

厚生労働科学研究費補助金（食品の安全性確保推進研究事業）  
「新たなバイオテクノロジーを用いて得られた食品の安全性確保とリスクコミュニケーション」  
分担研究報告書（令和3年度）

新規アレルゲン性予測手法開発のための基盤的研究

研究分担者 近藤 一成 国立医薬品食品衛生研究所生化学部 部長  
研究分担者 安達 玲子 国立医薬品食品衛生研究所生化学部 室長  
研究分担者 爲廣 紀正 国立医薬品食品衛生研究所生化学部 主任研究官

研究要旨：

本研究では、遺伝子改変技術応用食品のアレルゲン性について、より高い精度での評価・予測を可能とすることを目的として、アレルゲン性予測手法-1（allerStat）の検証、及びアレルゲン性予測手法-2（MHC-II 結合性予測手法）開発のためのデータセットの準備を行った。allerStat に関しては、アレルゲン特異的アミノ酸配列パターン（連結 ASP）と既知エピトープとのアミノ酸配列比較を行ったところ、連結 ASP の 24.3%は既知エピトープと高い相同性を有することが示された。また、国立医薬品食品衛生研究所にて運用・公開しているアレルゲンデータベース（Allergen Database for Food Safety, ADFS）に関して、令和2年6月から令和3年5月までの1年間に NCBI PubMed に掲載された論文から、エピトープ配列決定に関する9報のピアレビューを行い、6種のアレルゲンについて、総数27のエピトープ情報を ADFS に追加し、データベースの更新を行った。これらの情報更新により ADFS のアレルゲン及びイソアレルゲンのアミノ酸配列情報は2,379、エピトープ既知のアレルゲン数は260となった。また、ADFS ウェブページのデザインを全面的に更新し、遺伝子改変技術応用食品のアレルゲン性評価に有用なデータベースである ADFS を充実させることができた。

また、allerStat 実用化のためにプログラムの改良、および、判定器の拡充を行いより高精度なアレルゲン予測システムの構築を行った。

A. 研究目的

遺伝子改変技術を応用した食品開発は、技術的には、外来遺伝子導入による遺伝子組換え食品から、内在性遺伝子の改変を行うゲノム編集技術応用食品へ、また、酵母等に多数の外来遺伝子を導入し新規食品機能成分を産生させる合成生物学の利用へと変化している。現在、ゲノム編集技術では多様な手法が生み出されており、これらの手法による意図しない塩基変化も一様ではないことが明らかになりつつある。従って、このような意図しない変化、及びそこから生じる代謝成分の変化を検出または予測し、その変化が与える影響を正確に評価することは、食品の安全性確保において急務の課題である。

バイオテクノロジー技術を用いて開発された食

品のリスクの1つに、アレルゲン性増大の可能性がある。本研究では、国立医薬品食品衛生研究所生化学部にて管理・公開している、アレルゲン性予測機能（FAO/WHO 法等）を装備したアレルゲン・エピトープ情報データベース（Allergen Database for Food Safety, ADFS）について、新規アレルゲン及びエピトープ情報の収集・解析等によりアレルゲン性評価に関する検討を行い、遺伝子改変技術応用食品のリスク評価に資するデータベースとなるよう、情報を更新し内容を充実させる。

また、AI を活用した新規高精度アレルゲン性予測手法の開発を進める。令和2年度までの先行研究班では、機械学習によりアレルゲン及び非アレルゲンタンパク質から抽出した特徴的なアミノ酸配列パターンを利用してアレルゲン性を予測する手法（アレルゲン性予測手法-1、allerStat）を開発してきた。本研究班では、この予測システムを検

た。

## B. 研究方法

### アレルギー性予測手法-1 (allerStat) の検証

先行研究班において開発を進めた allerStat の検証を行うため、抽出されたアレルギー特異的アミノ酸配列パターン (allergen-specific pattern, ASP) と既知のエピトープとの比較を行った。抽出された ASP のうち、アミノ酸配列がオーバーラップしているものは連結させて整理し、1 つのパターンとして取り扱った (連結 ASP)。連結 ASP と ADFS に記載されているエピトープ配列との配列の類似性を BlastP により検討した。

### アレルギー性予測手法-2 (MHC-II 結合性予測手法) 開発のためのデータセットの準備

機械学習/深層学習を利用して MHC-II とペプチドとの結合性予測手法を開発するための学習データセットについて精査し、他の研究分担者に情報提供を行った。

### ADFS 登録アレルギー (アミノ酸配列情報) のアップデート

米国ネブラスカ大学リンカーン校が運営しているアレルギーデータベース (AllergenOnline) における登録アレルギーのアップデート内容を、ADFS に反映させた。

### ADFS エピトープ情報の追加

令和2年6月から令和3年5月までの1年間に NCBI PubMed に掲載された論文から、キーワード検索により、エピトープ配列決定に関するものを抽出した。キーワードとしては、IgE、epitope、linear、conformational、sequence、recognition 等々のワードを使用し、これらを複数組み合わせで6通りの検索式を作成して検索を行った。この検索により抽出されてきた論文についてピアレビューを行った。その結果エピトープ情報を報告していると判断された論文について、そのエピトープ情報を整理し、ADFS のデータに追加した。

### ADFS ウェブページのデザイン更新

ADFS をより活用しやすいものとするため、ADFS ウェブページのデザインを全面的に更新した。

## C. 研究結果

### アレルギー性予測手法-1 (allerStat) の検証

先行研究班において開発を進めた、機械学習によりアレルギー及び非アレルギータンパク質から抽出した特徴的なアミノ酸配列パターンを利用してアレルギー性を予測する手法 (allerStat) の検証を行うため、アレルギータンパク質のアミノ酸配列から抽出された ASP と、ADFS に記載されている既知のエピトープとの配列比較を行った。allerStat において、アレルギーから抽出された ASP は 5,994 個であった。これらの中にはアミノ酸配列がオーバーラップしているものもあったため、このような配列を連結させて整理したところ、整理後の ASP (連結 ASP) の数は 1,072 個であった。連結 ASP のアミノ酸配列と既知エピトープとの配列を BlastP により比較したところ、連結 ASP のうちの 260 個 (24.3%) がエピトープ配列と高い相動性を示した。図 1 に配列比較例を挙げる。例えば、パターン 537 は、種実類のアレルギーにおいて同定された 7 つのエピトープと高い相同性を有しており、アレルギー間の交差反応性が示唆された。さらに、連結 ASP のうち 225 個 (20.1%) はエピトープ配列と完全に一致していることが示された。

### allerStat の実用化のための改良と高精度化

allerStat の結果表示において、現行システムでは数値のみが表示される。この値の意味付けが一般公開したときの課題であった。そこで、アレルギー及び非アレルギータンパク質の中で、自身の予測したタンパクアミノ酸配列が、その位置に相対的に位置するかを表示できるように改良した。これにより、自身の予測結果が、アレルギー性があるかどうかの判断が視覚的に可能になった。

### アレルギー性予測手法-2 (MHC-II 結合性予測手法) 開発のためのデータセットの準備

アレルギーによる感作の第一段階である、アレルギータンパク質由来のペプチドが抗原提示細胞表面の MHC-II 上に提示されるステップに着目し、ヒト MHC-II とペプチドとの結合性をアレルギー性予測に活用するため、機械学習/深層学習を利用して MHC-II とペプチドとの結合性予測手法を開発するための学習データセットについて精査した。既報の予測手法で使用されているデータセッ

ト等に関して、開発担当の研究分担者に情報提供を行った。

#### ADFS 登録アレルゲン（アミノ酸配列情報）のアップデート

米国ネブラスカ大学リンカーン校が運営しているアレルゲンデータベースである AllergenOnline 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おり、登録データの信頼性が非常に高いデータベースである（但しエピトープ情報は含まない）。ADFS における登録アレルゲンは平成 20 年度に AllergenOnline の登録アレルゲンと統合し、その後も AllergenOnline のアップデートに伴って ADFS 登録アレルゲンのアップデートを行っている。令和 3 年度においても引き続きこのアップデート作業を実施した。

#### ADFS エピトープ情報の追加

令和 2 年 6 月から令和 3 年 5 月までの 1 年間で、キーワード検索により抽出された論文は 18 報であった。その中からエピトープ情報が記載されていると思われる 9 報を選択し（表 1）、ピアレビューを行った。その結果、6 報の論文から 6 種のアレルゲンについて、総数 27 種のエピトープ情報を新たに追加した（表 2）。

上記のアレルゲン及びエピトープ情報更新作業により、最終的に、ADFS のアレルゲン及びイソアレルゲンのアミノ酸配列情報は 2,379、エピトープ既知のアレルゲン数は 260、構造既知のアレルゲン数は 183、糖鎖付加アレルゲン数は 127 となった。

#### ADFS ウェブページのデザイン更新

ADFS をより活用しやすいものとするため、ADFS ウェブページのデザインを全面的に更新した。その際、"Search by Country (Food Allergen Labeling)"や"Search by Taxonomy"という検索項目を新たに設定し、日本・米国・EU の食品表示制度における食物アレルゲン表示品目や主要なアレルギー原因食品のアレルゲンに関する検索がより容易に行えるようにした。

更新版 ADFS は、これまでと同じ URL にて令和 4 年 2 月 21 日に公開した。

(<https://allergen.nih.gov/jp/ADFS/database>)

## D. 考察

本研究では、AI を活用した新規アレルゲン性予測手法開発に向けて、アレルゲン性予測手法-1(allerStat)の検証、及びアレルゲン性予測手法-2(MHC-II 結合性予測手法) 開発のためのデータセットの準備を行った。allerStat に関しては、連結 ASP と既知エピトープとのアミノ酸配列比較を行ったところ、連結 ASP の 24.3%は既知エピトープと高い相同性を有することが示された。allerStat に関しては、現在細かな点の修正等プログラムの整備を進めており、今後の実用化・公開を目指す。

また、令和 2 年 6 月から令和 3 年 5 月までの 1 年間で、アレルゲン及びイソアレルゲンのアミノ酸配列情報を 40 種追加、また、6 種のアレルゲンについて総数 27 のエピトープ情報を追加した。

本研究では、AI を活用した 2 種のアレルゲン性予測手法の開発・実用化を進め、高精度アレルゲン性予測法の構築を目指し、また、アレルゲンデータベース ADFS のさらなる整備を行っている。本研究には、遺伝子改変技術を応用した食品のアレルゲン性をより高い精度で評価・予測するシステムの構築に資するものである。

## E. 結論

本研究では、遺伝子改変技術応用食品のアレルゲン性について、より高い精度での評価・予測を可能とすることを目的として、アレルゲン性予測手法-1(allerStat)の検証、及びアレルゲン性予測手法-2(MHC-II 結合性予測手法) 開発のためのデータセットの準備を行った。allerStat は、令和 4 年度内に一般公開予定で準備を進めており、より制度が高いアレルゲン予測システムとして広く活用されることが期待される。また、令和 2 年 6 月から令和 3 年 5 月までの 1 年間に NCBI PubMed に収載された論文から、エピトープ配列決定に関する 9 報のピアレビューを行い、6 種のアレルゲンについて、総数 27 のエピトープ情報を ADFS に追加した。また、AllergenOnline の登録アレルゲン（アミノ酸配列情報）に関するアップデートを ADFS に反映させた。これらの情報更新により、遺伝子改変技術応用食品のアレルゲン性評価に有用なデータベースである ADFS を充実させることができた。

## F. 研究発表

### 1. 論文発表

- 1) Okazaki F, Momma K, Hirakawa Y, Kawai N, Yamaguchi-Murakami Y, Adachi R, Mori Y, Kondo Y, Narita H. Determination of severe peach allergens, gibberellin-regulated protein, and lipid transfer protein, using monoclonal antibodies. J Nutr Sci Vitaminol, in press.

### 2. 学会発表

なし

## H. 知的財産権の出願・登録状況

### 1. 特許取得

なし

### 2. 実用新案登録

なし

### 3. その他

なし