



Guest Editorials

Evaluating proteins for potential allergenicity using bioinformatic approaches



Ever since the introduction of a tougher tomato¹ and, more recently, the discovery of a “DNA nanosurgery” system based on CRISPR-CAS,² we are inexorably headed to a new world where genetically modified (GM) animals will consume GM plants, both of which will be labeled GM foods. These in turn will be (futuristically) consumed by GM humans. It is natural for allergists to wonder whether this could increase the incidence of allergy as we strive to *primum non nocere* and apply bioinformatics approaches to assess this at the bedside.³ Greenhawt et al⁴ address this question rigorously in this issue of the *Annals*.

As outlined by Greenhawt et al, broadly speaking, there are 2 hypothetical mechanisms by which proteins introduced by a new gene could be allergenic. The foreign proteins could be allergenic or the introduced genes could cause the host organism to express sufficiently high levels of its own allergenic proteins. The following guidelines⁵ are used to assess the potential allergenicity of a GM food:

1. Is the source organism of the gene, taken as a whole, known to be a common allergenic food?
2. Is the introduced protein similar to known allergens?
3. Is the newly expressed protein abundant in the food?
4. Is the expressed protein highly resistant to digestion?

It is a truism that the allergenic nature of an offending food is attributable to a very small number of macromolecules, usually proteins, out of tens of thousands. This is the underlying rationale for the production of allergen-free versions of foods such as peanut.⁶ Therefore, if a gene has desirable properties for transfer, then its origin from an allergenic plant or animal should not prohibit its candidacy, especially when recombinant technology is used.

How similar should a protein be to a known allergen to be considered potentially allergenic? As our food supply grows increasingly diverse, allergists might have to make a judgment on a case-by-case basis using bioinformatics principles and approaches.

To enable the allergist to address this key question, a key concept to consider is the nuanced notion of similarity between a protein of interest and a database of known allergenic proteins. Ideally, one would like to compare the detailed 3-dimensional structure. Because this is not known for most proteins, similarity is often assessed by comparing the corresponding amino acid sequences as a proxy for the putative structures. There are 3 general ways of scoring similarity: (1) the length of a contiguous run of

exact matches, (2) percentage identity, and (3) statistical significance. To illustrate methods 1 and 2, consider the toy example of matching the peptide sequence STEAL (Ser-Thr-Glu-Ala-Leu) to the peptides TEAK and STALL. STEAL and TEAK have the sequence TEA in common, corresponding to a contiguous match length score of 3. In contrast, STEAL and STALL have a match length score of 2 (either ST or AL can be considered to be shared, but not both). Therefore, based on match length as a criterion, TEAK is more similar to STEAL than STALL is.

For percentage identity, STEAL and TEAK have a score of 60% (3 of 5 positions in the alignment). However, STEAL and STALL have a sequence identity score of 67% (4 of 6 positions in alignment). Therefore, based on percentage identity, STALL appears to be a better match for STEAL than TEAK. Sequence analysis algorithms such as FASTA⁷ and BLAST⁸ are commonly used to find the alignment between a pair of sequences that maximizes the similarity score.

The third method, the statistical significance of the score, is the probability that a given score could be due to coincidence. For example, using FASTA to search an allergen database⁹ for matches to the vegetative insecticide protein from *Bacillus thuringiensis* (a GM addition to corn) retrieves a match to a dust mite allergen with an *E* value of 0.42 for the reported score. The *E* value is the expected number of coincidental matches in a given database search (*E* values <1 are approximately equivalent to *P* values and can be interpreted as such). The high *P* value confirms the fact that vegetative insecticide protein has been a safe addition to crops.

In the context of allergy, it is important to note that short stretches of exact matches are more relevant for recognition by T-cell receptors, whereas larger noncontiguous stretches are epitopes more likely to interact with immunoglobulins G and E. Also, tools such as FASTA and BLAST can result in false negatives because they are heuristic searches unlike slower, but more thorough, approaches such as the algorithm by Smith and Waterman.¹⁰

Returning to the determination of the required threshold for allergenicity, the Codex Alimentarius Commission⁵ recommends a threshold of greater than 35% identity within a stretch of 80 amino acids. Although seeking an exact contiguous match of amino acids is not recommended at this time,⁹ this could change in the future as studies of antigen-receptor specificity result in databases of specific epitopes.

Guided by these criteria, Greenhawt et al conclude there is no evidence in the published literature that GM foods have contributed to an increase in the incidence of allergic reactions. From a wider perspective, it is worthwhile to consider the proportion of

Disclosures: Authors have nothing to disclose.

<http://dx.doi.org/10.1016/j.anai.2017.06.020>

1081-1206/© 2017 American College of Allergy, Asthma & Immunology. Published by Elsevier Inc. All rights reserved.

allergic diseases potentially attributable to GM foods when regulated by rational guidelines. The 21st century will likely be remembered as the period when humans mastered the manipulation of the genome to control the variety of extant biology. Urbanization makes us dwell in the paradox of alternating between biologically sterile homes and an atmosphere of tire particles and exhaust fumes. With increasing globalization, it is common to spend the formative years of our immune system in an environment very different from the rest of our life. In a few decades, we are likely to live in a world where the transition between natural and synthetic biology is seamless. Against that backdrop, the effect of GM foods on allergic disease is just one of the plethora of influences on the immune system to be considered.

Deendayal Dinakarpanid, MD, PhD*[†]

Chitra Dinakar, MD[†]

*University of Missouri

Kansas City, Missouri

[†]Stanford University

Palo Alto, California

cdinakar@stanford.edu

References

- [1] Bruening G, Lyons J. The case of the FLAVR SAVR tomato. *Calif Agric (Berkeley)*. 2000;54:6–7.
- [2] Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*. 2012;337:816–821.
- [3] Dinakarpanid D, Lee Y, Dinakar C. Applications of medical informatics in allergy/immunology. *Ann Allergy Asthma Immunol*. 2007;99:2–9. quiz 9–12, 41.
- [4] Dunn SE, Vicini JL, Glenn KC, Fleischer DM, Greenhawt MJ. The allergenicity of genetically modified foods from genetically engineered crops: a narrative and systematic review. *Ann Allergy Asthma Immunol*. 2017;119:214–222.
- [5] World Health Organization; Food and Agriculture Organization of the United Nations. CODEX Alimentarius. *Foods Derived From Modern Biotechnology*. Rome, Italy: Food and Agriculture Organization of the United Nations; 2009. <http://www.fao.org/docrep/011/a1554e/a1554e00.htm>.
- [6] Li H, Yu J, Ahmedna M, et al. Reduction of major peanut allergens Ara h 1 and Ara h 2, in roasted peanuts by ultrasound assisted enzymatic treatment. *Food Chem*. 2013;141:762–768.
- [7] Pearson WR. Finding protein and nucleotide similarities with FASTA. *Curr Protoc Bioinformatics*. 2016;53:3.9.1–3.9.25.
- [8] Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–3402.
- [9] Goodman RE, Ebisawa M, Ferreira F, et al. AllergenOnline: a peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. *Mol Nutr Food Res*. 2016;60:1183–1198.
- [10] Smith TF, Waterman MS. Identification of common molecular subsequences. *J Mol Biol*. 1981;147:195–197.

Genetically modified products and food allergy Beliefs, evidence, and opportunity



The effect of genetically modified (GM) products on human health is a timely and important topic. This month's CME Review does an excellent job providing background on this technology, outlining potential health concerns as they relate to food allergy, and summarizing evidence on this topic.¹ The article addresses 2 issues that practicing physicians face. First is the increasing disconnect between scientific evidence and popularized fixed health care beliefs and concerns. The second is the many questions we face for which evidence is lacking. Recognizing both these aspects of care is important to provide proactive, engaged, and collaborative care to our patients.

The concern about increased or novel allergenicity related to GM products dates back more than 20 years and was prominently documented in a *New England Journal of Medicine* study that found that the introduction of Brazil nut protein into soybeans resulted in specific IgE binding of that protein in vitro and positive skin test results in sensitive individuals.² In addition, in the early 2000s, cases of anaphylaxis were reported after suspected exposure to Starlink corn.³ Starlink corn was not developed for human consumption but found its way into the human food supply nonetheless. These reports led to a Centers for Disease Control and Prevention (CDC) study that relieved this concern through in vitro testing and oral food challenges.

The disconnect between public discourse and scientific evidence is apparent in the concern over the safety of GM products. Despite the CDC study and reassurance from the scientific community, media reports still presented the findings with skepticism, with one report stating, "If Starlink didn't make them sick, what did?"⁴ For allergists, this question seems rhetorical. In daily clinical practice, allergists routinely encounter patients with allergic symptom for which allergens are not identified. For many patients, particularly those with chronic idiopathic urticaria, not identifying

an external trigger for their symptoms is a difficult concept to internalize.

Effectively communicating this concept can be difficult, and patients sometimes are unreceptive to the practitioners' estimation of their situation. Fixed health care beliefs often present challenges in delivering optimal care. Perhaps the most notable example is the ongoing vaccine controversy. Despite strong evidence supporting the safety of vaccines, many parents continue to believe that vaccines present intolerable risks for their children. In the case of GM products, public concern over the safety of these products, even if not routed in evidence, can be difficult to overcome. At the bedside, patient concerns about GM products driving allergic symptoms may need to be addressed with the understanding that physician and patient belief in the evidence may not be aligned.

Another challenge that clinicians and patients often face is the lack of evidence. The authors' findings point out that the question of whether GM products will lead to increased allergy development has not been fully evaluated. Patients regularly ask their physicians questions for which evidence-based answers are unavailable. In allergy/immunology, our practice parameters contain a perhaps surprising number of recommendations based on low-level evidence.⁵ When faced with evidentiary uncertainty, clinicians are well served to apply shared decision-making techniques to ensure our patients are given the opportunity to accommodate ambiguity into their choices.⁶

These discussions with our patients can serve as fertile ground for practitioners to connect with the communities we serve and work toward a shared and integrated understanding of health. An example of a program that attends to patient and population health with the community is Rooted Community Health, a program of Vanderbilt University's Center for Biomedical Ethics and Society. Focusing on engaging the community's concerns about food supply, agricultural production, and distribution methods as they intersect with health care, Rooted