Quantitative prediction of the allergenicity of new or modified proteins

AG Kruizinga[1*], RH Klein Entink[1], B Fabriek[1], WM Blom[1], GF Houben[1], JHM van Bilsen[1]

1 TNO, PO Box 360, 3700 AJ Zeist

** astrid.kruizinga@tno.nl, 0031 888665048*

TNO innovation for life

BACKGROUND

Currently, sustainable food production, including alternative protein sources, is receiving a lot of attention and exploration of these sources is widely supported by international organizations (1). Existing guidelines to assess the allergic potential of these sources are restricted to a qualitative assessment of allergenicity (2,3,4). One of the major challenges is to accurately and quantitatively assess the allergenicity of new or modified proteins. Therefore, a quantitative statistical model which predicts the relative allergenicity of a new protein compared to known allergenic products and proteins is being developed.

METHODS

As a first step, a scale for the relative allergenicity has been developed using data on potency and prevalence from a set of high/low allergenic products. Data on prevalence were collected by performing a literature search. Data on potency were retrieved from previous work (5) and new (unpublished) data.

As a second step, the first version of the statistical predictive model has been developed. To this end, physico-chemical characteristics and biological experimental data were analyzed from a set of high/low allergenic proteins (figure 1 gives an example).

Joint model.R | Joint Scale Analysis.R | Predictive Model Product Scale.R | scale.R | PD | FCDATA — 76 observations of 39 variables

protein	AA	Ala	Aliphatic.index	Arg	Asn	Asp	C	Cys	Gln	Glu	Gly	H	His	Hydropathicity	Ile	Instability.index
ara h 3	507	6.3	61.76	11.0	7.7	4.9	2521.00	0.8	8.7	9.9	7.1	3902.00	2.0	-1.068	3.7	80.29
ara h 1	601	5.0	59.05	9.5	6.8	5.0	2949.00	1.2	7.2	11.3	7.2	4679.00	2.3	-1.164	4.0	63.94
ara h 2	151	2.6	38.15	13.9	4.6	7.9	740.00	5.3	15.9	9.3	5.3	1166.00	1.3	-1.681	1.3	84.63
ara h 4	538	6.3	64.37	9.7	7.2	5.0	2688.00	1.1	9.5	8.9	7.2	4138.00	1.9	-0.923	3.7	74.23
profilin	130	6.2	89.31	0.8	3.1	5.4	623.00	1.5	5.4	6.2	13.8	968.00	2.3	-0.093	10.0	30.06
ara h 6	124	1.6	45.56	13.7	4.8	8.1	594.00	8.1	16.1	8.9	4.0	975.00	0.8	-1.277	3.2	67.15
ara h 7	144	2.1	36.60	16.0	5.6	6.9	713.00	5.6	16.7	9.0	4.9	1128.00	1.4	-1.705	2.1	82.74
ara h 8	157	7.0	83.82	0.6	3.8	5.7	763.00	0.0	0.6	10.2	10.8	1211.00	1.3	-0.365	7.6	21.80
oleosin	176	8.0	105.74	5.1	0.6	5.1	818.00	0.6	5.7	1.1	10.8	1353.00	1.7	0.289	7.4	43.12
b-conglycinin alpha chain	577	4.0	64.70	6.6	6.4	4.9	2936.00	0.2	9.0	13.7	5.0	4566.00	3.5	-1.147	4.9	67.75
Glycinin	172	5.8	87.73	7.0	9.3	2.3	840.00	1.2	6.4	4.7	8.7	1332.00	1.7	-0.357	3.5	40.41
Glycinin G1	495	7.8	87.78	5.0	10.0	3.3	883.00	1.1	6.7	5.0	5.6	1387.00	1.7	-0.239	6.1	45.10
hull protein	39	5.1	120.00	2.6	2.6	10.3	NA	0.0	0.0	0.0	12.8	NA	0.0	0.559	10.3	28.94
Hydrophobic seed protein	80	6.2	130.50	6.2	7.5	5.0	357.00	10.0	2.5	1.2	10.0	607.00	0.0	0.605	11.2	64.28
kunitz-trypsin inh = sam22	158	11.4	83.42	0.0	5.1	7.6	753.00	0.0	0.6	8.2	9.5	1179.00	1.3	-0.272	5.1	21.44
profilin- 1	130	6.9	81.77	1.5	3.1	6.9	623.00	1.5	5.4	5.4	13.8	959.00	2.3	-0.198	8.5	26.63
profilin- 2	130	6.2	87.00	1.5	3.1	6.2	625.00	1.5	5.4	5.4	14.6	967.00	2.3	-0.118	8.5	27.63
soy lectin	253	8.3	89.05	2.4	6.3	6.7	1248.00	0.0	2.4	3.6	5.1	1915.00	2.4	-0.107	5.9	34.75
vacuolar protein (p34)	379	6.1	74.12	3.4	4.2	6.1	1902.00	1.8	4.2	7.4	8.2	2898.00	3.7	-0.524	5.5	41.68
Gly m Bd 28K	473	7.2	81.46	4.2	3.2	5.9	2367.00	0.6	2.7	7.4	7.8	3678.00	2.5	-0.265	4.9	46.48

Figure 1: Example of the physico-chemical characteristics and biological experimental data collected.

A Bayesian statistical model combined these data with information on possible relationships between these characteristics and allergenicity retrieved from literature, to identify relationships between the characteristics of known allergenic proteins and the relative allergenicity scale for allergenic products. Since the model is still in development, for now it is assumed that all proteins in one allergenic product equally contribute to the relative allergenicity of this product. As a result, the model establishes a relationship between the physico-chemical characteristics of a protein and the allergenicity scale of a protein (figure 2).

Figure 2: Overview of the model to quantitatively predict the allergenicity of new or modified proteins.

RESULTS

The two dimensions of the allergenic product scale were defined as 1) the prevalence expressed as the percentage of the general population that is allergic and 2) the potency of the allergenic product expressed as the ED50 in mg-protein of the allergen (the dose that provokes objective allergic responses in 50% of allergic subjects). In terms of hazard characterization, severity of symptoms was not included, since this is a result of the nature of the exposure and the sensitivity of the individual rather than an inherent and independent attribute of the allergen. The two-dimensional scale was developed by ranking 18 allergenic products for potency and prevalence (figure 3).

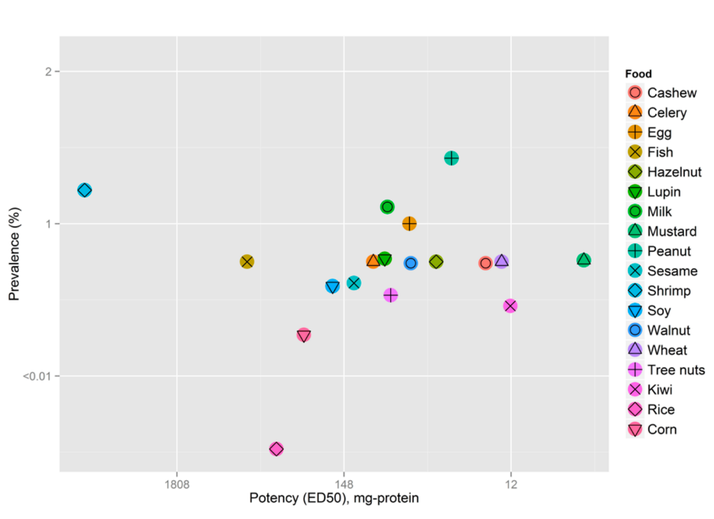

Figure 3: The allergenic product scale.

Relationships between the physico-chemical characteristics and biological experimental data and the allergenic product scale were established for both potency (figure 4) and prevalence. The relationships for prevalence were found to be less evident. By combining these relationships it is possible to place the new protein on the allergenic product scale.

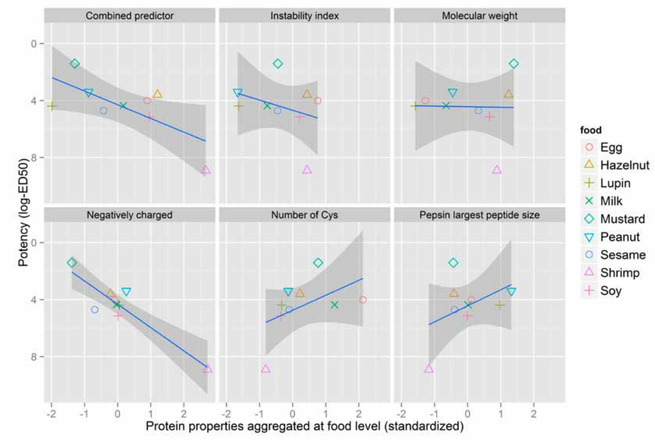

Figure 4: Relationship between physico-chemical characteristics and biological experimental data and the allergenic product scale. Combined predictor is based on aliphatic index, hydropathicity and the other factors in this figure.

CONCLUSION

The development of this model is promising in enabling the projection of new proteins on the scale and assess their allergenic hazard relative to the known allergens allowing a quantitative assessment of the relative allergenicity of a new protein.

REFERENCES

(1) Van Huis A. et al. Edible insects: future prospects for food and feed security. Food and agriculture organization of the United Nations. Rome, 2013.

(2) FAO/WHO: Report of Joint FAO/WHO Expert Consultation on Foods Derived from Biotechnology, 22-25 January. 2001.

(3) Stadler M.B. and Stadler B.M.. Allergenicity prediction by protein sequence. FASEB J, 2003;17:1141-1143.

(4) EFSA panel of Genetically Modified Organisms: Scientific opinion on the assessment of allergenicity of GM plants and microorganisms and derived food and feed. EFSA Journal 2010, 8(7):1700.

(5) Allen KJ et al.. Allergen reference doses for precautionary labelling (VITAL 2.0): clinical implications. JACI. 2014;133(1):156-164.

CONFLICT OF INTEREST

In relation to this presentation, I declare the following, real or perceived conflict of interest: This study was conducted within the framework of the Food Safety Program of TNO, with financial support from the Dutch Ministry of Health.