RESEARCH PREPRINT

A structure-guided hypothesis for sequential S8/PtpA cleavage of the Ara h 2 DPYSPS loop

Outcome-blind computational prioritization and a blinded validation protocol

Luqman Mahmood

Sfiniti AI | Independent research program

2026-07-18

Computational hypothesis and executable protocol | No wet-lab result

Abstract

Background. A secondary analysis of the IMPACT peanut oral immunotherapy trial associated failure to achieve remission with increased copy number of the gut microbial ptpA gene and showed that participant fecal communities could degrade Ara h 2. PtpA is an N-terminal Xaa-Xaa-Pro tripeptidyl peptidase, whereas the immunologically important DPYSPS motifs of Ara h 2 are internal. The initiating event that could make those motifs available to PtpA is therefore unresolved.

Methods. We performed an outcome-blind, receipt-tracked analysis of public shotgun metagenomes from BioProject PRJNA1261967, reviewed same-assembly S8/S53 and S9B/PtpA-like protein candidates from four organisms implicated by the parent study, corrected Ara h 2 sequence-to-structure mappings for PDB entries 3OB4, 8G4P, and 9RTG, and derived cleavage products from the Q6PSU2-1 sequence. We then designed a blinded, staged N-terminomics and targeted peptide assay with predeclared controls and stop rules.

Results. We propose that an S8/S53-family endoprotease could create a QDP-starting neo-N terminus at either S68|Q69 or S75|Q76. A subsequent PtpA-like reaction would remove the N-terminal QDP, producing Y72 or D79, respectively. The minimum sequential signature is linked precursor loss, downstream neo-N-terminus gain, and free-QDP gain. Y72 alone is insufficient because direct cleavage at P71|Y72 can produce the same terminus. Four same-assembly candidate pairs are retained as coequal first-screen candidates. The executable design contains 243 synthetic-peptide samples in Stage A and 150 Ara h 2 samples in Stage B across six 96-well plates; a 40-row Stage C remains disabled until the same pair passes both earlier stages on the same branch. No biochemical assay has been executed.

Conclusions. The work supplies a narrow, falsifiable molecular mechanism and a separation-of-duties validation packet. It does not establish enzyme activity, reduced allergenicity, or therapeutic benefit. Independent proteomics and enzymology laboratories are invited to test and report positive or negative results.

Plain-language summary

Previous research found that some gut microbiomes contain more copies of a gene called ptpA and also break down more Ara h 2, an important peanut allergen. The known PtpA enzyme can remove groups of three amino acids only from the exposed start of a protein or peptide. The relevant proline-rich segments of Ara h 2 are inside the protein, so PtpA cannot logically begin the process by itself.

We identified a testable two-enzyme possibility. A first, subtilisin-like enzyme may cut immediately before a QDP sequence. A PtpA-like enzyme could then remove that QDP. This predicts exact molecular

products that a mass spectrometry laboratory can look for. The prediction may be wrong; the protocol is intentionally designed to expose that quickly and report the negative result.

1. Introduction

Ara h 2 is a major peanut allergen and a protease-resistant 2S albumin. Its large loop contains repeated DPYSPS motifs that contribute to IgE recognition, with motif copy number differing between isoforms [1-5]. The loop can remain immunologically important even when the protein core is perturbed, and Ara h 2 digestion can leave stable, immunoreactive products [3,5].

Ozcam and colleagues analyzed the randomized IMPACT peanut oral immunotherapy trial and reported that non-remission was associated with an increased abundance of ptpA, encoding an Xaa-Xaa-Pro tripeptidyl peptidase [6]. Participant-derived fecal communities degraded Ara h 2 in vitro, with greater loss in cultures from children who did not achieve remission. The authors proposed that microbial degradation might reduce the antigen exposure needed to maintain remission. The study established an association and a community-level degradation phenotype; it did not identify the molecular cleavage sequence connecting PtpA to Ara h 2.

The biochemical constraint is important. The characterized PtpA reference from *Porphyromonas gingivalis* hydrolyzes Xaa-Xaa-Pro|Yaa, releasing the N-terminal tripeptide when Yaa is not proline [7]. An internal DPYSPS repeat is not itself an N terminus. A distinct endoprotease or prior processing event must first expose a compatible N-terminal tripeptide.

We therefore asked a bounded question: do the implicated microbial genomes encode plausible same-assembly endoprotease and tripeptidyl-peptidase candidates, and does the Ara h 2 sequence contain exposed bonds whose cleavage would produce an N-terminal PtpA substrate and unambiguous downstream products? The answer produced a concrete two-stage hypothesis. This paper reports that hypothesis, its computational provenance, and a blinded protocol for determining whether it is false.

2. Evidence classes and outcome firewall

Every statement in this work is assigned to one of four evidence classes:

1. **Observed public record:** sequence, structure, annotation, or published experimental result.
2. **Artifact-reviewed computation:** deterministic mapping, sequence-derived product, or measured runtime result with a retained receipt.
3. **Experimental hypothesis:** a predicted enzyme activity or molecular sequence requiring wet-lab testing.

4. **Excluded claim:** therapeutic, clinical, causal, or safety conclusions that the present evidence cannot support.

Metagenomic processing was outcome-blind. Run workers did not read clinical outcome metadata. Candidate abundance was not used to reorder the four enzyme pairs or the blinded plate plan. The 37-run computational cohort was incomplete at the release snapshot, so this paper reports no new outcome-association statistic.

3. Methods

3.1 Public sequence and metagenomic sources

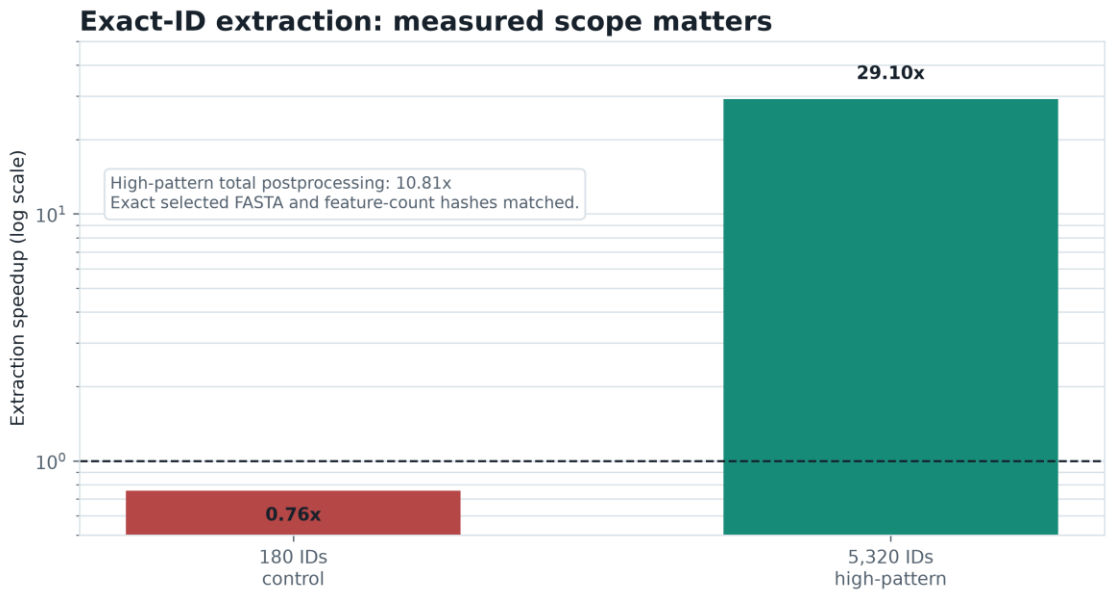
Public IMPACT shotgun metagenomes are available under NCBI BioProject PRJNA1261967 and SRA study SRP584678 [6]. Protein candidates were resolved against NCBI protein and assembly records. The PtpA substrate-class reference was UniProt Q7MUW6 and its underlying biochemical report [7].

Four assemblies were considered because the parent study attributed much of the `ptpA` signal to *Bacteroides uniformis*, *Phocaeicola dorei*, *Bacteroides caccae*, and *Bacteroides xylanisolvens* [6]. Within each assembly, one primary S8/S53-family candidate and one S9B/PtpA-like candidate were retained. Generic S9 annotation was not treated as proof of PtpA substrate specificity.

3.2 Exact metagenomic counting

The computational campaign used staged acquisition, quality control, host removal, target prefiltering, protein alignment, exact conditional counting, and receipt-bound fallback. Selected candidate reads were extracted by exact sequence identifier. DIAMOND supplied sensitive protein alignment [8]. Optional bounded proposal and payload-reduction steps ran in shadow; exact alignment owned every reported count and automatically retained ownership when the shadow result differed.

For the high-pattern SRR33533016 row, exact-identifier extraction reduced the critical extraction stage from 9,153.68 s to 314.60 s (29.10-fold) and total postprocessing from 9,456.98 s to 875.19 s (10.81-fold), with matching selected FASTA and feature-count hashes. A 180-identifier control row was slower (0.756-fold), so the optimization is explicitly limited to high-pattern extraction. Runtime measurements are engineering evidence, not evidence for the biological hypothesis.



Measured computational scaling for exact sequence-ID extraction. The low-pattern negative control is retained and the speed claim is limited to the high-pattern row.

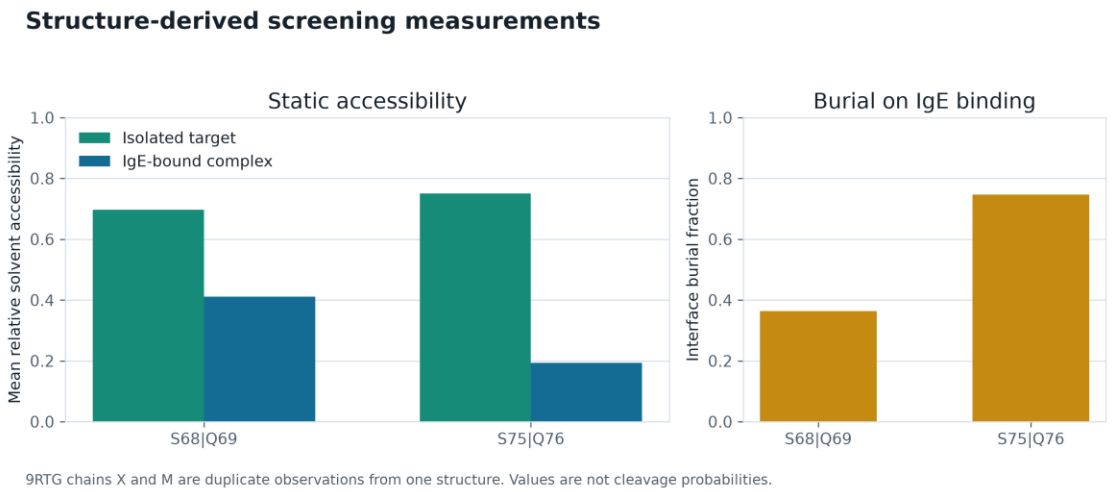
3.3 Structure correction and mapping

Ara h 2 candidates were mapped to:

- 3OB4, an X-ray MBP-Ara h 2 fusion [2];
- 8G4P, Ara h 2 chain E bound by antibodies 13T1 and 13T5 [4];
- 9RTG, two copies of the DPYSPS-rich loop bound by the D08 IgE Fab [5].

Mappings required sequence identity at the modeled positions. Missing coordinates were represented as missing, never as zero accessibility. The two Ara h 2 chains in 9RTG were treated as duplicate observations from one structure, not independent structures. Solvent-accessible surface area was computed in isolated-target and complex contexts. The resulting scores are screening-order heuristics, not probabilities of cleavage.

For Q6PSU2, the mean isolated and complex relative solvent accessibility values were 0.697 and 0.412 at S68|Q69, and 0.751 and 0.193 at S75|Q76. Interface-burial fractions were 0.364 and 0.747, respectively. These values support physical availability in the isolated structural model but do not model dynamics, assay pH, enzyme docking, or catalysis.



Static structure-derived screening measurements for the two Q6PSU2 branches. Chains X and M in 9RTG are duplicate copies in one complex and are not independent structures.

3.4 Derivation of the sequential signatures

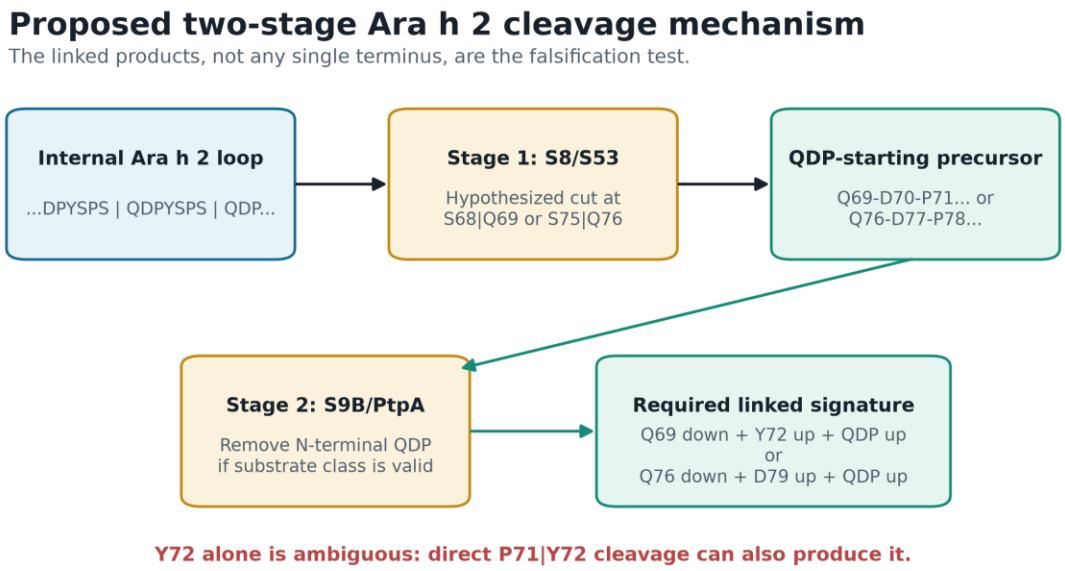
The Q6PSU2-1 sequence around the two proposed branches is:

```
... DPYSPS | QDPYSPS | QDPDRR ...
             ^ Q69       ^ Q76
```

Cleavage at `S68|Q69` creates a product beginning `Q69-D70-P71`. Cleavage at `S75|Q76` creates a product beginning `Q76-D77-P78`. Both are therefore `QDP`-starting substrates compatible with the substrate class reported for PtpA [7]. Removal of the N-terminal tripeptide predicts:

```
S68|Q69 initiation -> Q69 precursor -> release QDP -> Y72 product
S75|Q76 initiation -> Q76 precursor -> release QDP -> D79 product
```

Direct S8 cleavage at `P71|Y72` can also generate `Y72`. Consequently, `Y72` alone is not accepted as evidence of sequential action. The required signature is the linked change in the branch-specific precursor, branch-specific downstream terminus, and free `QDP`.



The proposed two-stage mechanism and its required linked molecular signatures. The diagram states a hypothesis, not an observed reaction.

N-terminal glutamine may cyclize to pyroglutamate, removing the free N-terminal amine and confounding PtpA action or detection [9,10]. The protocol therefore requires measurement of both Q-starting and pyroglutamate-starting species, controlled delay and buffer conditions, and a pyroglutamate analog.

3.5 Candidate enzyme pairs

Four pairs are admitted as coequal first-screen candidates:

Pair	Assembly and organism	Primary S8/S53	S9B/PtpA-like
1	GCA_000154205.1, *B. uniformis* ATCC 8492	EDO53097.1	EDO52549.1
2	GCA_000156075.1, *P. dorei* DSM 17855	EEB27374.1	EEB25557.1
3	GCA_000169015.1, *B. caccae* ATCC 43185	EDM19660.1	EDM21505.1
4	GCA_000210075.1, *B. xylanisolvens* XB1A	CBK68024.1	CBK67447.1

CBK67447.1 has the most explicit PtpA-like annotation, but annotation certainty does not establish activity on QDP, expression in the trial samples, or sequential activity with its paired S8 candidate. The pairs are not ranked biologically before assay.

3.6 Blinded validation design

The detailed protocol and machine-readable manifests are supplied with this paper. It separates laboratory operations from blinded analysis.

Stage A: S9/PtpA-like qualification

Stage A contains 243 synthetic-peptide reactions. It tests candidate S9/PtpA-like enzymes on QDP-starting substrates, position-3 Pro-to-Ala controls, pyroglutamate controls, inactive enzymes, no-enzyme controls, and a validated PtpA reference. A candidate passes a branch only if at least two of three valid replicates span both experimental blocks, product exceeds the laboratory-declared LLOQ, signal is at least 3.0-fold over inactive and no-enzyme controls, the matched Pro-to-Ala signal is no more than 0.20 of the QDP substrate, and replicate CV is no more than 0.30.

Stage B: S8 initiation

Stage B contains 150 reactions using native non-reduced recombinant Ara h 2 as the primary substrate and reduced/alkylated Ara h 2 as a mechanistic control. It tests Q69 and Q76 independently. Passing requires at least two of three valid replicates across both blocks, a neo-N-terminus signal at least 3.0-fold over controls, reproducible loss of intact Ara h 2, complete reporting of new N termini, and replicate CV no more than 0.30.

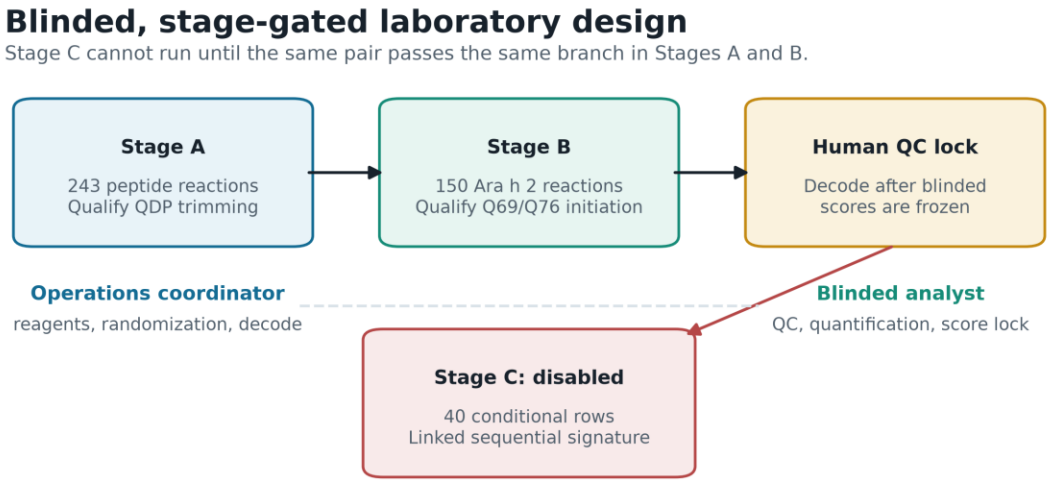
Stage C: conditional sequential test

Forty Stage C template rows are disabled. A branch may be unlocked only after human review confirms that the same pair passed Stage A and Stage B on that same branch. The analyst must lock quality-control decisions while blinded before the decode map is opened. Sequential attribution requires:

- loss of Q69 with gain of Y72 and free QDP; or
- loss of Q76 with gain of D79 and free QDP;

with every linked change exceeding matched vehicle, inactive-S9, and S9-only controls.

N-terminomics methods such as TAILS can identify protease-generated neo-N termini [11], while targeted LC-MS/MS with stable-isotope standards can quantify the specified small and branch-specific peptides. The executing laboratory must choose and predeclare its validated analytical implementation.



Stage-gated assay and separation of laboratory operations from blinded analysis.

4. Results

4.1 The public evidence leaves an initiating-cleavage gap

The parent study establishes a relationship among ptpA, failure of oral-immunotherapy remission, and community-level Ara h 2 degradation [6]. PtpA substrate chemistry requires an exposed N terminus [7]. The repeated Ara h 2 DPYSPS motifs are internal [1-5]. A prior initiating event is thus required for the proposed direct action of PtpA on this region.

4.2 Two Ara h 2 bonds generate the same diagnostic PtpA substrate

Sequence derivation identified two post-motif bonds that each generate QDP at the new N terminus: S68|Q69 and S75|Q76. Structural mapping shows both bonds in the isolated 9RTG target model. The second bond is more buried in the IgE-bound complex, emphasizing that antibody-bound accessibility cannot be equated with unbound enzymatic accessibility.

4.3 The two branches yield linked, falsifiable products

The first branch predicts Q69 loss, Y72 gain, and free-QDP gain. The second predicts Q76 loss, D79 gain, and free-QDP gain. The design also reports every other new terminus, preventing a target whitelist from hiding nonspecific proteolysis or an alternative mechanism.

4.4 Same-assembly candidates are available for empirical screening

Each of the four implicated organism assemblies contains a primary S8/S53-family candidate and a full-length S9B/PtpA-like candidate. Homology and annotation establish screening eligibility only. They do not establish co-expression, localization, activation, substrate specificity, or physical interaction.

4.5 A complete blinded assay packet is ready but unexecuted

Stages A and B contain 393 enabled samples across six plates. Randomized sample identifiers, an analyst-blinded manifest, a scoring template, an operations manifest, reagent specifications, configuration worksheets, and a decode map are supplied. Stage C is disabled by default. No laboratory result is reported in this paper.

5. Discussion

5.1 Contribution

The contribution is not the general observation that microbes can degrade Ara h 2, that subtilisin-like proteases can reduce peanut allergens, or that `ptpA` is associated with oral-immunotherapy outcome. Those points have published or patented precedent [6,12]. The contribution is the narrower mechanistic bridge:

> an S8/S53-family internal cut could expose a `QDP` N terminus in Ara h 2, allowing a PtpA-like enzyme to remove `QDP`, with linked products that make the sequence distinguishable from nonspecific degradation.

A bounded search conducted on 18 July 2026 did not identify a public experimental test of this exact two-stage mechanism, these two site branches, or the linked `Q69/Y72/QDP` and `Q76/D79/QDP` signatures. This is not a legal freedom-to-operate opinion or an exhaustive literature claim.

5.2 What a positive result would mean

A positive Stage A result would show only that a candidate behaves as a QDP-tripeptidase under the declared assay. A positive Stage B result would show only that a candidate S8 produces one of the proposed Ara h 2 termini. A positive Stage C result would support the linked sequential mechanism in vitro. None of these findings alone would establish that the pathway occurs in a human gut, explains oral-immunotherapy failure, or changes clinical risk.

The biological direction may also depend on context. Ex vivo processing that destroys IgE-binding structure could potentially reduce immediate allergenicity, while accelerated degradation during oral immunotherapy could reduce antigen exposure needed for durable remission [6]. Separate blinded IgE-binding and functional assays would be required after cleavage is proven.

5.3 What would falsify or downgrade the hypothesis

The hypothesis is rejected or narrowed if:

- no candidate S9B/PtpA-like enzyme removes QDP from the synthetic substrates;
- Q-starting products cyclize before the second reaction under relevant conditions;
- no candidate S8 creates Q69 or Q76 on folded Ara h 2;
- sequential treatment does not produce linked precursor/product/QDP changes;
- observed cleavage is broad and nonspecific;
- candidate genes are absent, unexpressed, or inactive in the relevant communities;
- cleavage products retain or increase IgE binding or effector-cell activity.

Negative results should be deposited with the same sample identifiers and protocol deviations so that the field does not repeatedly test an unsupported mechanism.

5.4 Limitations

This work is computational and artifact-reviewed, not independently replicated. Static solvent accessibility does not represent protein dynamics, solution conditions, glycation, proline hydroxylation, folding heterogeneity, or enzyme docking. 9RTG contains an IgE-bound loop and duplicate Ara h 2 copies in one complex. 8G4P does not resolve the DPYSPS-rich loop. 3OB4 is an MBP fusion and resolves only part of the repeat region.

Candidate functions are inferred from annotation and homology. Generic S9 members may not exhibit PtpA activity. Enzyme constructs, signal peptides, propeptide processing, activation, pH optima, and concentrations remain to be configured by a qualified laboratory. Metagenomic DNA assignments do not measure expression or activity. The outcome-blind cohort was not complete at the release snapshot, and no new clinical association is reported.

6. Independent validation request

We invite laboratories with N-terminomics, targeted proteomics, protease biochemistry, and allergen-handling capability to:

1. quote and configure Stage A and Stage B;
2. preserve operations/analyst separation;
3. execute both positive and negative controls;
4. return raw instrument files, acquisition methods, quality-control decisions, and hashes;
5. publish positive, negative, or ambiguous results.

Contact may be initiated through the associated Zenodo record. The NIH-funded program led by Mustafa Ozcam, which is independently investigating microbial peanut-protein metabolism and PtpA, is a particularly relevant potential recipient [13].

Data and materials availability

The Zenodo record contains the manuscript, figures, public evidence tables, analyst-blinded files, laboratory-operations files, checksums, and validation script. The original shotgun metagenomes remain at NCBI under PRJNA1261967; PDB and UniProt records are referenced by accession rather than redistributed. No participant-level clinical outcomes are included.

Ethics statement

This report performs secondary computational analysis of public, de-identified sequence data and public protein structures. It includes no new human enrollment, intervention, specimen collection, or wet-lab experiment. The original IMPACT study and its secondary analysis report their own institutional approvals and consent procedures [6]. Any laboratory receiving the protocol is responsible for its institutional biosafety, allergen, recombinant-protein, and analytical approvals.

AI assistance disclosure

OpenAI Codex was used under direct human supervision to inspect artifacts, cross-check claim boundaries, organize the protocol, assist literature retrieval, and draft and validate release materials. The named author selected the hypothesis, reviewed the evidence, directed the analysis, and accepts responsibility for the manuscript. AI assistance is not listed as authorship.

Competing interests

The author develops MFabric, a private computational campaign and execution system used in the analysis. The runtime implementation is not required to execute the laboratory protocol. No enzyme candidate reported here has been commercially validated.

Funding

No external funding was received for this computational hypothesis paper. Cloud-compute expenses were paid privately.

References

1. Chatel JM, Bernard H, Orson FM. Isolation and characterization of two complete Ara h 2 isoforms cDNA. *Int Arch Allergy Immunol.* 2003;131:14-18. doi:10.1159/000070429.
2. Mueller GA, et al. Ara h 2: crystal structure and IgE binding distinguish two subpopulations of peanut allergic patients by epitope diversity. *Allergy.* 2011;66:878-885. doi:10.1111/j.1398-9995.2010.02532.x.
3. Apostolovic D, et al. Conformational stability of digestion-resistant peptides of peanut conglutins reveals the molecular basis of their allergenicity. *Sci Rep.* 2016;6:29249. doi:10.1038/srep29249.
4. Min J, et al. Design of an Ara h 2 hypoallergen from conformational epitopes. *Clin Exp Allergy.* 2024;54:46-55. doi:10.1111/cea.14433.
5. Parkkinen T, et al. Structural basis for trivalent cross-linking of a patient-derived IgE antibody by the major peanut allergen Ara h 2.0201. *Allergy.* 2026;81:2036-2046. doi:10.1111/all.70258.
6. Ozcam M, et al. Gut microbial bile and amino acid metabolism associate with peanut oral immunotherapy failure. *Nat Commun.* 2025;16:6330. doi:10.1038/s41467-025-61161-x.
7. Banbula A, et al. Prolyl tripeptidyl peptidase from *Porphyromonas gingivalis*. *J Biol Chem.* 1999;274:9246-9252. doi:10.1074/jbc.274.14.9246.
8. Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods.* 2021;18:366-368. doi:10.1038/s41592-021-01101-x.
9. Seifert F, et al. Phosphate ions and glutaminyl cyclases catalyze the cyclization of glutaminyl residues by facilitating synchronized proton transfers. *Bioorg Chem.* 2015;60:98-101. doi:10.1016/j.bioorg.2015.04.005.
10. Bochtler M, et al. The Bacteroidetes Q-rule: pyroglutamate in signal peptidase I substrates. *Front Microbiol.* 2018;9:230. doi:10.3389/fmicb.2018.00230.
11. Kleifeld O, et al. Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat Protoc.* 2011;6:1578-1611. doi:10.1038/nprot.2011.382.
12. North Carolina State University. Enzymatic treatment of peanuts. US20180014562A1. Published 18 January 2018.
13. Ozcam M. Gut microbial protein metabolism in food allergy and oral immunotherapy efficacy. NIH RePORTER project 1DP2AI192747-01, 2026-2031.