

アレルゲンデータベースAllergen Database for Food Safety (ADFS) のデータ改訂とアレルゲン性予測ツールの信頼性評価

中村亮介[#], 中村里香, 手島玲子

Major revision of the Allergen Database for Food Safety (ADFS) and validation of the motif-based allergenicity prediction tool

Ryosuke Nakamura[#], Rika Nakamura and Reiko Teshima

We have been maintaining an integral web server system, the Allergen Database for Food Safety (ADFS), since 2005 (<http://allergen.nihs.go.jp/ADFS/>). Recently, a group at the University of Nebraska-Lincoln released a new version of an allergen database, AllergenOnline. This database includes more than 1,300 allergens, all of which have been peer-reviewed by an international board of allergology experts. Here, we have totally revised the dataset of the ADFS by comparing it with that of AllergenOnline to improve the reliability of our allergen data. Moreover, the performance of our web-based tool for predicting new allergens (motif-based method), which was developed according to a theory proposed by Stadler & Stadler (2003), was validated using three methods. As a result of the integration of this allergen data, the number of (iso)allergens in the ADFS has increased to 1340, and epitope information is now available for 76 allergens. Using model datasets, the precision, recall, and specificity of our motif-based allergenicity prediction tool was proved to be 100.0%, 99.4%, and 100.0%, respectively. These results were similar to those for the original motif-based prediction model that was previously reported and are much better than those of the method recommended by FAO/WHO, especially with regard to the precision of predictions.

Keywords: allergen, database, allergenicity prediction, epitope, motif

1. 背景

アレルゲンタンパク質のエピトープや立体構造等に関するデータベースの開発は、既知アレルゲン同士の交差反応性の予測や、遺伝子組み換え作物に新規に組み込まれるタンパク質の潜在的アレルゲン性の予測など、様々な情報を容易に取得することを可能とする。この目的のため、我々は2005年に、当時の既存アレルゲンデータベースの情報と独自に精査した文献情報とを元に、新しいアレルゲンデータベース、Allergen Database for Food Safety (ADFS) を作成し、公開した¹⁾。さらに、その内容を本誌にて詳述した²⁾。以降、PSI-BLASTに代表される様々なインターフェースの追加・修正作業を行なうと共に、毎年1回、アレルゲンとエピトープに関する最新のデータを追加してきたが、一度登録されたデータの再校正や削除などは行なってこなかったため、いくつかのアレルゲンデータが最新の情報と食い違う可能性が懸念

されていた。また、そもそも2005年当時に存在していた他のデータベースにおけるアレルゲン登録に関するクライテリアも各データベースにより様々であり、統一されてはいなかった。

近年、米国ネブラスカリンカーン大学がアレルゲンデータベース「AllergenOnline (以下AOL)」を作成し、公開している³⁾。このデータベースは、登録アレルゲンの全てが国際的なアレルギーの専門家チームによるピアレビューを経ていることを大きな特徴としている。よって、登録タンパク質が真にアレルゲンであるかどうかというエビデンスの信頼性において、AOLは現在知られているアレルゲンデータベースの中でも際だっていると思われた。

しかし、AOLのアレルゲンデータはIgEエピトープに関する情報を含んでおらず、また、潜在的アレルゲン性予測ツールとしても、80残基のアミノ酸配列スライディングウィンドウ中35%以上の一致性を示した場合または連続する8残基の完全一致をもって陽性とする、いわゆるFAO/WHOの方法⁴⁾しか提供していない。これに対しADFSでは、筆者らによる査読を通じて得たエピトープ情報(線形・コンフォメーション・ミモトープ(エピ

[#] To whom correspondance should be addressed:

Ryosuke Nakamura; 1-18-1 Kamiyoga, Setagaya-ku, Tokyo
158-8501 Japan
Tel: +81-3-3700-9437, Fax: +81-3-3707-6950
E-mail: ryosnak@nihs.go.jp

トープ模倣配列)・糖鎖) や、UniProtの情報に基づくタンパク質の構造情報(ドメイン・立体構造・糖鎖等)などが一部のアレルギーに付与されており、また、アレルギー性予測法についても、FAO/WHOの方法に加え、アレルギーの共通モチーフに着目したStadlerらの理論⁵⁾を元にした、モチーフに基づく予測ツール(Motif-based法)も提供している。

このように、登録アレルギーの信頼性はAOLの優位性が認められるものの、そこに付随する情報や解析ツールの充実度は、多くの面でADFSが上回っていると思われる。そこで、今回AOLのアレルギーデータベースをADFSのそれと比較し、ADFSのアレルギーデータをピアレビューを経たものに統合することを試みた。また、これに伴い新たに作製したアレルギーモチーフを用い、ADFSのモチーフに基づく予測ツールの精度、感度および確度に関する評価を行なった。

2. 方法

2.1 AOLアレルギーデータの取得

AOLにおけるデータは、2008年7月の時点で取得した。ADFSはUniProtデータベースに基づく運用システムを採用しているが、AOLはアレルギーシークエンスをすべてNCBIのgene identifier (gi) 番号により管理している。よって、AOLとADFSのデータの比較・統合を行なうため、AOLのgi番号を対応するUniProt IDへと変換した。UniProtに対応するエントリが存在しない場合は、RefSeq, GenPept, PDB等のデータベースよりシークエンス情報を取得した。

アレルギー情報はアレルギー名とそのアミノ酸配列(UniProtのアクセッション番号)によって整理され、同一名で同一シークエンスを持つものは基本的に一つのエントリにまとめられる。なお、この条件の中でも、エピトープ情報や立体構造情報等の有無に違いがある場合は、エントリを分けて表示している。International Union of Immunological Societies (IUIS) が発行する正式な(イソ)アレルギー名が存在しない場合は、由来する生物の属と種の頭文字をそれぞれ3文字と1(または2)文字で表し、末尾に「?」を付した。

ADFSでは登録アレルギーを従来8種のカテゴリに分類していたが、AOLのデータを取り込むにあたり、アレルギーのカテゴリも、AOLによる分類に合わせ、全13種(aero animal, aero fungi, aero insect, aero mite, aero plant, contact, food animal, food fungi, food plant, gliadin, protozoan, venom/salivary, worm)とした。

2.2 エピトープ情報の取得

NCBI PubMedにおいて、下記のキーワード群を一定の検索式に当てはめて検索した文献から、アレルギーの

IgEエピトープ情報が記載されているものを筆者らが実際に査読し、エピトープ情報とした。

Immunoglobulin E, IgE-binding, Epitope, Identification, Epitope Mapping, Linear (Sequence Conformational, Structural, Discontinuous, three-dimensional), Analysis, Peptide, Recognition, Mimotope (Bacteriophages, Phage, Display) エピトープは、その性状から線形(L)、コンフォメーション(C)、および糖鎖(S)に分類し、さらに、実験方法などの情報を記載した。なお、ミモトープに関しては、コンフォメーションエピトープの一種として分類した。糖鎖エピトープは、単なる交差反応により結合するいわゆるCCDではなく、サイトカイン産生や脱顆粒を誘導するなど、何らかの生物活性を持つと認められた場合のみを示した。

2.3 アレルギーモチーフの抽出

Stadlerらの報告⁵⁾に従い、ADFSが収載する923本のアレルギーシークエンスから、複数のアレルギーに共通なモチーフ構造を抽出した。モチーフ抽出ツールはMEME(ver. 3.0.13)を用いた。その際、アミノ酸残基数は50、E-valueは0.01、モードはzoopsを使用した。1つのモチーフが抽出されるごとに、抽出に用いられたアレルギーシークエンスを除き、残りのシークエンスにより再度モチーフを抽出するという工程を繰り返し、最終的にE-value 0.01で一つもモチーフがヒットしなくなるまでこの作業を繰り返した。

2.4 Motif-based法によるアレルギー性予測の信頼性評価

ADFSが独自に提供しているアレルギー性予測ツールであるMotif-based法による解析手法の信頼性について、下記の3通りの手法により評価を行なった。

1. 人工生成非アレルギーデータベースを用いた評価: ADFS収載アレルギーのうち、アミノ酸残基数が26残基以上の902種を真のアレルギーとし、この配列を、(A) 逆順に並べたもの、(B) window size 20でシャッフルしたもの、(C) window sizeなしでシャッフルしたもの、の3つを真の非アレルギーと定義した。アミノ酸残基長を26残基以上としたのは、Stadlerらが同法を発表した際と同一の条件にそろえるためであり、これは当時知られていた最もアミノ酸長の短いアレルギー(ハチ毒アレルギーApi m 3)の長さ(26残基)に基づいている⁵⁾。これらのアレルギーおよび非アレルギーデータベースを対象に、BLAST E-valueカットオフ値を様々に変えながらMotif-based法によるアレルギー性予測を行なった結果について、真陽性(TP)、真陰性(TN)、偽陽性(FP)および偽陰性(FN)を元に、次の式で定義される精度(Precision)、感度(Recall; 再現性)、および確度(Specificity)を算出した。

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP})$$

2. 10分割交差試験法：前述の902種のアレルゲンデータをランダムに10分割し、一つ一つのデータセットについて、残りの9個のデータセットを用いて抽出したモチーフとシーケンスにより試験を行なった。

3. とうもろこしタンパク質のアレルゲン性：とうもろこしが発現するタンパク質をランダムに50種選び、FAO/WHO法およびMotif-based法により解析を行なった。

3. 結果および考察

3.1 ADFSのデータ改訂後のスペック

今回のAOLとのアレルゲンデータ照合の後、ADFSにおける記載アレルゲンおよびイソアレルゲンの総数は1340種（うち、UniProtとの対応が付いたものは923種）となった。また、何らかのエピトープ情報を持つアレルゲンおよびイソアレルゲンの数は76種となり、これは現時点で存在するアレルゲンのエピトープ情報を集めたデータベースとしては世界でも最大である。

今回、アレルゲンの糖鎖自体がIgEと結合し、好塩基球からのヒスタミン遊離を起こす場合（Ole e 1）やIL-4産生を誘導する場合（Phl p 1）などについてエピトープ情報を追加した。多くのIgE結合性糖鎖は生物活性を持たず、*in vitro*の臨床試験におけるバックグラウンドの上昇に寄与していることが知られている⁶⁾。しかし少なくとも一部のアレルゲンにおける糖鎖は上記のようにアレルギーの発症に関与すると考えられており、アレルゲンデータベースが糖鎖エピトープの情報を記載することは望ましいと考えられる。このようなアレルゲンデータベースは、今のところADFS以外に存在しない。

3.2 Motif-based法

ADFSでは、任意のアミノ酸配列に関する潜在的アレルゲン性を予測するため、いわゆるFAO/WHOの方法⁴⁾とMotif-based法の2種類のインターフェイスを提供している。前者については他のアレルゲンデータベースにおいてもしばしば実装されているが、Motif-based法によるアレルゲン性予測については、Stadlerらによる報告⁵⁾があるのみで、ウェブ上でユーザが任意のクエリ（問い合わせ）配列を解析できるツールはこれまで開発されていなかった。

我々は、Stadlerらの報告に従い、MEMEソフトウェアにより923本のアレルゲンシーケンスから58種のアレルゲンモチーフを抽出した。その際、141本のシーケンスがモチーフに抽出されずに残った（Fig. 1）。本法は、クエリアミノ酸配列をpftoolソフトウェアにより58種のアレルゲンモチーフに対してプロファイル検索し、さらに141本のアレルゲンシーケンスに対してBLASTによる

ペアワイズ検索を行なう手法である。この手法について、次の3種類の方法によってその予測信頼性を解析した。

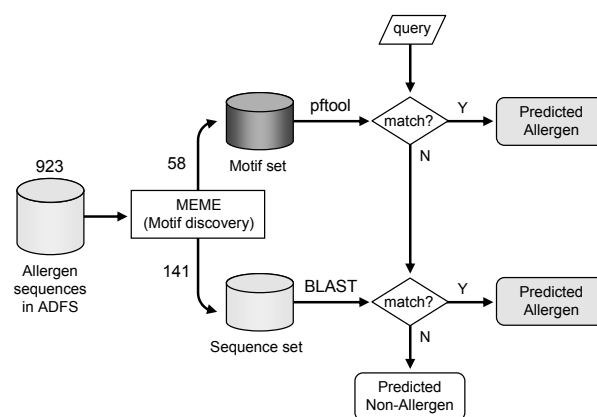


Fig. 1. Schematic of motif-based allergenicity prediction method.

Allergen sequences in the ADFS (923) were submitted to MEME, and 58 allergen motifs were extracted. Of the 923 sequences, 782 matched one or more of the allergen motifs, while the remaining 141 sequences did not match any of the motifs. The query sequence is first compared to the 58 motifs using the pftool; then, it is compared to the remaining 141 sequences using BLAST pairwise alignment tool. The query sequence is predicted to be a potential allergen if it is identical to either an allergen motif or an allergen sequence.

3.3 人工生成非アレルゲンデータセットを用いた評価

まず、ADFSに記載された902本の真のアレルゲンアミノ酸配列と、その3倍量に当たる真の非アレルゲンアミノ酸配列とを用い、本ツールがどれだけ正しくアレルゲン性を予測できるかについて調べた。Motif-based法では第一段階にクエリ配列とモチーフとの比較を行ない、第二段階ではBLASTによるペアワイズ検索を行なう。そこで、BLASTのマトリクスやギャップペナルティなどの各種パラメータ、およびE-valueカットオフ値を様々に変化させ、最もパフォーマンスの優れたBLASTのパラメータ設定を探索したところ、マトリクスとしてBlosum80、開始および伸張ギャップペナルティをそれぞれ13および2、フィルタをTrueに設定した場合であった。この条件でPrecision (%), Recall (%), Specificity (%)の合計が最大になるBLAST E-valueカットオフ値は 10^{-13} で、このときのそれぞれの値は100.0%, 99.4%, 100.0%と、極めて良好であった（Fig. 2）。この結果を受け、ADFSのMotif-basedアレルゲン性予測ツールにおけるBLASTのパラメータ設定は上記の組み合わせを使用することとした。E-valueカットオフ値のみはユーザが任意に変更できるが、これは後に述べる理由により 10^{-3} を初期値とした。

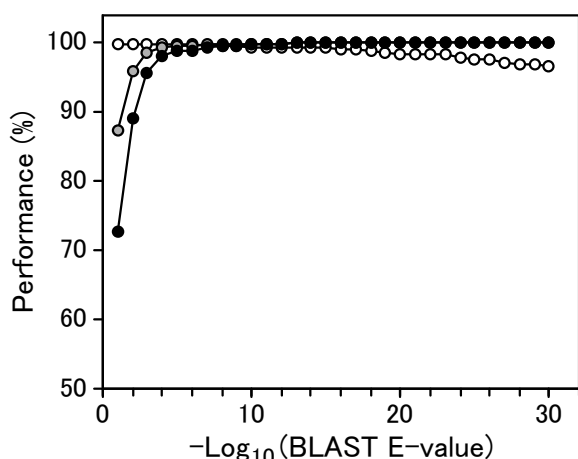


Fig. 2. Performance of motif-based allergenicity prediction method. Test sequences composed of 902 true allergens and 2706 true non-allergens were submitted to the motif-based allergenicity prediction tool in the ADFS. The prediction stringency was varied by adjusting the BLAST E-value cut-off. The best results were obtained when 10^{-13} was used as the E-value cut-off. See the Materials and Methods section for the definitions of Precision (○), Recall (○), and Specificity (●).

3.4 10分割交差試験法

上記の試験は、リファレンスとなるデータベース中にクエリアレルゲン配列自身が含まれていることになる。そこで、実際にアレルギー性が未知のタンパク質を本法で解析した場合を再現するため、次に10分割交差試験を行なった (Table 1)。これは、合計902種のアレルゲン配列をアルファベット順に10分割し、一つ一つのデータセットについて、自分自身を含まない他の9つのデータセットを用いて生成したモチーフおよびシークエンスにより解析を行なう手法である⁵⁾。各データセットにより Precision, Recall, Specificityの合計が最大になるBLAST E-valueカットオフ値は 10^{-15} から 10^{-3} までの間で変化した。これは、シークエンスを10分割した際のデータセットのばらつきによるものと推測される。実際にユーザが解析するクエリ配列は多様であることから、最適なE-valueカットオフ値を一つ設定することは困難であり、適切な値をユーザ自身が設定すべきであると考えられる。よってADFSでは、今回の最大値である 10^{-3} をもってBLAST E-valueカットオフ値の初期値とすることとした。この条件における10分割交差試験の解析結果を示したのがTable 1である。平均すると、Precision 96.5%, Recall 85.5%, Specificity 98.9%となった。Stadlerらが2003年に報告した時点では、同法のPrecisionおよびRecallはそれぞれ94.8%および86.2%であり、いわゆるFAO/WHOの方法

Set	Query sequence	Motif generation			Performance (%)		
		Reference	Motif	Sequence	Precision	Recall	Specificity
#01	93	809	65	133	95.6	93.5	98.6
#02	93	809	53	134	90.0	77.4	97.0
#03	93	809	57	133	94.7	77.4	98.6
#04	88	814	59	133	98.6	82.0	99.6
#05	84	818	56	126	98.6	85.7	99.6
#06	93	809	54	142	97.6	88.2	99.3
#07	93	809	55	140	100.0	90.3	100.0
#08	87	815	57	119	98.6	78.2	99.6
#09	93	809	55	134	93.2	88.2	97.8
#10	85	817	59	144	97.6	94.1	99.2
Total	902	-	-	-	96.5	85.5	98.9

Table 1. Ten-fold cross validation test

A test database (902 sequences) was randomly divided into 10 parts, and allergenicity prediction was performed for the sequences in each part, using the other nine parts as the allergen reference database. The appropriate BLAST E-value cut-off varied from 10^{-15} to 10^{-3} , depending on the dataset; only the results obtained using a cut-off of 10^{-3} are shown.

ではそれぞれ37.6～68.0%および92.2～97.0%であった (Specificityは不明)⁵⁾。これらの結果から、ADFSの Motif-based法によるアレルギー性予測ツールの信頼性

は、Stadlerの報告における成績とほぼ同程度で、FAO/WHO法よりも精度の面で大きく上回っていることが分かる。

ID	Corn Protein Locus	Description	Known Allergen	FAO/WHO			Motif-based			
				35%/80	cont. 6	cont. 7	cont. 8	Motif	BLAST	
1	A55092	catalase (EC 1.11.1.6) CAT-2 - maize (fragment).	No	0	1	0	0	0	0	
2	AAA33523	16-kDa zein protein.	No	0	15	3	0	0	0	
3	AAA68209	sus1 gene product.	No	0	6	0	0	0	0	
4	AAA73960	kaurene synthase A.	No	0	5	0	0	0	0	
5	AAA87580	cytosolic glyceroldehyde-3-phosphate dehydrogenase GAPC4.	No	0	4	0	0	0	0	
6	AAB71078	acidic ribosomal protein P3a [Zea mays].	No	0	25	7	4	0	0	
7	AAB86960	profilin [Zea mays].	Yes	65	52	49	49	#002	0	
8	AAC72193	pyruvate dehydrogenase E1 beta subunit isoform 2 [Zea mays].	No	0	2	1	0	0	0	
9	AAC78468	RNA polymerase sigma factor 2 [Zea mays].	No	0	11	0	0	0	0	
10	AAC79953	anionic peroxidase H [Zea mays].	No	0	2	1	0	0	0	
11	AAK26754	plasma membrane integral protein ZmPIP1-3 [Zea mays].	No	0	2	0	0	0	0	
12	AAK30114	teosinte branched protein 1 [Zea mays].	No	0	2	0	0	0	0	
13	AAK51777	MURB-like protein hMURB3 [Zea mays].	No	0	3	0	0	0	0	
14	AAK51786	MURB-like protein hMURB12 [Zea mays].	No	0	3	0	0	0	0	
15	AAK56122	alpha-expansin 4 [Zea mays].	No	0	3	2	2	0	0	
16	AAK56130	beta-expansin 7 [Zea mays].	Yes	26	36	27	17	#008	1	
17	AAK59898	kaurene synthase A [Zea mays].	No	0	3	0	0	0	0	
18	AAK60245	teosinte branched1 protein [Zea mays].	No	0	4	1	0	0	0	
19	AAK60502	sucrose export defective 1 [Zea mays].	No	0	8	0	0	0	0	
20	BAA05550	group 3 Lea protein MGL3 [Zea mays].	No	0	3	0	0	0	0	
21	BAA22410	calcium-dependent protein kinase-related kinase [Zea mays].	No	0	4	2	0	0	0	
22	CAA28734	40.1 kD A1 protein [Zea mays].	No	0	7	0	0	0	0	
23	CAA31221	unnamed protein product [Zea mays].	No	0	2	1	0	0	0	
24	CAA35589	pyruvate decarboxylase [Zea mays].	No	0	22	4	0	0	0	
25	CAA37038	dihydrodipicolinate [Zea mays].	No	0	1	0	0	0	0	
26	CAA39438	ribosomal protein S11 [Zea mays].	No	0	9	4	0	0	0	
27	CAA60366	hypothetical protein [Zea mays].	No	0	1	0	0	0	0	
28	CAA72196	cytochrome p450 [Zea mays].	No	0	13	1	0	0	0	
29	CAA87634	orf [Zea mays].	No	0	6	0	0	0	0	
30	CAB56627	SBP-domain protein 1 [Zea mays].	No	0	8	0	0	0	0	
31	CAC18100	putative legumain [Zea mays].	No	0	2	0	0	0	0	
32	CAC35069	VIP3 protein [Zea mays].	No	0	0	0	0	0	0	
33	JC1524	major allergen mI protein - maize.	Yes	23	28	25	24	0	3	
34	O24578	Adenylosuccinate synthetase, chloroplast precursor (AdSS) (IMP--aspartate ligase) (AMPSase).	No	0	5	0	0	0	0	
35	P04705	Zein-alpha PZ19.1 precursor (19 kDa zein PZ19.1).	No	1	11	0	0	0	0	
36	P15590	Globulin-1 S allele precursor (GLB1-S) (7S-like).	No*	6	16	8	2	0	1	
37	P33489	Auxin-binding protein 5 precursor (ABP) (ERABP5).	No	0	7	0	0	0	0	
38	P40280	Histone H2A.	No	0	2	1		0	0	
39	P46251	Actin-depolymerizing factor 1 (ZmADF1) (ADF-1) (ZmABP1).	No	0	0	0	0	0	0	
40	P51059	Phosphoenolpyruvate carboxylase 2 (PEPCase 2) (PEPC 2).	No	0	3	1	0	0	0	
41	PWZMAM	H+-transporting two-sector ATPase (EC 3.6.3.14) alpha chain - maize mitochondrion.	No	0	0	0	0	0	0	
42	S12043	leucoanthocyanidin dioxygenase (EC 1.14.11.-) - maize.	No	0	6	0	0	0	0	
43	S30062	polygalacturonase - maize.	Yes**	6	6	6	5	0	0	
44	S37379	catalase (EC 1.11.1.6) 3 - maize.	No	0	4	0	0	0	0	
45	S58532	matK protein (trnK intron) - maize chloroplast.	No	0	6	0	0	0	0	
46	T02242	wEST02963 Early embryo, Stratagene (cat. #937007) Caenorhabditis elegans cDNA clone CESAB12, mRNA sequence.	No	0	3	1	1	0	0	
47	T02763	0100M7 gmbP1HB3.1, G. Roman Reddy Plasmodium falciparum genomic clone PF0100M, genomic survey sequence.	No	0	6	1	1	0	0	
48	T02990	FB19G7 Fetal brain, Stratagene Homo sapiens cDNA clone FB19G7 3'end, mRNA sequence.	No	0	10	0	0	0	0	
49	T02993	FB1A10 Fetal brain, Stratagene Homo sapiens cDNA clone FB1A10, mRNA sequence.	No	1	9	3	0	0	1	
50	T03397	IB217 Infant brain, Bento Soares Homo sapiens cDNA clone IB217 3'end, mRNA sequence.	No	0	1	0	0	0	0	
Total positive:				4	7	47	21	9	2	4

Table 2. Comparison of FAO/WHO method and motif-based method using 50 corn proteins.

Of the 50 randomly selected corn proteins [6] that are shown, known allergens are indicated by a "Yes". For the FAO/WHO method, the numbers of matched allergens in the ADFS are presented according to the criteria shown in the top row (more than 35% identity over 80 or more sliding windows, exact match of 6, 7, or 8 contiguous amino acids). For the motif-based method, the motif ID in the ADFS and/or the numbers of matching allergen sequences (BLAST E-value, 10^{-3}) are presented.

*GLB1-S, a 7S-globulin, has been reported as an allergen found in several organisms other than corn.

**Corn polygalacturonase has several UniProt IDs, and some entries with longer amino acid sequences than S30062 are known as the allergen Zea m 13.

3.5 とうもろこしタンパク質のアレルゲン性

最後に、実際の作物由来タンパク質のアレルゲン性を本法により調べるため、Hilemanら⁷⁾が報告の中でランダムに抽出した50種類のとうもろこしタンパク質について、FAO/WHO法と本法とを比較した (Table 2)。FAO/WHO法においては、連続する6アミノ酸残基の完全一致をもって判定すると、50種のタンパク質中47種がアレルゲンであると判断され、偽陽性が極めて多いということが分かる。これに対しMotif-based法では、アレルゲンであることが既知であるタンパク質は、一つ (polygalacturonase) を除きすべてについて陽性判定を出していると同時に、クエリタンパク質自体のアレルゲン性は知られていないが他の生物種におけるアナログがアレルゲンであるような場合 (7S-globulin like GLB1-S) を正しく検出できていることが分かる。Polygalacturonaseについては、本表で用いたクエリアミノ酸配列 (S30062; 95残基) より長いアミノ酸長を持つ複数のUniProtエントリが存在し (404 ~ 411残基)、これらがとうもろこしアレルゲン Zea m 13として知られている⁸⁾。ADFSではこれら長い方のエントリのみがアレルゲンとして登録されており、そのモチーフは#020として抽出されているが、その抽出モチーフ (50残基) がクエリアミノ酸配列の領域と一致しなかったため、陰性として判定されてしまった (Table 2)。しかし、このような場合でも、クエリアミノ酸長の十分長いエントリ (Q1ZYQ5等) を用いることにより、正しいモチーフがヒットすることを確認している。

FAO/WHO法は、遺伝子組み換え作物に新規に導入するタンパク質のアレルゲン性を予測する上では現在でも標準的な手法とされているが、その偽陽性の多さゆえに、批判も根強い⁹⁾。Stadlerらが開発したMotif-based法はFAO/WHO法の欠点を大きく改善した有望な手法であり、それをウェブツールとして公開しているのは現在のところADFSのみである。しかし、ADFSのMotif-based法のパフォーマンスを定量的に評価することはこれまで困難であった。今回、様々な角度から同法の評価を行ない、ADFSのMotif-based法がStadlerらが報告したものと同程度のパフォーマンスを有していることが分かった。ADFSは公開以来毎年一回のペースでデータのアップデートを行っており、そのたびにアレルゲンモチーフやシーケンスも更新されている。今後もこの作業を継続し、潜在的アレルゲン性予測の信頼性向上に努めたい。

4. 参考文献

- 1) URL: <http://allergen.nihs.go.jp/ADFS/>
- 2) Nakamura, R., Teshima, R., Tkagi, K., Sawada, J.: *Bull. Natl. Inst. Health Sci.*, **123**, 32-36 (2005)
- 3) URL: <http://www.allergenonline.com/>

- 4) Codex Alimentarius Commission, Alinorm 03/34, pp47-60 (2003)
- 5) Stadler, M.B. & Stadler, B.M.: *FASEB J.*, **17**, 1141-1143 (2003)
- 6) van Ree, R.: *Int. Arch. Allergy Immunol.*, **129**, 189-197 (2002)
- 7) Hileman, R.E., Silvanovich, A., Goodman, R.E., Rice, E.A., Holleschak, G., Astwood, J.D., Hefle, S.L.: *Int. Arch. Allergy Immunol.*, **128**, 280-291 (2002)
- 8) Petersen, A., Dresselhaus, T., Grobe, K., Becker, W-M.: *Proteomics*, **6**, 6317-6325 (2006)
- 9) Goodman, R.E., Vieths, S., Sampson, H.A., Hill, D., Ebisawa, M., Taylor, S.L., van Ree, R.: *Nature Biotechnol.*, **26**, 73-81 (2008)