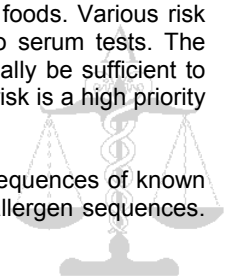
A Silvanovich et al. Toxicological Science (2006) Vol.90(1) p 252 – 258
The Value of Short Amino Acid Sequence Matches for Prediction of Protein Allergenicity

Inaccuracies in desktop risk assessment of Genetic Modifications could lead to excessive protection.

Genetic modification could result in the expression of novel allergens in traditional foods. Various risk assessment methodologies have been developed ranging from desktop studies to serum tests. The remotest chance that a new GM variety could contain a known allergen would usually be sufficient to deter commercialisation if there was any risk of human exposure. Detection of such risk is a high priority for GM researchers and civil society organisations.



One method of risk assessment is the comparison of genetic sequences with the sequences of known allergens. GM sequences are compared step by step with the full length of know allergen sequences. Perhaps self evidently such comparisons are known as 'sliding window searches'.

This paper concerns a theoretical analysis of the reliability of the sliding window search as a function of window size (number of sequential amino acids, n) and the size of the database. Databases of mock amino acid sequences and databases of allergen sequences were tested for reliability.

There were no reports of false negatives. I.e. no true matches were missed. On the other hand the rate of false positives was high when n = 5 (of the order of 50%) only dropping below 1% when n was ≥ 7 and reliably less than 1% when n = 9.

The author cites this as evidence that the sliding window technique identifies matches as a matter of chance when the window used has n ≤ 8 and concludes that:

> *this adds little value to allergy assessments for newly expressed proteins.*

Comment
The risk seems to us to be a commercial one. Potentially useful sequences are probably being left unused because of random matches with known allergens using a window size of 8 amino acids. This should be reassuring, especially as the research paper does not report any false negatives. The balance between commercial risk and health protection has led to strong proposals for the adoption of n = 8.

The more false positives the more likely desktop risk analysis will be ignored by developers. Regulators would then have to rely much more on serum and animal testing.

In practice, whether or not a GM product is allergenic or not cannot be determined with 100% reliability by such comparisons of genetic sequences. Many other processes can lead to the production of allergens in modified plants and animals, where none existed in the unmodified organism.