Allergenicity Evidence in U.S. GRAS Notices for Novel Proteins: A Document Audit

Bryan Quoc Le, Ph.D., Mendocino Food Consulting · Nov 3, 2026 · @Bryan Quoc Le

Abstract

Background. FDA's public response letters almost never show how a novel food protein was assessed for allergenicity: none of the 48 notices for which only the letter is available names an allergen database or a screening threshold.

Methods. This audit examined the 138 GRAS notices for protein, precision-fermentation protein and biomass ingredients in FDA's GRAS Notice Inventory. It coded the 83 for which an FDA letter, a notifier dossier, or both could be obtained (29 with both, 48 letter only, 6 pending with dossier only).

Each notice was coded on 13 fields covering protein identity, sequence screening, the database and thresholds used, host-cell proteins, digestion and heat testing, serum IgE or clinical testing, cross-reactivity, labeling, the framework cited, and what FDA's letter says. Coding used two independent AI passes, a blind third coder to resolve disagreements, a fourth coder for the one remaining split, and a brief review by the author.

Results.

- Among the 35 notices with a dossier, the mean transparency score was 2.5 of 5, and 11 (31%) scored zero.
- The score depended on the ingredient. Precision-fermentation proteins averaged 4.0 and proteins identical to a listed allergen 3.4, while conventional plant and animal protein isolates and concentrates averaged 0.5, with 9 of 11 scoring zero.
- Where a sequence screen was reported, it almost always used AllergenOnline (18 of 19) and the 35% identity over 80 amino acids criterion from Codex (2003). Only 6 of 17 added the 8-amino-acid exact match, and no notice built its assessment on any framework newer than Codex (2003) and FAO/WHO (2001).
- One notice of 83 reported any human immunological testing.

Conclusions. Because FDA's letter is the only public document for most notices, the public record does not let a reader judge how allergenicity was assessed. Two modest changes would close most of that gap: a standard allergenicity summary in every notice, and a sentence in FDA's letter naming the methods it reviewed.

1. Introduction

The allergenicity of a novel protein is usually argued, not tested, and the strength of the argument depends on what the notifier chose to report. Proteins made by precision fermentation, proteins isolated from new plant sources, and whole microbial or fungal biomass now make up a growing share of U.S. GRAS notices. Each one raises the same question: could this protein, or something carried along with it, sensitize consumers or cross-react with an existing allergy?

The accepted approach is a weight-of-evidence assessment set out by FAO/WHO (2001) and the Codex Alimentarius Commission (2003). It combines the source of the protein, sequence comparison against known allergens (an identity of more than 35% over a window of 80 amino acids, with a short exact-match search as an option), resistance to pepsin digestion, and, where the source is allergenic, serum IgE testing. Under the GRAS rule (21 CFR 170, subpart E), a notifier must provide a narrative that supports its conclusion of safety, but the rule does not prescribe how allergenicity is to be assessed or reported.

FDA's response letter is the public face of each notice. For most notices filed before about 2024, it is the only document a reader can easily obtain, because the full notices for older filings are archived rather than posted. This audit asked a simple question: from the documents the public can actually read, how transparent and how complete are the allergenicity assessments behind GRAS notices for novel proteins?

2. Data and methods

Scope. Notices were drawn from FDA's GRAS Notice Inventory as exported on November 2, 2026, using the classification from the author's earlier GRAS backlog study. Three classes were included: precision-fermentation proteins, other proteins (isolates, concentrates and hydrolysates) and biomass or whole-cell ingredients, for 138 notices. Enzymes (177 notices) were left for a later phase.

Documents. For each notice, the FDA response letter and any notifier dossier linked from FDA's inventory page were collected. Letters hosted on fda.gov and dossiers hosted on fda.gov or regulations.gov were saved; letters held only in FDA's web archive (mostly notices closed before 2017) could not be retrieved. This left 83 notices: 29 with both a letter and a dossier, 48 with a letter only, and 6 pending notices with a dossier only. Text was extracted from each PDF with page numbers retained. For long dossiers, the coder received the first pages of each part plus every page containing an allergenicity term.

Codebook. Each notice was coded on 13 fields: protein identity, protein source, sequence screen, database named, thresholds stated, host-cell or co-isolated proteins, digestibility, heat or processing stability, serum IgE or clinical testing, cross-reactivity with the production organism, allergen labeling, framework cited, and the content of FDA's letter. Every code other than an absent code carries a verbatim quote and page. The codebook was revised once after the first merge (version 1.1): a protein counts as identical to a known allergen only if it is itself a listed allergen; prevalence literature alone does not count as IgE or clinical evidence; any reported screen or test is coded, including for whole biomass; and any reported in vitro digestibility assay, including nutritional protein-digestibility kits, and any thermal characterization such as differential scanning calorimetry, count as tests.

Coding and review. Two AI coders worked independently from the same documents. Where they disagreed on a key field, a third coder that did not see their answers coded the notice, and the majority answer was kept. One field on one notice remained split three ways and was settled by a fourth coder, who also re-read one dossier that an earlier coder had only skimmed. The codes then received a brief review by the author.

Transparency score. Each notice received one point for each of five items a reader needs to judge a sequence-based assessment: a described screen, a named database, stated thresholds, quantified host-cell or co-isolated proteins, and a digestion test. Scores run from 0 to 5. Letter-only and dossier-based notices are reported separately, because a letter summarizes what the notifier filed rather than reproducing it.

3. The public letter record

None of the 48 letter-only notices names an allergen database or a threshold, and none describes how a sequence screen was run. Seven of the 48 letters say that a bioinformatic comparison was done, in phrases such as "sequence-alignment approaches" or "no significant similarity to known allergens", without naming a method. Three report a digestion test and six give an amount for co-isolated or host proteins. The mean transparency score for letter-only notices was 0.2 of 5, and 39 of 48 (81%) scored zero.

This is not because FDA's letters ignore allergenicity. Of the 77 letters obtained, 43 (56%) contain two or more sentences on it, 13 (17%) contain one, and 21 (27%) say nothing. The difference lies in the outcome of the notice:

FDA outcome	Letters	Two or more sentences	One sentence	Silent
No questions	58	43 (74%)	10 (17%)	5 (9%)
Ceased at the notifier's request	19	0	3 (16%)	16 (84%)

The substance of these passages is the limitation. Most consist of FDA restating the notifier's conclusion that the ingredient is unlikely to be allergenic, together with the standard paragraph reminding the notifier of its labeling obligations under FALCPA and the FASTER Act. FDA's letters rarely say what evidence the conclusion rested on. The share of letters with a substantive passage rose from 46% for notices closed through 2020 to 61 to 62% from 2021 onward, but the content of those passages did not change in kind.

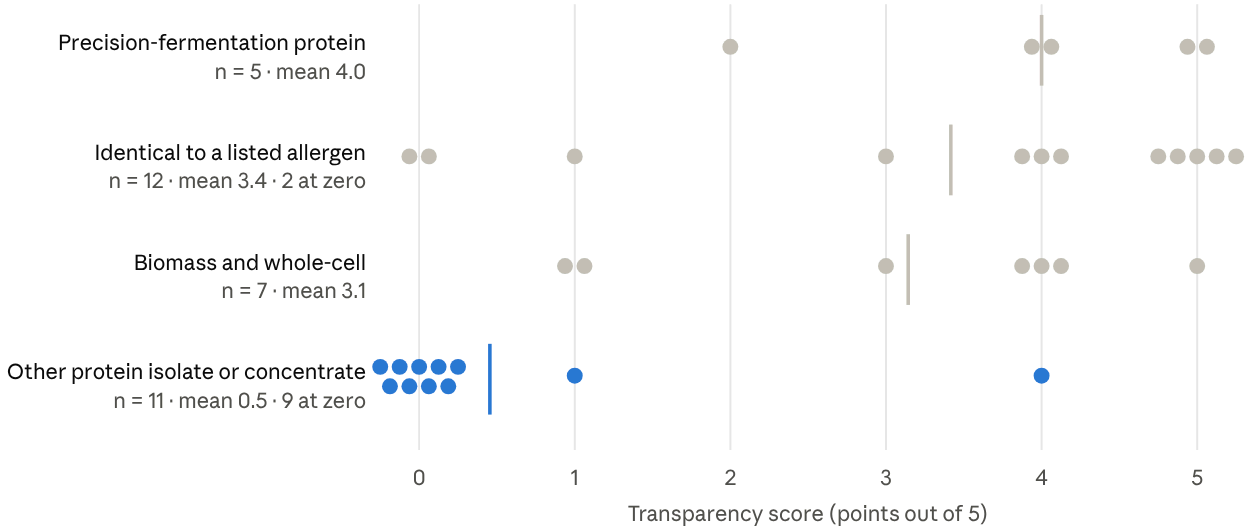
For the 19 notices that the notifier withdrew, the letter usually records only the withdrawal. Two of them list allergenicity among FDA's open questions: one cites "inadequate bioinformatics and in silico approaches" (GRN 1268), another the lack of discussion of allergens in the production medium (GRN 1061). These are the only places in the public letter record where FDA names a weakness in an allergenicity assessment.

4. What the dossiers show

Where a dossier is available, the ingredient type predicts how much of the allergenicity assessment a reader can see. Across the 35 notices with a dossier, 19 (54%) described a sequence screen, 19 named a database, 17 (49%) stated thresholds, 16 (46%) quantified host-cell or co-isolated proteins and 17 (49%) reported a digestion test. The mean score was 2.5; 11 notices (31%) scored zero and 8 (23%) scored five.

Conventional protein isolates scored lowest on allergenicity transparency

Transparency score (0 to 5) for each of the 35 notices with a dossier. One dot per notice; tick = group mean.



Source: author's coding of 35 GRAS notices with a dossier (29 closed, 6 pending), FDA GRAS Notice Inventory, November 2026

Precision-fermentation proteins and proteins identical to a listed allergen were documented most fully. Their notifiers screened the production organism's residual proteins, quantified them, and usually tested digestion. Biomass notices filed since 2024 increasingly ran genome-wide screens of the organism's predicted proteins. Conventional protein isolates and concentrates (rice, hemp, fava, almond, goat whey, eggshell membrane and similar) rarely did any of this: their safety arguments rested on the history of eating the source food. That argument is reasonable for a familiar food, but the dossiers seldom said so explicitly, and they did not address whether concentrating, hydrolyzing or fermenting the protein changes its allergen content.

5. Methods, testing and labeling

The sequence screens that were reported used one database and one rule almost without exception.

Item	Notices	Share
Database named	19	of 83
AllergenOnline (FARRP, University of Nebraska)	18	of 19 named

Item	Notices	Share
Database version or date given	12	of 19 named
Thresholds stated	17	of 83
35% identity over 80 amino acids	16	of 17 stated
8-amino-acid exact match added	6	of 17 stated

The one notice that used a different database, COMPARE, was the Lemna leaf protein (GRN 1256). Seven of the 19 notices that named a database gave no version or search date, which matters because AllergenOnline is updated each year and a screen cannot be repeated without knowing which release was searched. Where a framework was cited, it was Codex (2003) in 13 notices, FAO/WHO (2001) in 8 and EFSA guidance in 2. No notice built its assessment on a framework newer than those, even though EFSA's 2022 scientific opinion on allergenicity assessment questioned whether the 35%/80-amino-acid rule is fit for purpose; one dossier (GRN 1249) cited that opinion, but only in passing, and still applied the 35% rule.

Testing beyond the screen was uncommon. Twenty notices (24%) reported a digestion test, counting nutritional protein-digestibility kits as well as pepsin assays. Five (6%) reported thermal characterization of the protein. Only one notice of 83, the heat-killed Xanthobacter biomass (GRN 1351), reported any human immunological testing, a basophil activation test in 50 volunteers. Sixteen notices assessed cross-reactivity with the production organism, mostly by screening its residual proteins.

Allergen labeling was not addressed in 46 of 83 notices (55%). Of the 21 proteins identical to a listed allergen, 15 (71%) committed to a major-allergen declaration such as "milk" or "egg". Ten notices stated that no allergen label was needed, usually because the source is not one of the nine major food allergens.

6. Discussion and recommendations

The gap this audit found is mostly a reporting gap, not evidence that unsafe proteins are reaching the market. For precision-fermentation proteins, the assessments that can be read are generally thorough. The difficulty is that a reader usually cannot read them: for 48 of 83 notices the only public document is a letter that gives FDA's conclusion and none of the method behind it. Even where dossiers exist, a third score zero, and these are concentrated among conventional protein isolates that rely on the source food's history of use.

Three changes would make the record far easier to evaluate, at little cost to notifiers or FDA:

1. **A standard allergenicity summary in every notice.** One page stating the protein's identity and source, the database and release searched, the thresholds applied, the host-cell protein content and how it was screened, and any digestion, heat or IgE testing. Most precision-fermentation dossiers already contain this information; it is simply scattered across sections and appendices.
2. **One sentence in FDA's letter naming the methods reviewed.** For example: "The notifier compared the amino acid sequence against AllergenOnline version 22 using a threshold of 35% identity over 80 amino acids and an 8-amino-acid exact match." This would let the letter, which is the document most people read, carry the method as well as the conclusion.
3. **An explicit argument for conventional isolates.** Where a notice relies on history of use, it should say so and address whether concentration, hydrolysis or fermentation could raise allergen levels or create new epitopes relative to the source food.

For notifiers, the practical lesson is that the elements reviewers look for are well known and inexpensive to report. The Allergenicity Evidence Checklist that accompanies this report lists them and shows how a draft notice compares with the 35 dossiers coded here.

7. Limitations

- **Coverage.** The 55 notices whose letters are held only in FDA's web archive, nearly all closed before 2017, are not included, so the findings describe the last decade of filings rather than the whole program.
- **What was read.** Long dossiers were read as excerpts: the opening pages of each part plus every page containing an allergenicity term. About 6% of dossier pages were scanned images with no extractable text, concentrated in four notices (GRN 1219, 1248, 1249 and 1268). Information on unread pages may have been missed, which would understate those notices' scores.
- **Coding.** Codes were assigned by AI coders working independently, with disagreements resolved by majority and a brief review by the author. No separate human accuracy sample was drawn, and no accuracy figure is claimed.
- **The score.** The transparency score counts what is reported, not whether a method is adequate. A notice can score five with a weak screen, or zero for a protein that is plainly safe; the score is a measure of what a reader can check.

- **Small groups.** The ingredient-type comparisons rest on 5 to 12 dossiers per group and should be read as patterns, not estimates.

8. Conclusions

For most GRAS notices on novel proteins, the public cannot see how allergenicity was assessed. FDA's letters state a conclusion but not the method, and none of the 48 letter-only notices names a database or a threshold. Where dossiers are available, precision-fermentation proteins are documented well and conventional protein isolates poorly, and nearly every screen rests on one database and a 2003-era rule with little testing beyond it. A one-page allergenicity summary in each notice, and one sentence of method in each FDA letter, would make the record checkable without adding new testing requirements.

Data availability

Notice-level codes and transparency scores for all 83 notices are published with the Allergenicity Evidence Checklist at mendocinofoodconsulting.com. The codebook, extraction scripts and coding files are available from the author on request.

References

1. Codex Alimentarius Commission (2003). Guideline for the Conduct of Food Safety Assessment of Foods Derived from Recombinant-DNA Plants (CAC/GL 45-2003), Annex 1: Assessment of Possible Allergenicity.
2. FAO/WHO (2001). Evaluation of Allergenicity of Genetically Modified Foods. Report of a Joint FAO/WHO Expert Consultation, Rome, 22 to 25 January 2001.
3. EFSA Panel on Genetically Modified Organisms (2022). Scientific opinion on development needs for the allergenicity and protein safety assessment of food and feed products derived from biotechnology. EFSA Journal 20(1), 7044.
4. Goodman, R. E., et al. (2016). AllergenOnline: A peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. Molecular Nutrition & Food Research 60(5), 1183 to 1198.
5. van Ree, R., et al. (2021). The COMPARE database: A public resource for allergen identification, adapted for continuous improvement. Frontiers in Allergy 2, 700533.
6. U.S. Food and Drug Administration (2016). Substances Generally Recognized as Safe; Final Rule. 81 Fed. Reg. 54960 (21 CFR 170, subpart E).

7. Food Allergen Labeling and Consumer Protection Act of 2004 (Public Law 108-282); Food Allergy Safety, Treatment, Education, and Research (FASTER) Act of 2021 (Public Law 117-11).
8. U.S. Food and Drug Administration. GRAS Notice Inventory. https://www.fda.gov/food/generally-recognized-safe-gras/gras-notice-inventory (exported November 2, 2026).