

内閣府食品安全委員会
平成 15 年度食品安全確保総合調査

タンパク質のアレルギー評価手法に関する 調査報告書

平成 16 年 3 月

CTC ラボラトリーシステムズ株式会社

目次 -

| | | |
|-------|--------------------------------|----|
| 1 | 序論 | 4 |
| 1.1 | はじめに | 4 |
| 1.2 | 本調査の目的 | 5 |
| 1.3 | 本調査の概要 | 5 |
| 1.3.1 | 既知アレルゲン配列および構造に関する調査 | 5 |
| 1.3.2 | アレルゲン性予測手法に関する調査 | 5 |
| 1.4 | 本調査の実施期間 | 5 |
| 1.5 | 本調査の実施体制 | 6 |
| 2 | 既知アレルゲン配列および構造に関する調査報告 | 8 |
| 2.1 | 調査の目的と背景 | 8 |
| 2.2 | 利用可能なアレルゲンデータソース調査結果 | 8 |
| 2.3 | 既存データベースにおけるアレルゲンデータエントリー調査結果 | 15 |
| 2.3.1 | アレルゲンと関連する公共配列アクセッション番号の数 | 15 |
| 2.3.2 | SDAPに登録されているアレルゲンのエピトープ数 | 17 |
| 2.3.3 | 立体構造情報を持つアレルゲン数 | 20 |
| 2.3.4 | アレルゲンデータエントリー調査まとめ | 23 |
| 2.4 | エピトープ情報に関する文献調査方法 | 24 |
| 2.5 | エピトープ情報に関する文献調査報告 | 26 |
| 2.6 | 既知アレルゲン配列および構造に関する調査のまとめ | 29 |
| 2.7 | 既知アレルゲン配列および構造に関する調査からの考察 | 29 |
| 3 | アレルゲン性予測手法に関する調査報告 | 32 |
| 3.1 | 現在考えられているアレルゲン相同性検索方法の文献調査 | 32 |
| 3.1.1 | FAO/WHOのガイドラインに沿った方法(29) | 32 |
| 3.1.2 | Two-track approach(32) | 32 |
| 3.1.3 | Motif-based prediction(30) | 32 |
| 3.1.4 | 各相同性検索方法の比較とサンプル計算に利用する検索方法の選定 | 33 |
| 3.2 | 本調査におけるアレルゲン予測によるサンプル計算 | 34 |
| 3.2.1 | サンプル計算の全体の流れ | 34 |

| | | |
|-------|---------------------------------------|----|
| 3.2.2 | 計算機サーバーとアプリケーションの用意 | 35 |
| 3.3 | アレルゲンデータベースの作成 | 36 |
| 3.3.1 | データソース | 36 |
| 3.3.2 | 方法 | 36 |
| 3.3.3 | 結果 | 37 |
| 3.4 | MEMEによるアレルゲンプロファイルとアレルゲンシーケンスの作成..... | 38 |
| 3.4.1 | プロファイルとシーケンス作成の流れ..... | 38 |
| 3.4.2 | MEMEの実行..... | 38 |
| 3.4.3 | MEMEの実行結果..... | 38 |
| 3.4.4 | シーケンスの抽出..... | 39 |
| 3.4.5 | 考察 - 文献(30)との比較 - | 40 |
| 3.5 | 作成したプロファイルおよびシーケンスの評価 | 41 |
| 3.5.1 | PrecisionとRecallの算出のためのテスト配列セット..... | 41 |
| 3.5.2 | とうもろこしタンパク質で構成したテストセット | 41 |
| 3.5.3 | 方法 | 44 |
| 3.5.4 | 結果 | 45 |
| 3.5.5 | 考察 | 47 |
| 3.6 | 遺伝子組換えタンパク質の相同性解析..... | 48 |
| 3.6.1 | 対象とした遺伝子組換えタンパク質..... | 48 |
| 3.6.2 | 結果 | 50 |
| 3.6.3 | 考察 | 51 |
| 3.7 | アレルゲン相同性検索についてのまとめ..... | 52 |
| 4 | 参考文献およびサイト一覧..... | 54 |
| | 参考資料..... | 60 |

1

序論

1 序論

1.1 はじめに

現在、様々な遺伝子・タンパク質の配列、特徴ならびにその生物学的な機能が分子生物学的手法により決定されており、アレルゲン特異的 IgE が認識する既知アレルゲン及びその抗原決定基（エピトープ）についても、その知見が数多く蓄積されつつある。

このようなアレルゲンならびにエピトープ情報を利用し、遺伝子産物（タンパク質）について、既知のアレルゲン等との一次配列相同性からそのタンパク質がアレルゲンとして働くか否かを評価するバイオインフォマテックスの手法が、これまでいくつかのグループから提案されてきた。しかしながら、組換えタンパク質のアレルゲン性の予測については、具体的にアミノ酸配列が何個以上相同でなければならぬかが、今のところ明確にされていない。ある種のアレルゲンにおいては、4～6 アミノ酸残基でエピトープを形成している場合もあり、現在標準となっている連続 8 アミノ酸残基の相同性検索を行った場合、アレルゲンの可能性がある組換えタンパク質が偽陰性となる可能性があることが指摘されている。

本調査は、調査仕様書に基づき、タンパク質アレルギー誘発性（アレルゲン性）評価の項目の一つであるアミノ酸相同性検索手法に関する調査を行い、今後の遺伝子組換え食品のリスク評価の取組に資することを目的とする。第 2 章に記したように、アレルゲンデータソースならびに文献を調査し、得られたアレルゲンならびにエピトープ情報をまとめ、また、第 3 章に記したように、これらの情報をベースとして、更にアミノ酸相同性検索手法を考察した。

CTC ラボラトリーシステムズ株式会社

1.2 本調査の目的

遺伝子組換え食品の安全性評価基準における、タンパク質のアレルゲン性評価の項目の一つであるアミノ酸の相同性検索手法に関して、既知アレルゲンのエピトープ、3次元データ等に係る国内外の最新の文献情報ならびに学術発表情報等を幅広く収集・整理するとともに、最新のバイオインフォマティクス手法を用いてタンパク質のアレルゲン性を予測・解析し、今後の遺伝子組換え食品のリスク評価手法の改善に資することを目的とする。

1.3 本調査の概要

1.3.1 既知アレルゲン配列および構造に関する調査

多数のタンパク質、アレルゲン及びエピトープ関係の科学技術文献、データベースを調査し、既知アレルゲンタンパク質一次配列、エピトープ、アレルゲンタンパク質三次構造データ、文献をそれぞれ関連付け、アレルゲン配列データベースとして整理した。

1.3.2 アレルゲン性予測手法に関する調査

既知アミノ酸配列相同性検索技術を調査し、その調査から得られた複数のアミノ酸配列相同性検索手法を実行し、各種タンパク質に関するアレルゲン性予測結果を比較・分析した。この場合のアミノ酸配列相同性検索に利用する参照・比較データベースは1.3.1で得られたアレルゲン配列を利用した。

1.4 本調査の実施期間

自 2004 年 1 月 27 日 至 2004 年 3 月 25 日

1.5 本調査の実施体制

- ・ 本調査の実施体制を以下に示す。CTC ラボラトリーシステムズ側の各担当者役割と担当業務内容は、次表のとおり。

CTC ラボラトリーシステムズ側調査担当者一覧

| 番号 | 役割 | 担当業務 | 備考 |
|----|---------------|--|--------------------------------|
| 1 | エキスパート | <ul style="list-style-type: none">・ 本調査の CTCLS 側の承認を行う・ 本調査において各調査結果のレビューを行う | |
| 2 | プロジェクトマネージャー | <ul style="list-style-type: none">・ 契約・文献調達などの事務手続きの取り纏めを行う・ 本調査の進捗管理を行う | |
| 3 | シニアコンサルタント | <ul style="list-style-type: none">・ 本調査の実施計画を立案する・ 本調査を実施する | 分子生物学ならびにバイオインフォマテックスの知識を保有する。 |
| 4 | アプリケーションエンジニア | <ul style="list-style-type: none">・ 本調査に必要なソフトウェアならびにハードウェアおよび文献に記載の相同性検索プログラムを維持管理する・ 本調査を実施する | 情報学ならびにバイオインフォマテックスの知識を保有する |
| 5 | アシスタント | <ul style="list-style-type: none">・ 本調査実施によって得られた結果を記録し、報告としてまとめる | |

2

既知アレルギー配列および 構造に関する調査報告

2 既知アレルゲン配列および構造に関する調査報告

2.1 調査の目的と背景

今日の分子生物学的な研究では容易にアクセスできるデータソースと解析のための洗練されたツールが要求されており、タンパク質アレルゲン性評価の研究でも同様である。これら、データベースならびにデータソースの包括的なリストが分子生物学データベースコレクション(1)にあるが、アレルゲン関連データは様々な公的分子生物学データベースに分散しているようである。

PIR(2)やSWISS-PROT(3)といったような、よく利用されるタンパク質配列データベースは数多くの既知のアミノ酸配列が登録されているが、必ずしも頻繁に更新されているわけではない。さらに、これらのデータベースは一般的に特定のアレルゲンに対して詳しい構造情報や機能情報を持っているわけではなく、あくまでもタンパク質データベースとして統一的なフォーマットと生物学的な情報を提供しているだけである。そのような事情から、特定のアレルギーに関連したデータベース及びそのデータレポジトリの数がここ数年で増加している傾向にある。

本調査では、はじめに、インターネット上で公開されているアレルゲンデータベース、もしくは文献で公表されているアレルゲンリストから、それらのデータソースを使い統合的なデータベースが作成できる可能性を念頭におき、それらのアレルゲンが持っているエピトープ情報を関連文献から探し出し、整理することを目的とした。

2.2 利用可能なアレルゲンデータベース調査結果

バイオインフォマテックスを利用したアレルゲン性解析のためには、既知アレルゲン配列ならびに構造情報を持ち、かつ信頼性のあるデータソースにアクセスする必要がある。いくつかのアレルゲンに特化したデータベースならびに、ほとんどのアレルゲンのリストはインターネット上でアレルゲン名、注釈、アクセッション番号、既知配列データベースへのリンクなどの情報を持ち、HTML上のフォーマットで公開されている。以下に、アレルゲンデータベース調査結果として現在利用可能なアレルゲンデータベースを示す。

(1) 【 IUIS 】

このデータベースには臨床に関連したアレルゲンとイソアレルゲン

のリストが記載されており、アレルゲンの名称は、IUIS(International Union of Immunological Societies) のアレルゲンの学術命名規約ガイドライン(4)によって維持管理されている(14)。

リスト中のアレルゲンは、IgE 反応が罹患率 5%以上であるか、もしくは、5 人以上の被験者で IgE 反応があったアレルゲンを示している。1 年に 2、3 回アップデートし、1 回のアップデートによって約 1-20 の新規アレルゲンが追加されているようである。現在のアップデート(2004 Feb)は 150 以上の生物種からの 445 アレルゲンから構成される(この中にイソアレルゲンは含まれていない)。

アレルゲンリストには生物種、アレルゲン名(命名法)、タンパク質名(生化学的名称)、分子量、cDNA がわかっているのかペプチドがわかっているのかを示す記号、SWISS-PROT または GenBank のアクセッション番号、およびアクセッション番号が決まっていなアレルゲンに関しては引用文献が記載されている。また、このサイトには 20 以上の生物種から 219 のイソアレルゲンが登録されている。(表 1)

(2) 【 SWISS-PROT index of allergen sequences 】

256 アレルゲンと 304 の SWISS-PROT エントリーから構成されているアレルゲンリストである。アレルゲン名、SWISS-PROT のアクセッション番号ならびに SWISS-PROT へのリンクがリストされている。(表 1)

(3) 【 IFBC/ILSI 】

1996 年に Metcalfe ら(5)がアレルゲン性評価のための決定樹アプローチを研究開発した時に利用したアレルゲンリストである。

当時、国際食品バイオテクノロジー諮問委員会と国際生命科学協会アレルギー免疫研究所(The International Food Biotechnology Council (IFBC) and International Life Sciences Institute (ILSI) Allergy and Immunology Institute)が、食物組換えタンパク質のアレルゲン性評価のために、決定樹アプローチを支援したのが背景である。

この研究では、当時公共データベースに登録されていたタンパク質アレルゲン配列のリストを作成したようである。しかし、作成されたアレルゲンのリストは文献による印刷物だけであり、文献が公表されて以来更新されていない。

(4) 【BIFS Tables】

The Biotechnology Information for Food Safety (BIFS)のウェブページ(6)は食品の中に含まれているアレルゲンタンパク質を配列相同性比較によって評価するために、重複のないアレルゲンリストを提供している。

このBIFSのウェブページでは3つのテーブルを提供している。1番目のテーブルは食品アレルゲン、2番目のテーブルは食品以外のアレルゲン、3番目のテーブルは小麦タンパク質のアレルゲンとなっている。これらのテーブルは生物種、食品名、タンパク質名、アレルゲン名、SWISS-PROT アクセション番号、PIR アクセション番号、GenPept アクセション番号、および参照文献が記載されている。

なお、食品アレルゲンのテーブルには2004年3月時点で、289のSWISS-PROTのアクセション番号を持つアレルゲンが登録されている。(表1)

(5) 【CSL Allergen Database】

このサイトは検索可能なアレルゲンデータベースを提供している。データベースサイトはイギリスのSand HuttonにあるFood Safety and Quality Directorate of The Central Science Laboratoryによって開発およびメンテナンスされており、2004年3月現在で540のアレルゲンエントリーと1200以上のアイソフォーム配列にリンクしている。

データベースエントリーは、アレルゲン名(命名法)、生物種、アレルゲンタイプおよび注釈を持っている。CSLのデータベースは比較的よく研究されている20以上のアレルゲンから得られる120以上のエピトープ配列も含んでおり、さらに、アレルゲンのアミノ酸配列とPDBの立体構造データならびにPubMedのリファレンスにリンクしている。

このデータベースはSWISS-PROT, PIR, GenPept, RefSeq, PRF, PDBからのエントリーを選んでおり、重複エントリーや改訂されている複数のエントリーを1つの同じエントリーにまとめている。データベースの検索方法であるが、CSLデータベースはキーワードやアレルゲンタイプ(food allergen, inhalant allergen, contaminant allergen, other allergen)によって検索することができるようになっている。(表1)

(6) 【FARRP】

University of Nebraska-Lincolnに事務局のあるFARRP(The Food

Allergy Research and Resource Program)が管理・運営するアレルギーデータベース。

FARRP は、GenBank、EMBL、PIR、PDB、および SWISS-PROT から抽出したユニークな 650 以上のアレルギー (およそ 658 のアレルギー) を含んでおり、検索可能なデータベースシステムとなっている。このデータベースの開発とメンテナンスは、これまでモンサント社のグループによって行われてきたようである。

データベースの特徴として、文献にアミノ酸配列が記述されているが、公共の配列データベースには登録されていないようなアレルギーもこのデータベースの中に登録されている。

データベースは生物種、アレルギー名 (命名法)、およびアレルギーを抽出したソースデータベースのアクセッション番号を含むリストを見ることができるようになっている。また、これらデータベースに登録されているアレルギー配列に対して FASTA (7) を使った相同性検索が可能となっている。(表 1)

(7) 【 PROTALL 】

PROTALL データベースは Institute of Food Research of the United Kingdom に設置されており、もともと、食品アレルギーを研究するために臨床の専門家、食品科学者、植物学者への情報提供を目的としたサイトであった。

データエントリーは植物性食品タンパク質が持つアレルギーとしての可能性を示すデータ、ならびにそれらの分子構造と生物活性を含んでいる。このデータベースは、48 の植物種から得られる 77 のアレルギーを使い、免疫グロブリン E により誘導される過敏反応を調べることにより得られた植物性食品アレルギーと、そのアレルギーに関する生化学的かつ臨床に関する情報を含んでいる。

PROTALL のエントリーは、1 番目に一般的なアレルギー情報として、一般名、アレルギー名 (命名法)、公共データベースのアクセッション番号とリンク、ならびに構造情報を含んでいる。2 番目にアレルギーの物性として、分子量、エピトープ、安定性、交差反応性、タンパク質機能、単離精製情報、ならびに参照文献を含んでいる。3 番目に臨床データとして、経口誘発試験情報、プリックテスト、免疫グロブリンアッセイ、ならびに病歴などを含んでいる。

PROTALL のアレルギーデータとして登録するかどうかの判断基準は、免疫アッセイや免疫プロットによる免疫グロブリン結合の測定結果を

踏まえているようである。(表 1)

(8) 【ALLALLERGY】

このデータベースは、5200 以上のエントリーを含んでいるデータベースである(2004 年 3 月現在で 5227 エントリー)。

アレルゲン分類の名称(例えば、cereals, fish, fruit, pollens など)によって検索することができ、その物質の機能や副作用を結果として返してくる。

タンパク質しか含まれていない IUIS や PROTALL データベースと比較して非常に多くのエントリーを保有しているように見える理由は、化学物質アレルゲンも含んでいるためである。

ユーザーは、アレルゲン、副作用、交差反応、分類学的関連、患者の評価に有用な情報など包括的な情報をこのデータベースから得ることができる。ALLALLERGY データベースに含まれる情報は全て参考文献が付けられており、さらに定期的にデータ更新が行われている。

ALLALLERGY データベースの要約バージョンには Allergy Advisor™ ソフトウェア(Zing Solutions, Cape Town, South Africa)からアクセスすることができ、食品ならびに植物のアレルゲン、添加物、防腐剤、天然物に関する情報を得ることができる。(表 1)

(9) 【Asthma and allergy gene database】

このデータベース(8)はドイツ、ミュンヘンの National Research Centre for Environment and Health で開発されたようである。

喘息とアレルギー体質に対するリンケージ解析と変異解析のデータにアクセスすることができ、データベースは 19000 以上のリンケージ解析エントリーと 480 以上の変異解析エントリーを持っている。

この asthma and allergy データベースは染色体番号、試験タイプ、結果タイプ、統計的有意性によって検索することができるデータベースシステムとなっている。

データベースから得られるそれぞれのリンケージ解析のエントリーは、染色体上の位置、ヒトゲノムデータベースにリンクできるマーカー、*P*-value、表現型、ならびに試験の特徴に関するデータを持っている。また、それぞれの変異解析エントリーは遺伝子名、変異タイプ、相関タイプ、表現型ならびに、試験の特徴(試験に利用した集団など)に関するデータを持っている。(表 1)

(10) 【 SDAP (Structural Database of Allergenic Proteins) 】

このアレルギーデータベースは University of Texas Medical Branch で開発・管理されており、エピトープ配列情報とタンパク質アレルギーの立体構造情報を持つ包括的なデータベースである (9)。

このデータベースは SWISS-PROT、TrEMBL、PIR、NCBI Taxonomy Browser ならびに PDB からのエントリーを選んでおり、アレルギー名は IUIS を参照している。

データベースの検索方法であるが、SDAP データベースは、IUIS に従ったアレルギー名、生物種、注釈、アレルギータイプ、およびアミノ酸配列によって検索することができるようになっている。

データベースの検索結果からはアレルギータイプ、配列アクセッション番号、PDB アクセッション番号の他、エピトープ配列情報を得ることができ、参照文献からエピトープ配列に関しては文献から収集されていることがわかる。

アレルギー立体構造とエピトープ配列に関して参照できるようなデータベースはそれまでなく、エピトープ配列を問い合わせ条件として検索できる初の包括的なアレルギーデータベースと言える。(表1)

(11) 【 その他の data sources 】

その他、様々なアレルギーデータにアクセスすることのできるインターネット上のデータソースが存在している。吸入(空中)アレルギー(10)、ラテックスアレルギー(11)、ヨーロッパの花粉情報のウェブサイト(12)、European aeroallergen network pollen database (13)、やアレルギー用語集 (14)などが特徴的なデータベースとして広く知られている。

表 1 利用可能なアレルゲンデータベースとそのデータベースのバージョン

| Database | URL (http://) | バージョン または リリース日 |
|----------------------------|---|------------------------|
| IUIS | http://www.allergen.org | Mar 2004 |
| SWISS-PROT アレル ゲン Index | http://www.expasy.org/cgi-bin/lists?allergen.txt | Feb 2004 |
| BIFS | http://www.iit.edu/~sgendel/fa.htm | Release 4, 2003 |
| CSL | http://www.csl.gov.uk/allergen | 2003 |
| FARRP | http://www.allergenonline.com | Ver. 4.0 April 2003 |
| PROTALL | http://www.ifr.bbsrc.ac.uk/Protall | June 2001 |
| ALLALLERGY | http://www.allallergy.net | 2003 |
| Asthma & Allergy | http://cooke.gsf.de/asthmagen/main.cfm | December 2002 |
| SDAP | http://fermi.utmb.edu/SDAP/index.html | Ver. 1.0 Oct 2002 |

2.3 既存データベースにおけるアレルゲンデータエントリー調査結果

先に報告した通り、特定のアレルギーに関連したデータベースやデータレポジトリの数がここ数年で増加している傾向にあり、さらに、その特徴が決定されているタンパク質アレルゲンの数は着々と増え続けていると考えられる。

特徴が決定されているアレルゲンデータの数を明らかにするために、現在の各アレルゲンデータベースのエントリーに関して、アミノ酸一次配列、立体構造情報やエピトープ配列がどの程度わかっているのかを調査・整理したので以下に報告する。

2.3.1 アレルゲンと関連する公共配列アクセッション番号の数

以下表 2 には、Brusic らの文献 (15) による 2001 年 9 月時点での各アレルゲンデータベースが持つ配列データアクセッション番号の数を示す。これは、各アレルゲンデータベースが持つアレルゲン数と、登録されているアレルゲンが持つアクセッション番号の総数(公共の配列データベースへのリンク総数)を示している。

表 2 2001 年 9 月時点での各アレルゲンデータソースにおけるリンク、登録されているアクセッション番号の数 (1 アレルゲンに対して複数のアクセッション番号を許している)

| | IUIS | SPindex | BIFS | Total |
|------------------------|------|---------|------|-------|
| SWISS-PROT | 42 | 266 | 285 | 593 |
| PIR | 4 | | 344 | 348 |
| GenPept | 285 | | 375 | 660 |
| GenBank only | 8 | | | 8 |
| Total valid Accessions | 339 | 266 | 1004 | 1609 |
| Total Allergen | 589 | 266 | 559 | |

以下表 3 には、今回の調査による 2004 年 3 月時点での各アレルゲンデータベースが持つ配列データアクセッション番号の数を示す。これは、各アレルゲンデータベースが持つアレルゲン数と、登録されているアレルゲンが持つアクセッション番号の総数(公共の配列データベースへのリンク総数)を示している。

表 3 2004 年 3 月時点での各アレルギーデータソースにおけるリンク、登録されているアクセッション番号の数（1 アレルゲンに対して複数のアクセッション番号を許している）

| | IUIS | SPindex | BIFS (Food) | BIFS (Non-Food) | Total |
|---------------------------|------|---------|------------------|----------------------|-------|
| SWISS-PROT | 454 | 304 | 289 | 437 | 1030> |
| PIR | | | 235 | 193 | 428> |
| GenPept | | | 370 | 460 | 830> |
| GenBank only | | | | | |
| Total valid accessions | 454 | 304 | 894 | 1090 | 2742 |
| Total Allergen | 664 | 304 | 456 | 644 | |

（ IUIS が持つ 454 のアクセッション番号の内訳に関しては、IUIS データベース内の Allergen リストに含まれる 248 のアクセッション番号と Iso アレルゲンに含まれる 206 のアクセッション番号の総和となっている。 ）

表 2（2001 年 9 月）と表 3（2004 年 3 月）のアクセッション番号数を比較すると、Swiss-Prot と GenPept には同じぐらいの数のアレルゲンが登録されていることがうかがえる。また、トータルアレルゲンの数で比較すると、SPindex に関しては、約 15% 程度の増加であり、それは IUIS においても同じく約 13% 程度の増加であることがわかる。

Brusic らの文献（15）によると、BIFS データベース自身が報告しているデータエントリー数は約 557 エントリーとなっている。しかしながら、今回の調査では BIFS データベースの食品アレルゲンテーブルと食品以外のアレルゲンテーブルから多数のアレルゲンが得られており、食品以外のアレルゲンデータの増加がその理由ではないかと考えられる。

2.3.2 SDAP に登録されているアレルゲンのエピトープ数

先の「利用可能なアレルゲンデータソース調査報告」で述べたとおり、アレルゲンに対するエピトープ情報を持つデータベースは多くない。エピトープ情報を比較的多く含み最も利用しやすい包括的データベースは SDAP であろうと思われるため、SDAP から既知エピトープ情報を調査した。

SDAP に登録されているアレルゲンは合計 736 であり、その中でエピトープ情報の掲載されているアレルゲンの数は 19 であった。

以下表 4 には、SDAP データベースにおいて各アレルゲンが持つエピトープ数を示す。

SDAP データベースにおける各アレルゲンが持つエピトープ配列は本調査報告書の参考資料 1 を参考にして頂きたい。

表 4 SDAP データベースに記載されているエピトープ数（エピトープが記載されているアレルゲンだけを抜粋）

| | Allergen - Scientific Name | Species - Scientific Name | Species - Common Name | Allergen Type | Allergen Description | Number of Epitope | Link to Source | | |
|---|----------------------------|---------------------------|-----------------------|---|------------------------|-------------------|------------------|------------------|-----------|
| | | | | | | | GenBank GI | SwissProt ACC | PubMed ID |
| 1 | Alt a 1 | Alternaria alternate | | fungi (moulds) Ascomycota Dothidiales | | 2 | | P79085 | |
| 2 | Ara h 1 | Arachis hypogaea | peanut | foods | vicilin | 21 | | P43237 P43238 | |
| 3 | Ara h 2 | Arachis hypogaea | peanut | foods | conglutin | 10 | 15418705 | | 9186485 |
| 4 | Ara h 3 | Arachis hypogaea | peanut | foods | glycinin | 4 | 3703107 | 082580 | |
| 5 | Asp f 1 | Aspergillus fumigatus | | fungi (moulds) Ascomycota Eurotiales | | 13 | 166486 250902 | P04389 | |
| 6 | Asp f 12 | Aspergillus fumigatus | | fungi (moulds) Ascomycota | heat shock protein P90 | 1 | | P40292 | |

| | Allergen - Scientific Name | Species - Scientific Name | Species - Common Name | Allergen Type | Allergen Description | Number of Epitope | Link to Source | | |
|----|----------------------------|---------------------------|-----------------------|--|---|-------------------|----------------|----------------------------|-----------|
| | | | | | | | GenBank GI | SwissProt ACC | PubMed ID |
| | | | | Eurotiales | | | | | |
| 7 | Asp f 13 | Aspergillus fumigatus | | fungi (moulds) Ascomycota Eurotiales | alkaline serine protease | 5 | | P28296 | |
| 8 | Asp f 3 | Aspergillus fumigatus | | fungi (moulds) Ascomycota Eurotiales | peroxisomal protein | 7 | 664852 | 043099 | |
| 9 | Cry j 2 | Cryptomeria japonica | sugi | tree Pinales | | 4 | 506858 | P43212 | |
| 10 | Gal d 1 | Gallus domesticus | chicken | foods | ovomuroid | 17 | 212488 | P01005 | |
| 11 | Hev b 1 | Hevea brasiliensis | rubber (latex) | others | elongation factor | 8 | 18839 | P15252 | |
| 12 | Hev b 3 | Hevea brasiliensis | rubber (latex) | others | | 11 | | 082803 | |
| 13 | Hev b 5 | Hevea brasiliensis | rubber (latex) | others | | 11 | 1480457 | Q39967 | |
| 14 | Jug r 1 | Juglans regia | English walnut | foods | 2S albumin | 1 | 1794252 | | 11799381 |
| 15 | Jun a 3 | Juniperus ashei | mountain cedar | tree Pinales | homolog: pathogenesis related protein PR5 | 5 | | P81295 | |
| 16 | Par j 1 | Parietaria judaica | Spreading pellitory | weed Rosales | lipid transfer protein; homolog: pathogenesis | 5 | | P43217 004404 Q40905 | |

| | Allergen - Scientific Name | Species - Scientific Name | Species - Common Name | Allergen Type | Allergen Description | Number of Epitope | Link to Source | | |
|----|----------------------------------|--|-----------------------------|---|--------------------------------|----------------------|----------------|------------------|--------------|
| | | | | | | | GenBank GI | SwissProt ACC | PubMed ID |
| | | | | | s related protein PR14 | | | | |
| 17 | Par j 2 | Parietaria judaica | Spreading pellitory | weed Rosales | lipid transfer protein | 8 | 3914131 | P55958 | |
| 18 | Pen a 1 | Penaeus aztecus | shrimp | foods | tropomyosin | 9 | | | 11893851 |
| 19 | Pen ch 18 | Penicillium chrysogenum (formerly P. notatum) | | fungi (moulds) Ascomycota Eurotiales | vacuolar serine protease | 9 | 7963902 | | |

2.3.3 立体構造情報を持つアレルゲン数

アレルゲンの立体構造情報に関する調査は、Swiss-Prot、PDB(16)、HSSP(17)を利用した。Swiss-Protはタンパク質配列データベースであり、PDBはX線結晶構造解析やNMR構造解析によるタンパク質3次元座標のデータベースである。HSSP(homology-derived structures of proteins)は、類似しているタンパク質配列をそれぞれグループに分けることにより、タンパク質の三次構造とタンパク質の一次構造を構造的に相同なファミリーに分類したデータベースである。HSSPの中にあるそれぞれのファミリーは既知タンパク質三次構造が含まれており、各ファミリーの中心となっている(18)。

どのアレルゲンタンパク質配列の立体構造が決定されているのかという直接的な情報は、PDBを参照することによって得られる。一方、まだ立体構造の決定されていないアレルゲンタンパク質がどのような立体構造をとりうる可能性があるかという情報は、HSSPを参照することによって得られることになる。

以下、表5には、立体構造情報を持つアレルゲンの一部リストを示している。立体構造情報を持つアレルゲン159個の全リストは本調査報告書の参考資料2を参考にして頂きたい。

表 5 立体構造情報を持っているアレルゲンリスト（一部を記載）

| | Allergen - Scientific Name | SwissProt ID (ACC) | PDB | HSSP |
|----|----------------------------|--|------------------------------|------------------------------|
| 1 | Aca s 13 | FABP_ACASI (076821) | | P04117 (1LID) |
| 2 | Act c 1 | ACTN_ACTCH (P00785) | 2ACT, 1AEC, 1F3Q, 1FOG, 1FTV | |
| 3 | Aln g 1 | MPAG_ALNGL (P38948) | | P15494 (1BV1) |
| 4 | Aln g 4 | POC4_ALNGL (081701) | | P02593 (1AK8) |
| 5 | Alt a 10 | DHAL_ALTAL (P42041) | | P05091 (1CW3) |
| 6 | Alt a 11 | ENO_ALTAL (Q9HDT3) | | P00924 (1ONE) |
| 7 | Amb a 3 | MPA3_AMBEL (P00304) | | P00303 (2CBP) |
| 8 | Amb a 5 | MPA5_AMBEL (P02878) | | P10414 (1BBG) |
| 9 | Amb a 6 | NLT6_AMBAR (004004) | | P19656 (1MZM) |
| 10 | Amb p 5 | MP5A_AMBPS (P43174) MP5B_AMBPS (P43175) | | P10414 (1BBG), P10414 (1BBG) |
| 11 | Amb t 5 | MPT5_AMBTR (P10414) | 1BBG, 2BBG, 3BBG | |
| 12 | Api g 1 | ALL1_APIGR (P49372) | | P15494 (1BV1) |
| 13 | Api g 2 | ALL2_APIGR (P92918) | | Q24248 (1E09) |
| 14 | Api g 3 | CB23_APIGR (P92919) | | P04083 (1B09) |
| 15 | Api g 4 | PROF_APIGR (Q9XF37) | | P25816 (1CQA) |
| 16 | Api m 1 | PA2_APIIME (P00630) | 1POC | |
| 17 | Api m 2 | HUGA_APIIME (Q08169) | 1FCQ, 1FCU, 1FCV | |
| 18 | Api m 3 | MEL_APIIME (P01501) | 2MLT, 1BH1 | |
| 19 | Ara h 1 | AL11_ARAHY (P43237) AL12_ARAHY (P43238) | | P50477 (1CAU), P50477 (1CAU) |
| 20 | Ara h 5 | PROF_ARAHY (Q9SQ19) | | P25816 (1CQA) |

例えば、キーウィフルーツアレルギーの1つである Act c 1 は、タンパク質配列の Swiss-Prot ID は ACTN_ACTCH であるが、解明されている PDB の立体構造エントリーは、複合体やその予測モデルも含め5つも登録されている。

その他、ピーナッツアレルギーの1つである Ara h 5 は、タンパク質配列の Swiss-Prot ID は PROF_ARAHY であるが、関連している PDB の立体構造エントリーがないため、このアレルギータンパク質の立体構造はまだ決定されていないということがわかる。しかし、関連している HSSP のエントリーは、P25816 (1CQA) となっており、Ara h 5 は PDB エントリー 1CQA と構造的に同じファミリー、つまり 1CQA と同様なフォールドをとると類推できる。さらに、HSSP データベースの P25816 (1CQA) を詳しく調べてみると、立体構造の解明されている P25816 (1CQA) と Ara h 5 の間には 75% にわたり一次配列上のアミノ酸残基が同一であることがわかり、同様な立体構造をとる可能性が高いことがわかる。

今回の調査では、PDB または HSSP エントリー情報の得られたアレルギーは 159 あった。また、これらのアレルギーが持つ、Swiss-Prot エントリーは合計 196、PDB エントリーは合計 169、HSSP エントリーは合計 127 であることがわかった。IUIS に登録されているアレルギーの数はアイソフォームも含み 664 であることから計算すると、全体の約 25% 以下のアレルギーで立体構造情報が得られる可能性があると言える。

2.3.4 アレルゲンデータエントリー調査まとめ

今回の調査によると、「認定された」タンパク質アレルゲンリストを維持管理している The Allergen Nomenclature Sub-Committee of the International Union of Immunological Societies (IUIS) (19) は、2004年2月の時点で、445のタンパク質アレルゲンと219のアイソフォーム、計664のアレルゲンを分類登録している。この445という数は、数多くのアイソフォームを持つアレルゲンのほんの一部にしすぎないと考えられる。この認定されたリストは、その特徴が決定されている多くのアレルゲンを含んでいるが、一方、文献中に数多く見つけられるまだその特徴がはっきりとしないアレルゲンなどを網羅しているわけではないと考えられる。

2.4 エピトープ情報に関する文献調査方法

SDAP データベースの調査からわかるように、アレルゲンのエピトープ情報は文献から得るしか方法がない。そこで、本調査では、既存アレルゲンデータベースが持たない新たなエピトープ情報を付け加えるためにエピトープ情報の掲載されている既存文献を探した。

文献のソースは MedLine (20) を使い、MedLine アブストラクトをキーワード検索することによる文献抽出を行った。

これまで SDAP データベースなどで参照されているエピトープ情報の掲載されている文献 (21~28) の多くは、次のキーワードを含む傾向にある。

IgE-Binding

Epitope

Identification

Immunoglobulin E

Epitope Mapping

Sequence

Analysis

Peptide

Recognition

IgE-epitopes

Linear

本調査では、次に、これら各キーワードの組み合わせにより 5 回にわけて MedLine アブストラクトを検索した。

以下に、本調査で利用した MedLine アブストラクトを検索するためのキーワードならびに、そのキーワードを入力した時に Entrez システムが自動で作成する検索式を示す。

1) **検索キーワード** : IgE-Binding、Epitope、Identification

Entrez 検索式 :

```
((ige-binding[All Fields] AND ("epitopes"[MeSH Terms] OR epitopes[Text Word])) AND ("identification (psychology)"[MeSH Terms] OR identification[Text Word]))
```

2) **検索キーワード** : Immunoglobulin E、Epitope Mapping、Sequence

Entrez 検索式：

((("immunoglobulin e"[MeSH Terms] OR immunoglobulin E[Text Word]) AND ("epitope mapping"[MeSH Terms] OR epitope mapping[Text Word])) AND sequence[All Fields])

3)検索キーワード：Immunoglobulin E、Epitope、Sequence、Analysis、Peptide

Entrez 検索式：

(((((("immunoglobulin e"[MeSH Terms] OR immunoglobulin E[Text Word]) AND ("epitopes"[MeSH Terms] OR epitope[Text Word])) AND ("analysis"[MeSH Subheading] OR analysis[Text Word])) AND ("peptides"[MeSH Terms] OR peptide[Text Word])) AND sequence[All Fields])

4) 検索キーワード：IgE-Binding 、 Epitope、 Recognition

Entrez 検索式：

((IgE-binding[All Fields] AND ("epitopes"[MeSH Terms] OR epitope[Text Word])) AND ("recognition (psychology)"[MeSH Terms] OR recognition[Text Word]))

5) 検索キーワード：IgE-Epitope 、 Linear

Entrez 検索式：

(IgE-epitopes[All Fields] AND linear[All Fields])

上記各検索から多数の文献を結果として得たが、下に示す3つを選択のクライテリアとして使い、エピトープ情報の掲載されている文献をさらに精査した。

各検索条件でのリファレンス出現頻度

アブストラクトの内容（エピトープ情報が記載されている内容）

SDAP データベースが作成された 2002 年以降の文献

上記3点を考慮した結果、合計で21報の文献を得ることができた。参考資料5にその21報の文献リストを示す（PMIDはPubMedのIDを示している）。

2.5 エピトープ情報に関する文献調査報告

選択された 21 報の論文を読むことにより、既知データベースには取り込まれていないエピトープ情報も得ることができた。

以下、表 6 には、今回選択した 21 報の論文に記載されていたエピトープ配列と、そのエピトープ配列を持つアレルゲンの一部を示している。また、表 7 には、同じく 21 報の論文に記載されていたアレルゲンと、そのアレルゲンが持つエピトープの数を一部示している。表 6 および 7 に示したエピトープ情報に関しては SDAP データベースには記載されていない。

参考資料 3 は、今回選択した 21 報の論文に記載されていた全アレルゲンと、記載されていた全エピトープ情報を抽出し、示している。また、参考資料 4 は、今回選択した 21 報の論文に記載されていた全アレルゲンと、記載されていた全エピトープの数を示している。各表に記載されている文献番号は 2.4 で示した文献番号に対応している。

表 6 エピトープ情報を持つアレルゲンリスト（一部を記載）

| Allergen - Scientific Name | Species - Scientific Name | Epitope | Position | Type | Description | References |
|----------------------------|---------------------------|-------------------------------|-----------------------|------|---|------------|
| Hev b 6.02 | <i>Hevea brasiliensis</i> | | 5, 10, 29, 30, 35, 38 | IgE | Arg5, Lys10, Glu29, Try30, His35, Gln38 という hevein の 6 つの残基が, IgE との相互作用に重要である。 | 1 |
| (omega-5 gliadin) | <i>Triticum aestivum</i> | QQIPQQQ (Dominant epitope) | | IgE | QQIPQQQ および QQFPQQQ のアミノ酸を部分的に異なるアミノ酸に置換してエピトープ性を調べた結果, 位置 1(Q),4(P),5(Q),6(Q) および 7(Q)のアミノ酸が IgE に結合するためには重要である。 | 2 |
| | | QQFPQQQ (Dominant epitope) | | IgE | | |
| | | QQSPEQQ (Dominant epitope) | | IgE | | |
| | | QQSPQQQ (Dominant epitope) | | IgE | | |
| | | QQLPQQQ | | IgE | | |
| | | QQYPQQQ | | IgE | | |
| | | PYPP | | IgE | | |
| Pru p 3 | <i>Prunus persica</i> | APCIPYVRG GGAVPP | 11-25 | IgE | identified major IgE epitope | 3 |
| | | IRNVHNLAR TTPDRQ | 31-45 | IgE | identified major IgE epitope | |
| | | GKCGVSIPIY K | 71-80 | IgE | identified major IgE epitope | |

表 7 アレルゲンが持つエピトープの数 (一部を記載)

| Allergen - Scientific Name | Species - Scientific Name | Species - Common Name | Allergen Type | Allergen Description | Number of Epitope | Link to Source | | |
|----------------------------|---------------------------|-----------------------|---------------|------------------------|-------------------|----------------|---------------|-----------|
| | | | | | | GenBank GI | SwissProt ACC | PubMed ID |
| Hev b 6.02 | <i>Hevea brasiliensis</i> | rubber (latex) | others | hevein | | 123062 | P02877 | 14764736 |
| | <i>Triticum aestivum</i> | wheat | foods | omega-5 gliadin | 7 | | | 14699123 |
| Pru p 3 | <i>Prunus persica</i> | Peach | foods | lipid transfer protein | 3 | 3287877 | P81402 | 13679821 |

結果として、参考資料 5 に示した 21 の論文から、17 生物種 (*Hevea brasiliensis*, *Triticum aestivum*, *Prunus persica*, *Glycine max*, *Bos domesticus*, *Penaeus aztecus*, *Blomia tropicalis*, *Gallus domesticus*, *Anacardium occidentale L*, *Helianthus annuus*, *Penicillium notatum*, *Juglans regia*, *Juniperus ashei*, *Penicillium chrysogenum*, *Aspergillus fumigatus*, *Parietaris judaica*, および *Cryptomeria japonica*) の 21 アレルゲン (Hev b 6.02, omega-5 gliadin, Pru p 3, G2a, Bos d 8, Bos d 6, Pen a 1, Blo t 10, Gal d 1, Ana o 2, Hel a 2, Pen n 18, Jug r 1, Jun a 1, Pen ch 18, Asp f 3, Gal d 2, Par j 1, Par j 2, Blo t 11, および Cry j 2) に対して、合計 118 の IgE エピトープ情報を得ることができた。

これらの文献の中では、一次配列上で連続したアミノ酸残基に注目しているリニアエピトープだけでなく、立体構造上の不連続なアミノ酸残基に注目しているコンフォメーションナルエピトープに関して記載されているものもあった。

例えば、ラテックスアレルゲンである Hev b 6 とエピトープに関する論文では、Hev b 6 に関してリニアエピトープに関しては言及しておらず、コンフォメーションナルエピトープにつながる重要なアミノ酸残基に関して述べていた。Hev b 6 に関しては既に立体構造が 1Hev であることがわかっており、これら立体構造情報からエピトープに関する重要な知見が得られる可能性がある。

2.6 既知アレルゲン配列および構造に関する調査のまとめ

今回の調査では、2002 年以降に公表された 21 の論文から 21 のアレルゲンと 118 の IgE エピトープ情報を得ることができた。これら 118 エピトープの内、71 エピトープが、公的アレルゲンデータソースである SDAP にはまだ収録されておらず、さらに今回の調査で得た 21 アレルゲンの内、3 アレルゲンが同じく SDAP には収録されていなかった。これらエピトープならびにアレルゲンに対しては、今後、SDAP などの公的データソースへの登録が待たれる。

また現在、多くのアレルゲンタンパク質の配列および立体構造が決定されてきているのが今回の調査でわかった。これら、アミノ酸配列や構造決定されたアレルゲンタンパク質を分類してみると、比較的少ないファミリーにしか分布していないことから、これらのアレルゲンタンパク質には IgE 結合能に関与する共通の特性があって、アレルギー反応を誘発していることが示唆される。

2.7 既知アレルゲン配列および構造に関する調査からの考察

本報告書で調査したデータソースに対して、アミノ酸残基特性を利用した配列（モチーフ）検索アルゴリズムを利用することにより、研究者の持っているペプチド配列に類似した既知アレルゲンタンパク質、あるいは、類似した既知エピトープ配列を得ることができると考えられる。また、このような比較プログラムとデータソースの組み合わせは、既知アレルゲンに潜在的な交差反応性の有無を判定することができたり、たとえ機能のわからない新規なタンパク質であったとしても既知 IgE エピトープ配列と一致する配列を持っていないかどうかをスクリーニングすることができる 1 種のツールになりえると考えられる。

アレルゲン名や食品名に代表されるようなキーワード検索のみならず、タンパク質配列またはペプチド配列によるデータベース検索は、何らかの近縁関係にあるアレルゲンを特定するのに有効な手段と言えそうである。例えば、あるアレルゲンタンパク質の IgE エピトープが、他のアレルゲンタンパク質に対しても臨床的に観察された交差反応性を認め得るような場合、このような統合データベースに対する検索は、近縁関係にあるアレルゲンを見つけるのに有効な手段であると思われる。

また同時に、既知アレルゲンの中から迅速に交差反応性のあるアレルゲンを見つけ出すことも配列によるデータベース検索で可能と思われる。

例えば、エビのトロポミオシンPen i 1タンパク質の IgEエピトープ配列であるMQQLENDLDQVQESLLKを、データベースに対して、配列相同性検索すると、全く同じエピトープ配列が、別種のエビのMet e 1というアレルギータンパク質内に存在しており、また、ロブスターのPan s 1というタンパク質内にも存在していることがわかるが、この配列相同性検索結果によって、臨床的にもこれらの甲殻類の間に交差反応が見られるという事実と合致する結果が得られることになると想像できる。

これらのことから、配列、構造データベースや文献を基本としたデータだけでなく、そのデータを利用するためのシステムの重要性と必要性も同時に高まることが考えられる。

3

アレルギー性予測手法に関する調査報告

3 アレルゲン性予測手法に関する調査報告

3.1 現在考えられているアレルゲン相同性検索方法の文献調査

ここでは、現在考えられているアレルゲン相同性検索方法についての、簡単な要約と、今回のアレルゲン性予測手法としてどの方法を選定するかについて述べる。

3.1.1 FAO/WHO のガイドラインに沿った方法(29)

(以下本書では FAO/WHO の方法と呼ぶ。)

2001 年 FAO と WHO の合同専門家会議の中で、1996 年に IFBC と ILSI の定めた Decision Tree を改訂する勧告を出した。その会議報告では、遺伝子組み換え食品のアレルゲン性の有無を評価することを目的として、ガイドラインが出された。そのガイドラインでは、対象タンパク質に実施する生物学的テストのほかに、配列を比較する標準的方法が提案された。

その標準的方法とは、80 残基以上のアミノ酸部分で、35%以上の同一性を有するか、または連続するアミノ酸の数が 6~8 ($n=6\sim 8$) 以上が既知アレルゲン配列と一致していれば、問い合わせタンパク質がアレルゲン性を持つ可能性があるるとみなす方法である (29)。しかし、その後、2003 年の Codex 総会で採択された遺伝子組換え食品の安全性評価のガイドラインの中では、連続アミノ酸検査の規模は、偽陰性または偽陽性結果が生じる可能性を最低限に抑えるために科学的正当性に基づくべきであるとされている(36)。

3.1.2 Two-track approach(32)

(以下本書では Two-track 法と呼ぶ)

まず、アレルゲンと認められるエントリー(文献 32 内では Entrez でキーワードとして、"アレルゲン"を含むもの)に対して BLAST(33)を実行し、連続する 6 残基以上が一致しているアミノ酸配列を選び出す。IgE と結合するエピトープ配列かどうかを予測するため、既知の IgE エピトープ配列に関する文献データと、Hopp and Woods の親水性プロットによる抗原性予測演算法が用いられる。

3.1.3 Motif-based prediction(30)

(以下本書では Motif-based prediction と呼ぶ。)

Motif-based prediction とは、MEME というツールにより既知アレルゲンのアミノ酸配列を解析し、多くのアレルゲンに共通した配列

(モチーフ)をアレルゲンプロファイルとして抽出し、アレルゲン性未知タンパク質のアミノ酸シーケンスをこのプロファイルと比較することによりアレルゲン性を予測しようという手法である。シーケンス同士を比較するのではなく、シーケンスをモチーフと比較することから、Motif-based prediction とよばれる。なお、モチーフの抽出に漏れた既知アレルゲンは、そのアミノ酸配列がアレルゲンシーケンスとして別に抽出され、プロファイルとの比較で漏れた問い合わせ配列のシーケンスベースな比較に用いられる。最終的に、アレルゲン性未知の問い合わせタンパク質は、アレルゲンプロファイルおよびアレルゲンシーケンスのいずれかにマッチすればアレルゲン性を持つ可能性があるとして推定される。

3.1.4 各相同性検索方法の比較とサンプル計算に利用する検索方法の選定

Stadler らの報告(30)によると、FAO/WHO の方法、Two-track 法の、どちらの予測方法でも、どちらも見かけ上は非常に感度の良い予測であるように見える。しかしながら、実際の所は、文献(30)で述べられているモチーフ-ベースによる予測精度(95.5%)が、2001年FAO/WHO ガイドラインに沿った予測精度(36.6%)をはるかに上回っている。

Stadler らの報告(30)を参考にして、本書ではアレルゲン性予測手法として、Motif-based prediction に改変を加えた方法で行うことにした。

Stadler らの方法はモチーフに対する検索方法として、pftools を利用しているが、本調査では、MEME(34)と同様のサイトから配布されている MAST(34)を利用することとした。詳細な方法は、3.2~3.5で述べる。

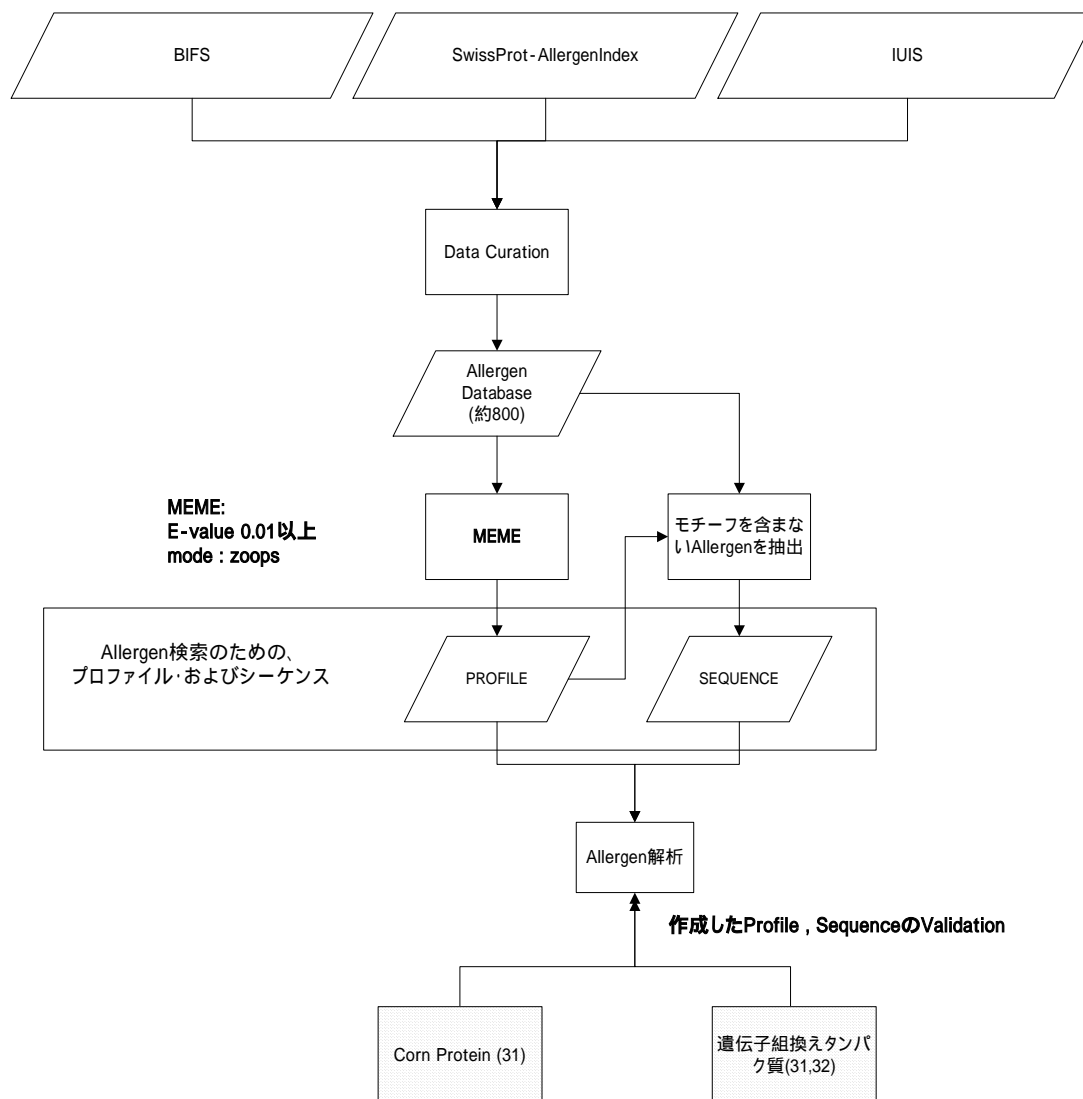
尚、今回行った調査方法を、本報告書では、「本調査におけるアレルゲン性予測」と呼ぶ。

3.2 本調査におけるアレルゲン予測によるサンプル計算

3.2.1 サンプル計算の全体の流れ

今回のサンプル計算の Protocol を図 1 に示す。

図 1 【 アレルゲンデータベース、プロフィール、シーケンス作成の流れ 】



本調査では、Stadler らの Motif-based Prediction を参考にして、アレルゲン予測を行った。

BIFS、SwissProt-AllergenIndex、IUIS をデータソースとして、重複のないデータベースを作成し、それらを実験データベースとした。その実験データベースから、MEME を用いてモチーフを抽出し、さらにどのモチーフにもマッチしないアレルゲンのエントリ

ーをアレルゲンシーケンスとした。アレルゲン予測の判定方法は、MASTによるモチーフのマッチングまたはBLASTによるシーケンスの相同性検索において、有意な結果が得られたものをアレルゲンと予測するというものである。詳細なMASTおよびBLASTのE-valueについては、3.5 作成したプロファイルおよびシーケンスの評価 の章で述べる。

3.2.2 計算機サーバーとアプリケーションの用意

(1)サーバー

以下のサーバーを相同性検索サーバーとして使用した。

表 8 【計算機サーバーのスペック】

| | |
|------|-----------------|
| メーカー | SGI |
| 型 式 | Origin2100 |
| メモリ | 4GB |
| CPU | Mips 250MHz x 4 |
| ディスク | 2TB |

(2)アプリケーション

- MEME3.0.4

入手先：

<http://meme.sdsc.edu/meme/>

アプリケーション概要：

核酸配列やアミノ酸配列から、モチーフ(= 保存された領域)を抽出するプログラム。SDSC(San Diego Super Computer Center) より 1995年にリリースされた。

- MAST3.0.4

入手先：

<http://meme.sdsc.edu/meme/>

アプリケーション概要：

MEMEで抽出したモチーフに対して、モチーフ検索をするプログラム。MEMEの出力結果をそのまま利用できる。

- blast2.2.8 (<http://www.ncbi.nih.gov/BLAST/>)

入手先：

<http://www.ncbi.nih.gov/BLAST/>

アプリケーション概要：

NCBI が作成した 相同性検索プログラム。

・ SRS (<http://www.lionbioscience.com/>)

入手先 : CTCLS の提携先である独 LION Bioscience 社より

アプリケーション概要

データベースインテグレーションツール。各データベースの ID からシーケンスを取得するために利用している。

3.3 アレルゲンデータベースの作成

3.3.1 データソース

今回アレルゲンデータのソースとして利用したものは表 1 内の SwissProt-AllergenIndex・IUIS・BIFS の 3 つである。

3.3.2 方法

3.3.1 のデータソースから同じアレルゲンを指していると思われるものを削除する必要がある。今回は、SwissProt、TrEMBL、PIR、Genbank/Genpept の ID を識別のための ID として利用した。エントリーとして、相補的な cDNA のエントリーが記載されているものに関しては、対応するアミノ酸配列を TrEMBL から取得した。また、複数のデータベースからのエントリーが取得可能なエントリーに関しては、SwissProt、Pir、TrEMBL、Genbank/Genpept の順に優先順位をつけた。これらのエントリーおよびそのシーケンスの取得には、SRS を利用した。

3.3.3 結果

抽出されたエントリーをデータベースごとに集計した。結果を表 9 に示す。なお、アレルゲンデータベースにおいて、ホモロガスな SwissProt の ID が複数記載されているものがあるが、それらは別々に登録した。また、シーケンスの長さが 8 以下のエントリーに関しては、次のプロファイル 作成に利用する MEME にて利用できないため、エントリーから削除した。

表 9 【 データベースとエントリー数 】

| データベース名 | エントリー数 |
|-------------------|--------|
| SwissProt | 509 |
| Pir | 71 |
| TrEMBL | 121 |
| Genbank / GenPept | 2 |
| 合計 | 703 |

(同じアレルゲンとの重複を除いた non-redundant なエントリー数を示す)

3.4 MEME によるアレルゲンプロファイルとアレルゲンシーケンスの作成

3.4.1 プロファイルとシーケンス作成の流れ

大まかな流れは図 1 に示した通りである。3.3 で作成したアレルゲンデータベースに対して MEME を実行した。MEME により抽出されたプロファイルに一致しないアレルゲンデータベースのエントリーのシーケンスをアレルゲンシーケンスとした。

3.4.2 MEME の実行

文献(32)の情報を参考にして、以下のパラメータを用いて MEME を実行した。結果を下表に示す。

表 10【パラメータリスト】

| パラメータ内容 | 値 |
|----------------|-------|
| Motif のアミノ酸残基数 | 50 |
| E-value | 0.01 |
| モード | zoops |

3.4.3 MEME の実行結果

MEME の実行におよそ 20 時間を要した。抽出されたプロファイルの数は 100 種類で、その一部を表 11 に示す。

表 11【 MEME を用いて抽出されたアレルゲンモチーフ例 】

| ID | Matching allergens | E-value | Multi Level Consensus Sequence |
|-------|--------------------|-----------|--|
| N0001 | 65 | 1.5e-2281 | YMVIQGEPGA VIRGKKGSGGITVKKTGQALIIGIYDEPMTPGQCNMVVER GA G P V I N VL NK IA |
| N0002 | 51 | 1.1e-201 | TAAAIIGHDGSVWAQSATFPQFKPEEITGIMKDFDEPGSLAPTGLFLGGT S L Q SN L VAA N N T Y H H |
| N0003 | 57 | 1.5e-1926 | TTSVIPAARLFKAFVLDADNLIPKVAPQAIKSVEILEGNGGPGTIKKITF F E PP I G T I S A N I D |
| N0004 | 61 | 6.5e-1238 | RSQVLQQSTYQLLQQCCQQLAQIPEQSRCQAIHNVVHAIILQQQQQQQ P C VML L E W Q CQ E LR I S QHE G P |
| N0005 | 45 | 1.3e-1326 | KVAAPGGGSIKSTSHYHTKGD AEIKEEHVKAGKEKAHGLFKAIEAYLLA LLV SDS TL II KF A NVVN EI A LV S KG TS P |

| ID | Matching allergens | E-value | Multi Level Consensus Sequence |
|-------|--------------------|----------|---|
| N0006 | 41 | 1.1e-953 | SLTLPALRDLDSAQYVVLYKNAIYVPHWNLNANSIIYVTRGEGRVQVVN P KN Q QFQ VS G I EGGLFL Y SK HVVL ANE RA EL G I R MLS R |
| N0007 | 34 | 7.0e-869 | PQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQILQQILQQQLPCKDVVLQQHNIAH QPPPPS P P P LP IP I PS L N RVFLQ CSPVA S I V |
| N0008 | 28 | 8.1e-800 | QPSQQNPQAQGSVQPQQLPQFEEIRNLALQTLPAMCNVYIPPYCSTTIAP Q QV QGTFL H IA L V TSI R T S NV L TIAPVG S |
| N0009 | 34 | 1.5e-751 | NRVFDGELQEGQVLVVPQNFVAEQAQSEGFEYVVFKTHDNAVTSYLAGV VA Y A RR DLFI AG V V KSQEQLL W GINNHRNQI FIKD D I P N DN A PSR N |
| N0010 | 20 | 5.3e-733 | HNDFRQKVARGLETRGNPQPPAKNMNNLVWDELAKIAQVWANQCQYG E I Q A S K Y S NFL R K |

3.4.4 シーケンスの抽出

抽出されたプロファイルのどれにもマッチしないものをシーケンスとして抽出した。抽出されたアミノ酸配列は213本である。表12に10例を挙げる。

表 12【アレルゲンシーケンス】

| | Accession Number | Description |
|----|------------------|---|
| 1 | swissprot:P04149 | Arachin 25 kDa protein |
| 2 | swissprot:P10846 | Triticum aestivum (Wheat). |
| 3 | swissprot:P80740 | Olea europaea (Common olive). |
| 4 | swissprot:P27042 | Gallus gallus (Chicken). |
| 5 | swissprot:P49277 | Dermatophagoides pteronyssinus (House-dust mite). |
| 6 | swissprot:P35775 | Solenopsis invicta (Red imported fire ant). |
| 7 | swissprot:Q9SDL2 | Apium graveolens (Celery). |
| 8 | pir:A03364 | lectin - peanut (tentative sequence) |
| 9 | Swissprot:P42059 | Cladosporium herbarum. |
| 10 | trembl:Q7XYF3 | Cynodon dactylon (Bermuda grass). |

3.4.5 考察 - 文献(30)との比較 -

Stadler らの報告 (30) に比べ、本調査におけるアレルゲン予測で抽出されたアレルゲンのプロファイルおよびシーケンスの数は、ともに約 2 倍 (それぞれ 100 および 213) にのぼった。これは、Stadler らがモチーフ抽出を行った時期 (2002 年 2 月) より現在に至るまでに、公共データベースが拡充され、用いたデータセットが異なっていること等が寄与したものと思われる。

3.5 作成したプロファイルおよびシーケンスの評価

3.5.1 Precision と Recall の算出のためのテスト配列セット

3.4 で作成したアレルゲンプロファイルとアレルゲンシーケンスのパフォーマンスを調べるため、以下の4種類のデータを用意した。

表9であげている703種のアレルゲンデータのうちシーケンスの長さが26残基以上のもの(682種)

のセットのアミノ酸配列を逆順に並べたもの

のセットを Window Size 20 で Shuffle したもの

Window Size なしで Shuffle したもの

このうち は、既知のアレルゲンシーケンスのうち最も短いものが26残基(Bee venom melittin)であることから、25残基以下のエントリーを除いてある。以上の4つのセット(計2728種)を、を True Allergen、 ~ を True Non-Allergen として本調査におけるアレルゲン性予測を行った。

3.5.2 どうもろこしタンパク質で構成したテストセット

プロファイルとシーケンスの評価のサンプルとして、文献(31)では、どうもろこしのタンパク質からランダムに抽出した50種類のタンパク質を用いて、FAO/WHOの方法の妥当性を検討している。そこで本報告においても、同様のどうもろこしタンパク質テストセットを用い、本調査におけるアレルゲン性予測手法を適用した。用いた50種類のタンパク質を表13に示す。

表13 【どうもろこしタンパク質】

| | Corn protein | Description | Allergen |
|---|--------------|------------------------|----------|
| 1 | A55092 | catalase | |
| 2 | AAA33523 | 16 kDa zein | |
| 3 | AAA68209 | Sus1 gene product | |
| 4 | AAA73960 | kaurene synthase | |
| 5 | AAA87580 | GAPDH | |
| 6 | AAB71078 | ribosomal protein P3a | |
| 7 | AAB86960 | profilin | |
| 8 | AAC72193 | pyruvate dehydrogenase | |

| | Corn protein | Description | Allergen |
|----|--------------|--|----------|
| 9 | AAC78468 | RNA polymerase sigma factor | |
| 10 | AAC79953 | anionic peroxidase H | |
| 11 | AAK26754 | plasma membrane integral protein ZmPIP-3 | |
| 12 | AAK30114 | teosinte branched protein 1 | |
| 13 | AAK51777 | MURB-like transposable element | |
| 14 | AAK51786 | MURB-like protein hMURB12 | |
| 15 | AAK56122 | alpha-expansin 4 | |
| 16 | AAK56130 | beta-expansin 7 | |
| 17 | AAK59898 | kaurene synthase A | |
| 18 | AAK60245 | teosinte branched protein | |
| 19 | AAK60502 | sucrose export defective 1 | |
| 20 | BAA05550 | Lea group 3 protein | |
| 21 | BAA22410 | Ca ²⁺ dependent protein kinase-kinase | |
| 22 | CAA28734 | 40.1 kDa A1 protein | |
| 23 | CAA31221 | cytochrome oxidase su III | |
| 24 | CAA35589 | pyruvate decarboxylase | |
| 25 | CAA37038 | dihydrodipicolinate | |
| 26 | CAA39438 | ribosomal protein S11 | |
| 27 | CAA60366 | hypothetical plastid protein | |
| 28 | CAA72196 | cytochrome p450 | |
| 29 | CAA87634 | unknown function w/APELTALA2-like binding domain | |
| 30 | CAB56627 | SBP domain protein 1 | |
| 31 | CAC18100 | putative legumain | |
| 32 | CAC35069 | VIP3 protein | |
| 33 | JC1524 | major allergen m1,pollen allergen | |
| 34 | 024578 | adenylosuccinate synthetase precursor | |
| 35 | P04705 | zein alpha-precursor | |
| 36 | P15590 | Embryo globulin-1 | |
| 37 | P33489 | auxin-binding protein-5 | |
| 38 | P40280 | histone H2a | |
| 39 | P46251 | actin depolymerizing factor | |
| 40 | P51059 | phosphoenolpyruvate carboxylase 2 | |
| 41 | PWZMAM | F1 ATPase mitochondria | |

| | Corn protein | Description | Allergen |
|----|--------------|---|----------|
| 42 | S12043 | leucoanthocyanidin dioxygenase | |
| 43 | S30062 | polygalacturonase | |
| 44 | S37379 | catalase 3 | |
| 45 | S58532 | chloroplast matK protein | |
| 46 | T02242 | GTP-binding protein rab2 | |
| 47 | T02763 | probable malate dehydrogenase | |
| 48 | T02990 | cinnamyl alcohol dehydrogenase | |
| 49 | T02993 | Ca ²⁺ dependent protein kinase | |
| 50 | T03397 | hypothetical protein | |

3.5.3 方法

3.5.1 に示したテスト配列のセットについては、本解析法を適用して、Cut-off 値を変動させた場合のアレルゲン予測精度 (Precision) および感度 (Recall) を検討した。なお、計算式は以下の通りである。

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

モチーフとのマッチングを検索する際に用いるMASTのE-value Cut-off 値に関して $10^{-2} \sim 10^{-30}$ まで変化させ、それぞれについてシーケンス相同性検索に用いるBLAST のCut-off 値を $10^{-1} \sim 10^{-30}$ まで変化させたときのPrecision およびRecall を計算し、最適な値を決めた。

また、3.5.2 に示したとうもろこしタンパク質のテストセットについては、先の検討結果により採用したCut-off 値 (MAST= 10^{-7} , BLAST= 10^{-8}) を用いて本解析法によるアレルゲン性予測を行い、FAO/WHOの方法 (n=8) と比較した。

3.5.4 結果

MASTのE-valueは、False Negativeを最小限に抑えることを考慮して Cut-off 値 10^{-7} を採用した（変動させたデータは示していない）。MASTのE-value Cut-off 値 10^{-7} で解析した場合のBLAST の Cut-off 値の影響を図2に示した。やはりFalse Negative を最小限に抑えることを優先して、BLAST のCut-off 値 10^{-8} を採用した。このときPrecision=75.5%、Recall=97.8%という値が得られた。

とうもろこしタンパク質 50 エントリーのアレルゲン性予測の一部を表14に示す。全ての結果は別紙資料6に記載した。本調査におけるアレルゲン性予測で、実際にアレルゲン性が報告されている2件（AAB86960 : profilin、JC1524 : pollen allergen Zea m 1）についてアレルゲン性を予測できた。また、FAO/WHOの方法では12件がアレルゲン性を持つと予測され、そのうち、11件が本解析法の結果と同様であった。

図2 【 Blast の E-value を変化させたときの Precision および Recall 】

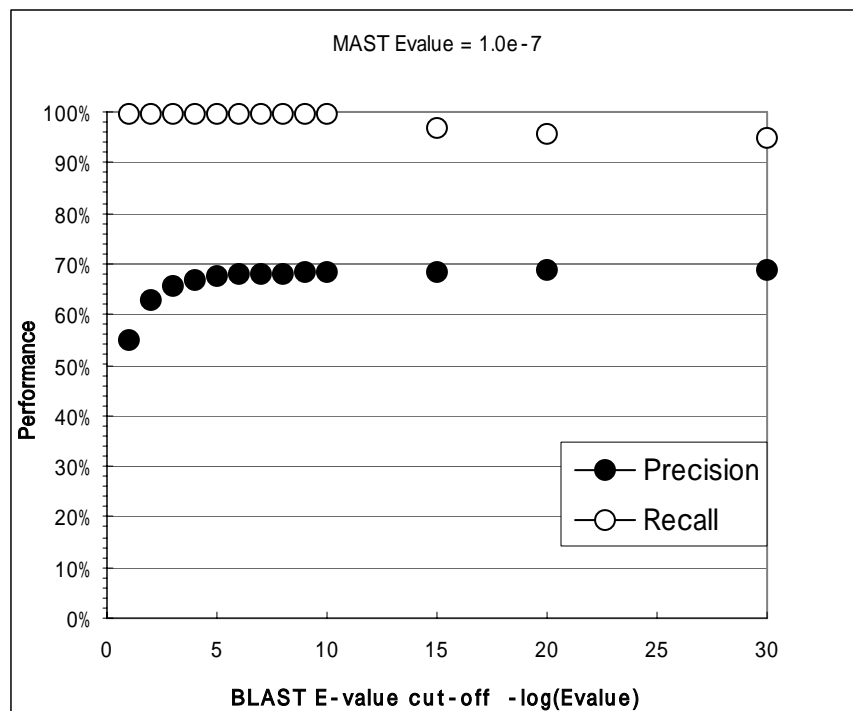


表 14 【結果とうもろこしタンパク質】

| No. | Corn protein | | Predict | |
|-----|---------------|------|---------|-----|
| | Accession No. | True | FAO/WHO | 本調査 |
| 2 | AAA33523 | | | |
| 6 | AAB71078 | | | |
| 7 | AAB86960 | | | |
| 15 | AAK56122 | | | |
| 16 | AAK56130 | | | |
| 20 | BAA05550 | | | |
| 33 | JC1524 | | | |
| 36 | P15590 | | | |
| 43 | S30062 | | | |
| 47 | T02763 | | | |
| 48 | T02990 | | | |
| 49 | T02993 | | | |

- 1 本報告書 2.3 により得られたアレルゲンと一致するものには「True」の列に記載
- 2 「FAO/WHO」の列は、FAO/WHO の方法でアレルゲンと予測されるもの。「本調査」の列は、本調査での解析方法でアレルゲンと予測されるもの。

3.5.5 考察

FAO/WHO の方法 (n=8) でテスト配列の実験をした結果では、Precision =67%、Recall = 99.7%という結果となり、Precision すなわち False Positive の少なさにおいては、本調査のほうが 8%ほど優れていた。この結果から、単なるシーケンスの相同性からアレルゲンを予測するよりも、モチーフを抽出して、相同性検索を行ったほうが、False Positive を低く抑えられる可能性が示唆された。また、Recall に関しては検索するとき用いる Cut-off 値を適切に定めることで、False Negative を最小限に抑えることができた。

とうもろこしタンパク質のエントリーのうち、FAO/WHO の方法で Positive と判断され、本解析で Negative と予測した AAA33523(Zein) は、とうもろこし中に多量に含有され、従来多くの人に摂食されてきたタンパク質であるが、現在のところアレルゲンとしての報告はない。このことから本解析方法で False Positive を低くすることができた可能性がある。

3.6 遺伝子組換えタンパク質の相同性解析

3.6.1 対象とした遺伝子組換えタンパク質

遺伝子組換え食品に利用されているタンパク質 34 種について、本調査におけるアレルゲン性予測手法を適用した。表 15 に用いたタンパク質の一覧を示す。なお、No.1 ~ 3 はアレルゲン性が認められないものとして知られている Negative Control である。

表 15 【遺伝子組換えタンパク質】

| 遺伝子組換えタンパク質 | | |
|-------------|------------|--|
| No. | Genbank ID | Name |
| 1 | 1070586 | myoglobin bovine |
| 2 | 132157 | Ribulose biphosphate carboxylase small chain (EC 4.1.1.39) (RuBisCOsmall subunit) |
| 3 | 668133 | ribulose-biphosphate carboxylase (EC 4.1.1.39) large chain precursor - spinach chloroplast |
| 4 | 80474 | parasporal crystal protein Bt2 - Bacillus thuringiensis subsp. kurstaki (strain HD-1) |
| 5 | 117326 | Pesticidal crystal protein cry1Ac (Insecticidal delta-endotoxin CryI(A(c)) (Crystalline entomocidal protoxin) (133 kDa crystal protein). |
| 6 | 231457 | Pseudomonas 6G5 |
| 7 | 124372 | Acetolactate Synthase |
| 8 | 124369 | Acetolactate Synthase II |
| 9 | 125463 | NTPII |
| 10 | | Barnase |
| 11 | 114831 | Barstar |
| 12 | 7428191 | Beta-glucuronidase |
| 13 | 460936 | Beta-lactamase |
| 14 | 115192 | Bromoxy nitrilase |
| 15 | 593495 | Coat protein (Cucumber mosaic virus) |
| 16 | 593497 | Coat protein (Papaya ringspot virus -P) |
| 17 | 593498 | Coat protein (Watermelon virus 2) |
| 18 | 593499 | Coat protein (Zucchini yellow mosaic virus) |
| 19 | 117533 | Crystal protein Cry1Ab |
| 20 | 117547 | Crystal protein Cry1Ac |

| 遺伝子組換えタンパク質 | | |
|-------------|------------|--|
| No. | Genbank ID | Name |
| 21 | 8469138 | Crystal protein Cry1Fa |
| 22 | 117328 | Crystal protein Cry2Ab |
| 23 | 117325 | Crystal protein Cry3Aa |
| 24 | 8469155 | Crystal protein Cry3Bb |
| 25 | 8469160 | Crystal protein Cry9Ca |
| 26 | 118682 | DNA adenine methyltransferase |
| 27 | 8469107 | Enoylpyruvate shikimate phosphate synthase |
| 28 | 1252836 | Glyphosate oxidoreductase |
| 29 | 8469218 | Lauroyl acyl carrier protein thioesterase |
| 30 | 5042199 | Nopaline synthase |
| 31 | 9629161 | Open reading frame 1 |
| 32 | 9629162 | Open reading frame 2a |
| 33 | 9629163 | Open reading frame 2b |
| 34 | 80811 | Phosphinothricin acetyltransferase (bar gene , Streptomyces hygrosopicus) |
| 35 | 8474524 | Phosphinothricin acetyltransferase |
| 36 | 3410642 | S-adenosylmethionine hydrolase |
| 37 | 67077 | ADD |

3.6.2 結果

3.5 で決定したMASTおよびBLASTのE-value Cut-off値(それぞれ 10^{-7} および 10^{-8})を適用し、表 15 のタンパク質に対してアレルゲン性予測を行った(表 16)。アレルゲンプロファイルおよびアレルゲンシーケンスのいずれかとマッチした場合はE-valueを示してある。マッチしなかったエントリーについては「-」と示した。BLASTに関しては、最も相同性の高かったエントリー(データベース名: Accession Number)を示した。

本手法によれば、表 15 の遺伝子組換えタンパク質 34 種の中には、アレルゲン性が予測されるものは存在しなかった。

表 16 【遺伝子組換えタンパク質に対する本調査の予測結果】

| No. | Genbank ID | 本調査 | | |
|-----|------------|----------|---------------|----------|
| | | MAST | BLAST | |
| | | E-value | hit | E-value |
| 1 | 1070586 | 1.00E-01 | trembl:Q94507 | 1.40E-02 |
| 2 | 132157 | 9.00E+00 | - | - |
| 3 | 668133 | 8.80E+00 | - | - |
| 4 | 80474 | 1.30E+00 | - | - |
| 5 | 117326 | 2.30E+00 | - | - |
| 6 | 231457 | - | - | - |
| 7 | 124372 | - | - | - |
| 8 | 124369 | - | - | - |
| 9 | 125463 | - | - | - |
| 10 | 133149 | - | - | - |
| 11 | 114831 | - | - | - |
| 12 | 7428191 | - | - | - |
| 13 | 460936 | - | - | - |
| 14 | 115192 | - | - | - |
| 15 | 593495 | - | - | - |
| 16 | 593497 | - | - | - |
| 17 | 593498 | - | - | - |
| 18 | 593499 | 4.3E-02 | - | - |
| 19 | 117533 | - | - | - |
| 20 | 117547 | - | - | - |
| 21 | 8469138 | - | - | - |

| No. | Genbank ID | 本調査 | | |
|-----|------------|---------|---------------|---------|
| | | MAST | BLAST | |
| | | E-value | hit | E-value |
| 22 | 117328 | - | trembl:P02538 | 5.5E-02 |
| 23 | 117325 | - | - | - |
| 24 | 8469155 | - | - | - |
| 25 | 8469160 | - | - | - |
| 26 | 118682 | - | - | - |
| 27 | 8469107 | 4.4E-01 | - | - |
| 28 | 1252836 | 4.8E-01 | - | - |
| 29 | 8469218 | - | - | - |
| 30 | 5042199 | - | - | - |
| 31 | 9629161 | - | - | - |
| 32 | 9629162 | 2.2E-04 | - | - |
| 33 | 9629163 | 1.2E+00 | - | - |
| 34 | 80811 | - | - | - |
| 35 | 8474524 | - | - | - |
| 36 | 3410642 | - | - | - |
| 37 | 67077 | - | - | - |

3.6.3 考察

本章で用いたエントリーの中には、すでに実用化され、食品中に組み込まれて食されているタンパク質も含まれるが、Negative Controlとして用意されたタンパク質と同様にアレルゲン性は予測されず、本解析法によってはアレルギー性を誘発する可能性は見出されなかった。

3.7 アレルゲン相同性検索についてのまとめ

第3章では、アレルゲン性予測手法について調査ならびに解析を行った。本調査では、FAO/WHOの方法、Two-track法およびMotif-based predictionの三種の方法のうち、もっともパフォーマンスが高いとされるStadlerらのMotif-based prediction(30)を参考に、独自のアレルゲン性予測手法を開発した。その方法は、アレルゲンデータベースからモチーフを抽出し、モチーフにマッチしないアレルゲンをアレルゲンシーケンスとし、対象とするタンパク質とモチーフとのマッチング、またはアレルゲンシーケンスとの相同性を考慮してアレルゲン性を予測するというものである。

その結果、FAO/WHOの方法に比べ、同程度のRecallを確保しながらもPrecisionを向上することに成功した。利用したサンプルセットに違いがあるが、テスト配列を用いて本解析とFAO/WHOの方法とを比較すると、Precision[FAO/WHO(n=8) = 67%、本調査 = 75.5%]、Recall[FAO/WHO(n=8) = 99.7%、本調査 = 97.8%]という結果であった。

PrecisionとはすなわちFalse Positiveの少なさをさすが、シーケンスベースの相同性比較であるFAO/WHOの方法は、比較的False Positiveが多いことが知られている。本調査で採用したMotif-based predictionでは、多くのアレルゲンに共通して存在するアミノ酸配列をモチーフとして抽出することにより、より信頼性の高い相同性比較を実現した。

今後、公共データベースのさらなる拡充、および、より適正なモチーフの抽出・補正手法等の確立により、Motif-based predictionはより高いPrecisionやRecallを実現できるものと期待される。

4

参考文献およびサイト一覧

4 参考文献およびサイト一覧

1. Baxevanis AD.
The molecular biology database collection: 2002 update.
Nucleic Acids Res 2002;30: 1-12.
2. Wu CH, Huang H, Arminski L, Castro-Alvear J, Chen Y, Hu ZZ et al.
The protein information resource: an integrated public resource of functional annotation of proteins.
Nucleic Acids Res 2002;30: 35-37.
3. O'Donovan C, Martin MJ, Gattiker A, Gasteiger E, Bairoch A, Apweiler R.
High-quality protein knowledge resource: SWISS-PROT and TrEMBL.
Brief Bioinform 2002;3: 275-284.
4. King TP, Hoffman D, Lowenstein H, Marsh DG, Platts-Mills TA, Thomas W.
Allergen nomenclature.
Allergy 1995;50: 765-774.
5. Metcalfe DD, Astwood JD, Townsend R, Sampson HA, Taylor SL, Fuchs RL.
Assessment of the allergenic potential of foods derived from genetically engineered crop plants.
Crit Rev Food Sci Nutr 1996;36: S165-S186.
6. Gendel SM.
Sequence databases for assessing the potential allergenicity of proteins used in transgenic foods.
Adv Food Nutr Res 1998;42: 63-92.
7. Pearson WR, Lipman DJ.
Improved tools for biological sequence comparison.
Proc Natl Acad Sci USA 1988;85: 2444-2448.
8. Wjst M, Immervoll T.

An internet linkage and mutation database for the complex phenotype asthma.

Bioinformatics 1998;14: 827-828.

9. Ivanciuc O, Schein CH, Braun W.
SDAP: database and computational tools for allergenic proteins.
Nucleic Acids Res. 2003 Jan 1;31(1):359-62.
10. <http://www.niaid.nih.gov/publications/allergens/title.htm>
11. <http://dmd.nihs.go.jp/latex/allergen-e.html>
12. <http://www.cat.at/pollen/index.en.html>
13. <http://www.ean.cc/>
14. <http://www.hon.ch/Library/Theme/Allergy/Glossary/allergy.html>
15. Brusic V, Millot M, Petrovsky N, Gendel SM, Gigonzac O, Stelman SJ.
Allergen databases.
Allergy. 2003 Nov;58(11):1093-100.
16. <http://www.rcsb.org/pdb/>
17. <http://srs.ebi.ac.uk/>
18. Dodge C, Schneider R, Sander C.
The HSSP database of protein structure-sequence alignments and family profiles.
Nucleic Acids Res. 1998 Jan 1;26(1):313-5.
19. Allergen Nomenclature Sub-Committee, International Union of Immunological Societies. Allergen Nomenclature.
(<http://www.allergen.org/List.htm>)
20. <http://www.nlm.nih.gov/hinfo.html>

21. D. S. Shin, C. M. Compadre, S. J. Maleki, R. A. Kopper, H. Sampson, S. K. Huang, A. W. Burks, and G. A. Bannon, Biochemical and structural analysis of the IgE binding sites on Ara h 1, an abundant and highly allergenic peanut protein, *J. Biol. Chem.* 1998, 273, 13753-13759.
22. J. S. Stanley, N. King, A. W. Burks, S. K. Huang, H. Sampson, G. Cockrell, R. M. Helm, C. M. West, and G. A. Bannon, Identification and mutational analysis of the immunodominant IgE binding epitopes of the major peanut allergen Ara h 2, *Arch. Biochem. Biophys.* 1997, 342, 244-253.
23. P. Rabjohn, E. M. Helm, J. S. Stanley, C. M. West, H. A. Sampson, A. W. Burks, and G. A. Bannon, Molecular cloning and epitope analysis of the peanut allergen Ara h 3, *J. Clin. Invest.* 1999, 103, 535-542.
24. V. P. Kurup, B. Banerjee, P. S. Murali, P. A. Greenberger, M. Krishnan, V. Hari, and J. N. Fink, Immunodominant peptide epitopes of allergen, Asp f 1 from the fungus *Aspergillus fumigatus*, *Peptides* 1998, 19, 1469-1477.
25. L.-P. Chow, S.-L. Liu, C.-J. Yu, H.-K. Liao, J.-J. Tsai, and T.-K. Tang, Identification and expression of an allergen Asp f 13 from *Aspergillus fumigatus* and epitope mapping using human IgE antibodies and rabbit polyclonal antibodies, *Biochem. J.* 2000, 346, 423-431.
26. K. V. Soman, T. Midoro-Horiuti, J. C. Ferreon, R. M. Goldblum, E. G. Brooks, A. Kurosky, W. Braun, and C. H. Schein, Homology modeling and characterization of IgE binding epitopes of mountain cedar allergen Jun a 3,

- Biophys. J. 2000, 79, 1601-1609.
27. B. Banerjee, K. Kanitpong, J. N. Fink, M. Zussman, G. L. Sussman, K. J. Kelly, and V. P. Kurup,
Unique and shared IgE epitopes of Hev b 1 and Hev b 3 in latex allergy,
Mol. Immunol. 2000, 37, 789-798.
 28. D. H. Beezhold, V. L. Hickey, and G. L. Sussman, Mutational analysis
of the IgE epitopes in the latex allergen Hev b 5, J. Allergy Clin.
Immunol. 2001, 107, 1069-1076.
 29. FAO/WHO (2001) guidelines for allergenicity evaluation for foods
derived from biotechnology
 30. Michael B. Stadler and Beda M. Stadler
Allergenicity prediction by protein sequence.
The FASEB Journal express article 10. 1096/fj.02-1052fje. Published
online April 22 , 2003
 31. Ronald E. Hileman Andre Silvanovich Richard E. Goodman Elena A. Rice
Gyula Hollescahak James D. Astwood Susan L. Hefle
Bioinformatic Methods for Allergenicity Assessment Using a
Comprehensive allergen database.
Int Arch Allergy Immunol 2002;128;280-291
 32. Gijs A Kleter and Ad ACM Peijnenbug
Screening of transgenic proteins expressed in transgenic food crops
for the presence of short amino acid sequences identical to
potential , IgE binding linear epitopes of allergens.
BMC Structural Biology, 2, 8, 2002
 33. Altschul, Stephen F., Thomas L. Madden, Alejandro A.
Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J.
Lipman
Gapped BLAST and PSI-BLAST: a new generation of protein database
search programs

Nucleic Acids Res. 25:3389-3402. 1997

34. <http://meme.sdsc.edu/meme/website/meme-intro.html>

35. <http://meme.sdsc.edu/meme/website/mast-intro.html>

36. ftp://ftp.fao.org/codex/alinorm03/al03_34e.pdf

参考資料

- 資料 1 SDAP データベースにおける各アレルゲンが持つエピトープ配列
- 資料 2 立体構造情報を持つアレルゲンの全リスト
- 資料 3 調査で選んだ論文に記載されていた全アレルゲンと全エピトープ配列
- 資料 4 同じく調査で選んだ論文に記載されていた全アレルゲンと全エピトープ数
- 資料 5 調査で選んだ論文リスト
- 資料 6 表 14 の詳細結果

平成 16 年 3 月

タンパク質のアレルギー評価手法に関する調査報告書

原本作成 CTC ラボラトリーシステムズ株式会社 ライフサイエンス部
〒154-0012 東京都世田谷区駒沢 1-16-7 中村ビル 6F
TEL 03-3419-9171
FAX 03-3419-9179

無断複写・転載を禁止します。