibody specificity.

Naturally, even parasites with a vocabulary identical to that of their host may still be detected. Peptides that are rare in host proteins, or expressed at low levels, including in the thymus, may be presented too infrequently to delete the corresponding T cell clones [35,36]. Such clones can persist and later respond when the same peptides are abundant in a parasite. This mechanism may also help to explain why infections sometimes trigger autoimmune diseases such as type 1 diabetes [37] or multiple sclerosis [38].

Box 2. Peptidome mimicry versus molecular mimicry

It is essential to distinguish *peptidome mimicry* from the more familiar concept of *molecular mimicry*. The latter concept, originally formulated by Damian in the 1960s, became widely recognized in the 1980s [39,40]. Molecular mimicry describes how parasites evade immune surveillance by substituting or masking parasite-specific antigens with determinants resembling those of the host. In its classical form, the focus was on shared **epitopes**: if a parasite antigen closely resembles a host antigen, immune recognition of the parasite may be reduced. Later, the concept was broadened to include adverse side effects for the host, such as autoimmunity. Here, cross-reactivity of T cells or antibodies against shared epitopes can lead to immune attacks on host tissues [41], damaging the host and, indirectly, sometimes also the parasite [42].

Peptidome mimicry, by contrast, refers to a different and more fundamental strategy. Rather than *sharing* epitopes, parasites *remove* their unique epitopes – peptides present in parasite proteins but absent from the host proteome. By eliminating such unique peptides from their peptide vocabulary, parasites reduce the repertoire of non-self targets that the host immune system can recognize. The apparent enrichment of host-present peptides in parasite proteins may occur as a side effect, for example when a parasite-specific motif mutates into an already existing host motif. However, this enrichment is not the main driver of the process. The driving force is selection against parasite-unique peptides that label the pathogen as foreign.

This distinction matters because it implies different evolutionary dynamics. Molecular mimicry typically involves selection on a limited number of antigenic determinants and may lead to localized cross-reactivity. Peptidome mimicry, in contrast, operates genome-wide. It can reshape large parts of the parasite proteome and, over long coevolutionary timescales, lead to pronounced convergence of peptide vocabularies between hosts and their parasites.

Peptidome mimicry operates at multiple levels: at the population level, strains lacking strongly immunogenic peptides spread, while within a host, immune pressure selects mutants that lose parasite-unique peptides. In viruses, including human coronaviruses, selection operating both at the population and infrapopulation (within-host) levels can drive a very rapid erosion of host-absent peptides [43].

Taken together, molecular mimicry and peptidome mimicry both describe ways in which parasites reduce their visibility to host immunity. Yet because peptidome mimicry operates at a broader, proteome-wide scale, it warrants recognition as a distinct conceptual framework in host–parasite coevolution.

Glossary

Antigen-presenting cells (APCs): cells that display antigens to activate T cells. Professional APCs express MHC class II and present exogenous antigens.
Basic reproduction number (R_0): the average number of secondary infections caused by one infected individual in a fully susceptible population.
Bridge host: an organism that transfers a pathogen from a reservoir to a target host when direct transmission is inefficient; sometimes misnamed 'intermediate host'.
Clade: a group consisting of a common ancestor and all its descendants.
Epitopes: molecular regions recognized by antibodies or T cell receptors.
Heterosis (hybrid vigor): enhanced performance of first-generation hybrids from genetically distinct parents, often reduced in later generations.
Host-switching event: successful colonization of a new host species by a parasite/pathogen with sustained onward transmission in that host.
Hybridogenesis: recurrent hybrid formation in which offspring arise by hybridization, but only one parental genome enters gametes.
Major histocompatibility complex (MHC) molecules: cell-surface molecules that present peptides to T cells.
Molecular mimicry: structural similarity between parasite and host molecules, aiding immune evasion but sometimes causing autoimmunity.
Peptide: a short chain of amino acids. Longer chains form proteins, which may consist of multiple peptide chains.
Peptide vocabulary: the set of all unique peptides encoded in the proteins of a species, irrespective of how often they occur.
Peptidome: the complete set of peptides encoded in the proteins of a species, including both their presence and their relative frequencies.
Peptidome mimicry: mimicry of the host's peptide repertoire, where parasite-unique peptides are eliminated or replaced with host ones, reducing T cell targets.
Proteome: the entire protein set expressed by a genome, cell, tissue, or organism.
Pseudoreplication: statistical error that occurs when non-independent data are treated as independent; the error is common in comparative studies.

Peptidome simplification in parasites and hosts

Each cell expresses multiple MHC genes, often with numerous variants, giving individuals unique combinations of MHC proteins. This diversity lets individuals recognize the same parasite via different peptides, which hinders parasites from escaping detection via single-peptide mutations. However, despite this, long-term or recurrent interactions between host and parasite species lead to a coevolutionary arms race. During this process, parasites develop specific adaptations, often including the modification of identifiable peptides, to counter host defenses. One strategy is to simplify their peptide vocabulary (e.g., collapse diverse motifs into fewer variants) so that functions previously carried by several peptides are maintained by a smaller set. This reduces repertoire size relative to free-living species, but at a cost: protein efficiency and stability may decline, limiting the extent of reduction.

Even free-living hosts with vertebrate-type immunity are selected to trim their peptide vocabulary, though less strongly than parasites. Failure to recognize parasite reduces host fitness, whereas failure of evasion is often fatal for the parasite (the 'Life–Dinner Principle'). Fewer peptides mean fewer T cells deleted during thymic selection, preserving a broader repertoire and enhancing parasite recognition. Therefore, on theoretical grounds, vertebrates that rely on peptide-based recognition are expected to have smaller peptide vocabularies than invertebrates, fungi, or plants. However, because vertebrates constitute a single monophyletic **clade** ($n = 1$ for comparative purposes), this prediction is difficult to test rigorously and, to our knowledge, has not been directly evaluated.

In contrast, the peptidome mimicry hypothesis predicts that interspecies hybrids and individuals from distant intraspecific crosses likely have larger peptide vocabularies due to the combination of parental repertoires. This may lead to more extensive deletion of T cells during thymic selection and thus reduced immune resistance. However, hybrid susceptibility is not a universal pattern; outcomes vary depending on genetic background, parasite and host species, and environmental context. A recent comprehensive review [3] found that, in mammals, increased infection susceptibility was reported in 57% of studies, whereas in fish, hybrids most often showed increased resistance (40% of studies). The PMH offers a plausible contributing mechanism in cases where hybrid susceptibility is observed, especially in later-generation hybrids after initial **heterosis** wanes. Naturally, additional factors, such as misregulation or gene incompatibilities, may also play a role.

Since the 1990s, it has been known that the peptide vocabularies of studied species lack a significant number of peptides that should theoretically be present under simple combinatorial expectations [4,5]. For instance, one can infer expected frequencies of longer peptides from shorter ones – pentapeptides from tripeptides and tetrapeptides, or hexapeptides from tetra- and pentapeptides. However, observed frequencies often diverge from these expectations; some peptides, especially pentapeptides expected to be common, are entirely absent [6]. This discrepancy reflects more than physicochemical constraints: some peptides are missing broadly across taxa, suggesting structural limitations, but many others are absent only in certain species or clades, consistent with lineage-specific selection pressures [4,6]. Selection by parasites via the direct elimination of parasite-unique peptides during the host–parasite arms race is a plausible contributor, though alternative explanations remain possible.

In sum, although the diversity of MHC molecules complicates parasite evasion, long-term host–parasite interactions still drive an evolutionary arms race. A recurring outcome is peptide vocabulary simplification, benefiting parasites by reducing detectability and hosts by preserving a broader T cell repertoire. The next section turns to an even more refined strategy: peptidome mimicry, in which parasites not only reduce peptide diversity but also replace their peptides with host-present ones.

RaTG13: a coronavirus related to SARS-CoV-2, sequenced in the same laboratory; it differs mainly by lacking the furin cleavage site.
Reinforcement: strengthening of reproductive barriers between species by natural selection.
SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2): a novel coronavirus causing COVID-19. An enveloped, positive-sense single-stranded RNA virus (~30 kb) in the family Coronaviridae, genus *Betacoronavirus* (subgenus *Sarbecovirus*).
Speciation: an evolutionary process by which new species arise.
Synanthropic species: species living near and adapting to human environments, for example, the house mouse.
T cells: lymphocytes maturing in the thymus; cytotoxic T cells kill infected cells, helper T cells coordinate responses.
Zoonoses: diseases caused by pathogens transmitted from animals to humans.

Beyond reduction: peptidome mimicry in parasite evasion

In addition to peptide vocabulary reduction, parasites that specialize on a narrow host range may adopt a strategy of peptidome mimicry [7]. To avoid detection, they accumulate mutations that replace their unique peptides with host-present peptides. Like vocabulary reduction, this adaptation carries costs, since mutations that enhance mimicry can impair protein efficiency or stability, forcing a balance between immune evasion and functionality [8]. Current data suggest that peptidome mimicry is most evident in viruses, particularly those causing chronic or latent infections. A 2024 survey of 134 human-infecting viruses [9] found widespread presence of short peptides identical or nearly identical to host peptides across diverse human viruses. Large DNA viruses such as members of the Herpesviridae (e.g., human herpesvirus 1, Epstein–Barr virus) and members of the Poxviridae (e.g., variola virus, molluscum contagiosum virus) showed the strongest enrichment for host-present peptides. This contrasts with acute 'hit-and-run' viruses [10], exemplified by influenza and norovirus, which have broad host ranges and rely on transmission speed rather than long-term immune evasion [9].

Constraints, host specificity, and zoonotic transfers

Parasites face limitations in adapting their vocabularies to multiple phylogenetically distant hosts, because proteins must retain enough peptides to remain functional. This constraint helps to explain host specificity, where many parasites infect a limited host range, sometimes just one species [11,12]. For example, *Plasmodium falciparum* infects only humans, and papillomaviruses are typically highly species-specific. Parasites with low host specificity, that cannot achieve close convergence with any single host's peptide vocabulary, may instead use a 'hit-and-run' strategy [13]. These parasites multiply quickly and produce infectious stages, allowing transmission before immune clearance.

Receptor-binding change is typically a necessary step for host switching [14,15] but is rarely sufficient. Most spillovers fail because the **basic reproduction number** (R_0) in the new host remains below 1. Repeated spillover events, including occasional reversals to the original host, can enable gradual adaptation. Progressive peptidome convergence with the new host can contribute to raising R above the epidemic threshold and facilitate the transition from occasional spillover to sustained, species-specific parasitism.

Phylogenetically related host species share similar peptide vocabularies, allowing parasites to thrive in both their primary and closely related hosts. This is why primates, our evolutionary relatives, can be especially risky sources of **zoonoses** [16]. Nevertheless, most human infections originate from domesticated animals [17]. The number of transferred parasite species correlates with domestication duration [18], suggesting that long timescales and repeated attempts are typically required for a parasite to cross into unrelated hosts. After zoonotic incidents, antibodies against the parasite were often detected in samples from both humans and domestic animals taken long before the outbreak [19,20]. This implies that many failed transmission attempts preceded successful spillover.

Unrelated species that share habitats, and occasionally exchange parasites, may, in theory, converge in their peptide vocabularies. This alignment could arise if host species progressively eliminate peptides that, while non-essential to them, are shared with local parasites, thereby increasing parasite visibility to immunity. This trend should be most evident in domesticated and **synanthropic** species, whose vocabularies are predicted to most closely resemble those of humans, and in ecological communities with shared environments and overlapping parasitic pressures. Testing these predictions is feasible with existing proteomic databases, but such analyses have yet to be systematically performed.

The number of peptides shared between species reflects phylogenetic relatedness, whereas shared losses of peptides may indicate both relatedness and similar parasite-mediated selection. If mutations were purely random, phylogenetic trees derived from shared [21] and absent [22] peptides should broadly coincide with morphology-based trees. Therefore, deviations between these two types of phylogenies may signal peptidome mimicry and provide an analytical route to detect its influence using existing genomic data.

At the population level, host–parasite coevolution is shaped by selection favoring rare variants. These allele-level dynamics likely also influence the proteome-wide convergence predicted by the PMH, potentially accelerating the coevolution of host and parasite peptide vocabularies.

Testing the peptidome mimicry hypothesis: case studies

Parasites have reduced peptide vocabularies

The PMH not only clarifies already known phenomena but also offers specific, testable predictions (Table 1, Key table). As a relatively new hypothesis introduced in 2017, PMH has only recently begun to be empirically tested. The following case studies illustrate how modern sequencing data and bioinformatic approaches have opened new avenues for evaluating its predictions. An analysis of 38 parasitic and 33 free-living organisms showed that parasites have significantly fewer pentapeptides. Notably, this reduction was not attributable to proteome size or gene number, as models including these variables still identified parasitism as the dominant factor

Key table

Table 1. Tested and untested predictions of the peptidome mimicry hypothesis[a]

Facts known before the peptidome mimicry hypothesis was formulated
Missing peptides in vocabularies, varying across different clades ☻ [4–6]
Host specificity of most parasites, often very strict ☻ [11,12]
Lower resistance to parasites in some interspecies hybrids, especially in vertebrates ☻ [32–34]
Predictions tested after the peptidome mimicry hypothesis was formulated
Impoverished peptide vocabularies in parasites of vertebrates ☻ [23]
Similarity in pentapeptide vocabularies between a parasite and its main hosts ☻ [22]
Similarity in hexapeptide vocabularies between a recently captured parasite and its presumed bridge host ☻ [22]
T cell targeting of parasite peptides absent from the host vocabulary ☻ [31]
Predictions of the peptidome mimicry hypothesis yet to be tested
Greater host specificity in parasites of vertebrates compared to other parasites ☺
Little or no peptide impoverishment in parasites of non-vertebrates (plants, fungi, invertebrates) ☺
Richer pentapeptide vocabularies in non-parasitic non-vertebrates compared to vertebrates ☺
Richer pentapeptide vocabularies in interspecies hybrids
Richer pentapeptide vocabulary in individuals from distant intraspecific crosses
Lower resistance of individuals from distant intraspecific crosses
Similarity in peptide vocabulary between humans and domesticated and synanthropic species ☺
Richer peptide vocabulary in solitary species (with fewer parasites) than in gregarious or social species ☺
Phylogenies from missing peptides show weaker signal than those from shared peptides ☺

[a]Symbol ☻ denotes a prediction already tested and supported while ☺ denotes a prediction not yet tested, but for which all necessary data are already available in public repositories.

[23]. Although MHC molecules typically bind longer peptides, only the central residues contact the T cell receptor, with flanking residues embedded in the MHC groove [24]. To avoid **pseudoreplications**, five clades were analyzed separately; in each, parasites had more limited pentapeptide vocabularies, with the probability of this pattern arising by chance = 0.031 (Fisher's exact test). The results also supported the prediction that this reduction in peptide vocabulary size is absent in parasites that target invertebrates. However, only one such parasite, *Perkinsus* (an oyster parasite), was included in the study. Notably, the pentapeptide deficit co-occurred with a mild increase in hexapeptides, which Zemková *et al*. interpreted as a compensatory adjustment to preserve protein function under a restricted pentapeptide repertoire.

Coronaviruses as a case study of host specificity and switching

A 2022 study examined **SARS-CoV-2** to assess parasite–host vocabulary similarities [22]. SARS-CoV-2 is thought to be a recombinant of a bat-adapted virus, with another virus contributing the human-adapted Spike protein [25–27]. Because of this dual origin, it has been recommended to analyze the Spike region separately from the rest of the genome [28]. Analysis of 11 human and bat coronaviruses against 38 phylogenetically diverse mammalian species showed that the Spike's pentapeptide vocabulary was most similar to that of humans. By contrast, the remainder of the proteome most closely resembled horseshoe bat – see Figure 1. This unique pattern, also seen in the presumed SARS-CoV-2 ancestor **RaTG13**, suggests recombination between bat- and human-adapted viruses and illustrates how vocabulary analysis can reveal original host specificity of a virus.

The same study extended the analysis to hexapeptides. Because each pentapeptide can be embedded in up to 40 distinct hexapeptides (20 for each possible flanking residue), the loss of a single pentapeptide causes the disappearance of many related hexapeptides. As a result, hexapeptide matches decay faster with sequence change and can be more informative about recent host switches, such as pinpointing likely **bridge hosts**, whereas pentapeptide patterns are more stable and tend to reflect longer-term adaptation. Consistent with this, the pentapeptide data pointed to deep ancestry in bats, whereas the hexapeptide profiles of non-Spike proteins in both SARS-CoV-2 and RaTG13 were closest to tree shrews. By contrast, the Spike protein hexapeptide profile diverged: in SARS-CoV-2 it was most similar to rats, whereas in RaTG13 it was most similar to mice. All three species co-occur in China in both natural and human-associated settings [29]. The coronavirus case study indicates that peptidome profiles can trace both deep host associations and recent **host-switching events**. Although further confirmation is needed, they exemplify how peptidome analysis can generate testable hypotheses about host-switching dynamics and provide a framework for targeted empirical studies.

Implications for vaccine design

PMH also has implications for vaccine development. A study of epitope mapping in 99 post-COVID patients [30] showed that conventional MHC–peptide binding predictions yield thousands of candidates. Only a small fraction proves immunogenic (14.3% for helper, 8.1% for cytotoxic T cells). In contrast, a 2023 study [31] identified 983 host-absent pentapeptides in SARS-CoV-2, 30% of which overlapped with empirically verified epitopes. This filter reduces the candidate pool from thousands to fewer than a thousand while retaining ~30% of confirmed T cell targets. Therefore, the peptide-vocabulary-based method represents a more focused and efficient prediction strategy than conventional pipelines. Importantly, it complements rather than replaces existing epitope prediction tools that rely on MHC-binding affinity. When the peptide-vocabulary method is used as an initial filter to exclude host-present peptides and is followed by binding predictions, the resulting pipeline yields a smaller, higher-value set of vaccine targets and thus improves vaccine design efficiency.

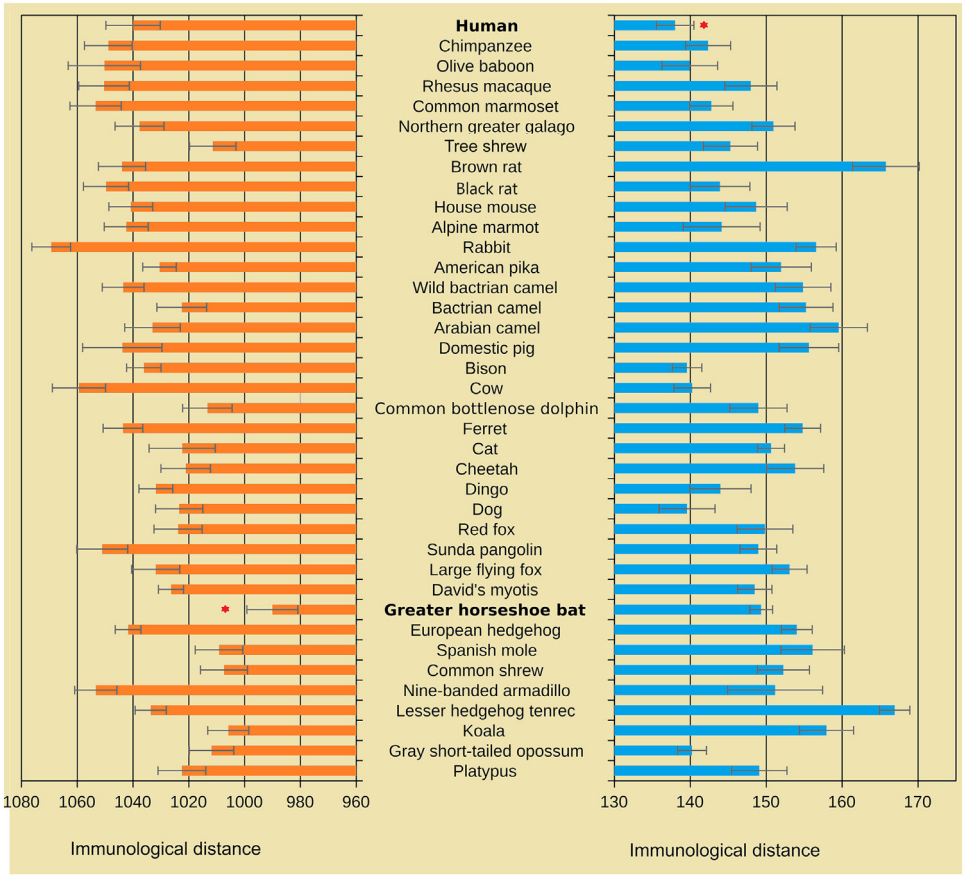

Figure 1. Immunological T-distance and host-specific adaptation of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Bars show the immunological T-distance, defined as the number of SARS-CoV-2 pentapeptides absent from the host proteome (standardized by host proteome size); error bars denote 95% confidence intervals. Lower T-distance indicates fewer virus-specific pentapeptides in that host and thus higher inferred mimicry (reduced immune visibility). Results are shown for two viral peptide sets: non-Spike proteins (left) and Spike (S) protein (right). *The pattern suggests that Spike is most adapted to humans, whereas non-Spike proteins are most adapted to horseshoe bats. Figure adapted from [22].

Outstanding questions

How widespread is peptidome mimicry across non-viral parasitic taxa? Does peptidome mimicry differ between parasites causing acute versus chronic infections?

To what degree can parasites reduce their peptide vocabularies without compromising essential protein function?

Are peptide impoverishment patterns found in parasites of vertebrates with peptide-based self–non-self recognition, but absent in those of plants, fungi, and invertebrates?

Do proteomic data confirm predicted similarities between humans and domesticated or synanthropic species, and what does this mean for zoonotic risk?

How does hybridization, both inter- and intraspecific, alter peptide vocabularies and influence T cell repertoire pruning and pathogen resistance?

Do genetic distances based on shared peptide presence reflect phylogenetic signal shaped by drift, whereas those based on shared peptide absences reflect parasite-driven convergence?

What role does peptidome mimicry play in reinforcement during speciation, and could it explain patterns of reduced hybrid fitness?

Can integrating PMH-based filters with MHC-binding predictions significantly improve epitope selection for vaccines?

Concluding remarks and future perspectives

The PMH is a novel conceptual framework that has begun to receive empirical attention with the advent of whole-proteome sequencing. Preliminary findings suggest that it may help to explain certain patterns, such as species-specific gaps in peptide vocabularies, the narrow host specificity of many parasites, and the increased susceptibility of many interspecies vertebrate hybrids to parasite infections. Some predictions derived from the hypothesis, such as reduced peptide vocabularies in parasites of vertebrates, similarities between parasite and host peptide vocabularies, and the enrichment of host-absent peptides among T cell targets, have shown empirical support, though more extensive and rigorous testing is needed.

Other predictions remain largely unexplored and invite further investigation (see Outstanding questions). The lower diversity of vertebrate peptide vocabularies compared to those of metazoans lacking peptide-based recognition of non-self is also largely unexplored. Further research is needed on the peptide vocabularies of distant intraspecific hybrids, including the effects on T

cell repertoire and parasite resistance. The potential impact of **hybridogenesis** on peptide vocabulary expansion (and the corresponding narrowing of the T cell repertoire) deserves attention in vertebrate taxa where it occurs, such as certain amphibians and fishes. Variation in peptide vocabularies driven by parasitic pressures may contribute to reduced hybrid resistance and therefore deserves closer examination, given its potential role in **reinforcement** during **speciation**.

Parasites employ a wide range of immune evasion strategies, from antigenic variation to the secretion of immunosuppressive molecules. Peptidome mimicry, if confirmed as a general phenomenon, would represent a complementary process operating at a more fundamental proteomic level. Rather than replacing other mechanisms, it may operate alongside them in long-term host–parasite coevolution, gradually reshaping peptide vocabularies through both simplification and convergence. Highlighting this possibility draws attention to an underexplored dimension of evolutionary conflict that deserves closer empirical scrutiny.

Acknowledgments

The author was supported by the Czech Science Foundation (grant no. 22-20785S).

Declaration of interests

The author has no conflict of interests.

References

1. Pishesha, N. *et al.* (2022) A guide to antigen processing and presentation. *Nat. Rev. Immunol.* 22, 751–764
2. Ashby, K.M. and Hogquist, K.A. (2024) A guide to thymic selection of T cells. *Nat. Rev. Immunol.* 24, 103–117
3. Theodosopoulos, A.N. *et al.* (2019) Parasites and host species barriers in animal hybrid zones. *Trends Ecol. Evol.* 34, 19–30
4. Doi, H. *et al.* (1995) Diverse incidences of individual oligopeptides (dipeptidic to hexapeptidic) in proteins of human, bakers' yeast, and *Escherichia coli* origin registered in the Swiss-Prot data base. *Proc. Natl. Acad. Sci. U. S. A.* 92, 2879–2883
5. Strelets, V.B. *et al.* (1994) Analysis of peptides from known proteins: clusterization in sequence space. *J. Mol. Evol.* 39, 625–630
6. Tuller, T. *et al.* (2007) Forbidden penta-peptides. *Protein Sci.* 16, 2251–2259
7. Flegr, J. (2016) Watch Out for Toxo!: The Secret Guide to Practical Science. In *Faculty of Science* (1st edn), Charles University Press
8. Hurford, A. and Day, T. (2013) Immune evasion and the evolution of molecular mimicry in parasites. *Evolution* 67, 2889–2904
9. Maguire, C. *et al.* (2024) Molecular mimicry as a mechanism of viral immune evasion and autoimmunity. *Nat. Commun.* 15, 9403
10. Ewald, P.W. (1994) *Evolution of Infectious Disease*, Oxford University Press
11. Poulin, R. and Keeney, D.B. (2008) Host specificity under molecular and experimental scrutiny. *Trends Parasitol.* 24, 24–28
12. Esser, H.J. *et al.* (2016) Host specificity in a diverse Neotropical tick community: an assessment using quantitative network analysis and host phylogeny. *Parasit. Vectors* 9, 372
13. Flegr, J. (2006) Evolutionary parasitology: the development of invasion, evasion, and survival mechanisms. In *Food Consumption and Disease Risk: Consumer–Pathogen Interactions* (Potter, M., ed.), pp. 251–270, Woodhead Publishing Limited
14. Galaway, F. *et al.* (2019) Resurrection of the ancestral RH5 invasion ligand provides a molecular explanation for the origin of *P. falciparum* malaria in humans. *PLoS Biol.* 17, e3000490
15. Wang, Q. *et al.* (2023) Key mutations on spike protein altering ACE2 receptor utilization and potentially expanding host range of emerging SARS-CoV-2 variants. *J. Med. Virol.* 95, e28116
16. Jiang, X. *et al.* (2023) A review on zoonotic pathogens associated with non-human primates: understanding the potential threats to humans. *Microorganisms* 11, 246
17. Weiss, R.A. and Sankaran, N. (2022) Emergence of epidemic diseases: zoonoses and other origins. *Fac. Rev.* 11, 2
18. Morand, S. *et al.* (2014) Domesticated animals and human infectious diseases of zoonotic origins: Domestication time matters. *Infect. Genet. Evol.* 24, 76–81
19. Zheng, B.J. *et al.* (2004) SARS-related virus predating SARS outbreak, Hong Kong. *Emerg. Infect. Dis.* 10, 176–178
20. Müller, M.A. *et al.* (2015) Presence of Middle East respiratory syndrome coronavirus antibodies in Saudi Arabia: a nationwide, cross-sectional, serological study. *Lancet Infect. Dis.* 15, 559–564
21. Kanduc, D. and Shoenfeld, Y. (2020) Molecular mimicry between SARS-CoV-2 spike glycoprotein and mammalian proteomes: implications for the vaccine. *Immunol. Res.* 68, 310–313
22. Flegr, J. *et al.* (2022) Thus spoke peptides: SARS-CoV-2 spike gene evolved in humans and then shortly in rats while the rest of its genome in horseshoe bats and then in treeshrews. *Commun. Integr. Biol.* 15, 96–104
23. Zemkova, M. *et al.* (2017) Parasitism as the main factor shaping peptide vocabularies in current organisms. *Parasitology* 144, 1–9
24. Vyas, J.M. *et al.* (2008) The known unknowns of antigen processing and presentation. *Nat. Rev. Immunol.* 8, 607–618
25. Zhu, Z. *et al.* (2020) Genomic recombination events may reveal the evolution of coronavirus and the origin of SARS-CoV-2. *Sci. Rep.* 10, 21617
26. Makarenkov, V. *et al.* (2021) Horizontal gene transfer and recombination analysis of SARS-CoV-2 genes helps discover its close relatives and shed light on its origin. *Bmc Ecol. Evol.* 21, 5
27. Li, X. *et al.* (2020) Emergence of SARS-CoV-2 through recombination and strong purifying selection. *Sci. Adv.* 6, eabb9153
28. Lytras, S. *et al.* (2022) Exploring the natural origins of SARS-CoV-2 in the light of recombination. *Genome Biol. Evol.* 14, evac018
29. Sirotkin, K. and Sirotkin, D. (2020) Might SARS-CoV-2 have arisen via serial passage through an animal host or cell culture? A potential explanation for much of the novel coronavirus' distinctive genome. *Bioessays* 42, 2000091
30. Tarke, A. *et al.* (2021) Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases. *Cell Rep Med* 2, 100204
31. Flegr, J. *et al.* (2023) Exposing and exploiting host–parasite arms race clues in SARS-CoV-2: a principally new method for improved T cell immunogenicity prediction. *Biol. Methods Protoc.* 8, bpad011
32. Fritz, R.S. *et al.* (1999) Resistance of hybrid plants and animals to herbivores, pathogens, and parasites. *Annu. Rev. Ecol. Syst.* 30, 565–591

33. Moulia, C. (1999) Parasitism of plant and animal hybrids: are facts and fates the same? *Ecology* 80, 392–406
34. Karvonen, A. and Seehausen, O. (2012) The role of parasitism in adaptive radiations – when might parasites promote and when might they constrain ecological speciation? *Int. J. Ecol.* 2012, 280169
35. Anderson, M.S. *et al.* (2002) Projection of an immunological self shadow within the thymus by the aire protein. *Science* 298, 1395–1401
36. Klein, L. *et al.* (2009) Antigen presentation in the thymus for positive selection and central tolerance induction. *Nat. Rev. Immunol.* 9, 833–844
37. Houeiss, P. *et al.* (2022) Environmental triggering of Type 1 diabetes autoimmunity. *Front. Endocrinol. (Lausanne)* 13, 933965
38. Lanz, T.V. *et al.* (2022) Clonally expanded B cells in multiple sclerosis bind EBV EBNA1 and GlialCAM. *Nature* 603, 321–327
39. Damian, R.T. (1964) Molecular mimicry: antigen sharing by parasite and host and its consequences. *Am. Nat.* 98, 129–149
40. Oldstone, M.B. (2014) Molecular mimicry: its evolution from concept to mechanism as a cause of autoimmune diseases. *Monoclon. Antibodies Immunodiagn. Immunother.* 33, 158–165
41. Cusick, M.F. *et al.* (2012) Molecular mimicry as a mechanism of autoimmune disease. *Clin. Rev. Allergy Immunol.* 42, 102–111
42. Gouttefangeas, C. *et al.* (2023) The good and the bad of T cell cross-reactivity: challenges and opportunities for novel therapeutics in autoimmunity and cancer. *Front. Immunol.* 14, 1212546
43. Venkatakrishnan, A.J. *et al.* (2020) Benchmarking evolutionary tinkering underlying human-viral molecular mimicry shows multiple host pulmonary-arterial peptides mimicked by SARS-CoV-2. *Cell Death Discov.* 6, 96