Wegweiser durch den Allergendschungel: Allergendatenbanken, ihre Merkmale und Anwendungsgebiete

Christian Radauer

Institut für Pathophysiologie und Allergieforschung, Zentrum für Pathophysiologie, Infektiologie und Immunologie, Medizinische Universität Wien, Wien, Österreich

Schlüsselwörter

Allergene · Datenbanken · Bioinformatik · Epitope · Proteinfamilien

Zusammenfassung

Der wachsende Umfang an verfügbaren Daten zu allergenen Proteinen hat die Einrichtung strukturierter, frei zugänglicher Allergendatenbanken erforderlich gemacht. In diesem Artikel werden die Funktionen und Anwendungszwecke von 6 der meistgenutzten Allergendatenbanken behandelt. Die *WHO/IUIS Allergen Nomenclature Database* ist die offizielle Quelle für die Benennung von Allergenen. *Allergome* ist die umfassendste Datensammlung zu Allergenen und Allergenquellen. *AllergenOnline* soll als Peer-Review-Datenbank für Allergensequenzen dazu dienen, die Allergenität von Proteinen vorherzusagen, die z.B. in gentechnisch veränderte Nutzpflanzen eingebracht werden sollen. Die *Structural Database of Allergenic Proteins (SDAP)* ist eine Da-

tenbank mit Allergensequenzen, -strukturen und -epitopen, verknüpft mit Bioinformatik-Tools für Sequenzanalysen und -vergleiche. Die *Immune Epitope Database (IEDB)* ist die größte Sammlung von T-Zell-, B-Zell- und Haupthistokompatibilitätskomplex-Epitopen einschließlich Allergen-Epitopen. *AllFam* ordnet Allergene verschiedenen Familien evolutionär verwandter Proteine zu, aufbauend auf Definitionen aus der Proteinfamilien-Datenbank *Pfam*. Diese Datenbanken enthalten großteils überlappende Daten, unterscheiden sich aber im Hinblick auf die Nutzerzielgruppe, die Kriterien für die Aufnahme von Allergenen, die für die einzelnen Allergene erfassten Daten und die Verfügbarkeit von Bioinformatik-Tools.

Einleitung

Seit in den späten 1980er Jahren die ersten Allergene entdeckt, kloniert und sequenziert wurden, sind Hunderte von Allergenen aus unterschiedlichsten Quellen identifiziert worden. Angesichts des rapide wachsenden Datenbestands über Allergene, ihre Sequenzen, biochemischen Merkmale, IgE-Bindungs- und sonstigen klinischen Eigenschaften sowie ihre evolutionären und immunologischen Verwandtschaften entstand der Bedarf nach öffentlich zugänglichen Ressourcen für Allergendaten. Als in den 1990er Jahren das World Wide Web allgemein verfügbar wurde, bereitete das den Boden für die Einrichtung verschiedener frei zugänglicher Allergendatenbanken. Diese Datenbanken enthalten weitenteils überlappende Daten, unterscheiden sich aber im Hinblick auf die Nutzerzielgruppe, die Kriterien für die Aufnahme von

Allergenen, die für die einzelnen Allergene erfassten Daten und die Verfügbarkeit von zusätzlichen Tools, z.B. für Sequenzvergleiche.

Die vorliegende Übersichtsarbeit gibt einen kritischen Überblick über die Merkmale, Stärken und Schwächen der sechs derzeit meistgenutzten frei zugänglichen Allergendatenbanken. Darüber hinaus werden einige Beispiele für Forschungsfragen gegeben, die sich mit diesen Datenbanken jeweils bearbeiten lassen. Tabelle 1 und Tabelle 2 zeigen eine Zusammenfassung ihrer wichtigsten Merkmale. Die vorliegende Arbeit beschränkt sich auf Datenbanken von Proteinallergenen, die IgE-vermittelte (Typ-I)-Allergien auslösen; nicht behandelt werden Allergene, die für Reaktionen vom verzögerten Typ (Typ IV) verantwortlich sind.

Tab. 1. Eckdaten der wichtigsten Allergendatenbanken

Name	Adresse	Erstmalige Freischaltung	Betreiber	Literatur
WHO/IUIS Allergen Nomenclature Database	www.allergen.org	2000	WHO/IUIS Allergen Nomenclature Sub-Committee	3, 7, 9
Allergome	www.allergome.org	2005	Allergy Data Laboratories, Latina, Italien	10, 11
AllergenOnline	www.allergenonline.org	2005	Food Allergy Research and Resource Program, Department of Food Science and Technology, University of Nebraska-Lincoln, Lincoln, NE, USA	19
Structural Database of Allergenic Proteins (SDAP)	fermi.utmb.edu/SDAP/	2002	Sealy Center for Structural Biology, Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX, USA	26, 27
Immune Epitope Database (IEDB)	www.iedb.org	2006	National Institute for Allergy and Infectious Diseases, Bethesda, MD, USA	36, 38
AllFam	www.meduniwien.ac.at/allfam/	2007	Division of Medical Biotechnology, Department of Pathophysiology and Allergy Research, Medical University of Vienna, Austria	55

Allergendatenbanken

World Health Organization/International Union of Immunological Societies Allergen Nomenclature Database

Ursprünglich waren die Wissenschaftler frei in der Benennung der Allergene, die sie entdeckt hatten, und jeder verwendete ein anderes Schema der Namensbildung. Die ersten Allergene, die in den 1970er und frühen 1980er Jahren identifiziert wurden, erhielten nicht-systematische Namen wie Antigen E (heute Amb a 1) und Ra5 (jetzt Amb a 5) aus Ambrosia (Ragweed), Allergen M (jetzt Gad m 1 und Gad c 1) aus Kabeljau oder cat-1 (jetzt Fel d 1) von der Katze. In einigen Fällen verwendeten verschiedene Gruppen sogar unterschiedliche Namen für dieselben Allergene. Um ein offizielles, einheitliches Benennungssystem einzuführen, wurde 1984 unter der Schirmherrschaft der Weltgesundheitsorganisation (WHO) und der International Union of Immunological Societies (IUIS) das WHO/IUIS Allergen Nomenclature Sub-Committee ins Leben gerufen. Diesem Unterausschuss gehören derzeit 22 führende Experten für die Charakterisierung, Struktur, Funktion, Molekularbiologie und Bioinformatik von Allergenen an. Seine Aufgabe ist es, eine einheitliche, eindeutige, systematische Nomenklatur für alle Proteine, die beim Menschen IgE-vermittelte Allergien induzieren können, zu entwickeln und zu verwalten.

Die offizielle Allergennomenklatur wurde erstmals 1986 veröffentlicht [1], 1994 erschien eine überarbeitete Fassung [2–6]. Diese noch heute geltende überarbeitete Nomenklatur fußt auf der Linnéschen binären Nomenklatur, die jeden Organismus durch Gattungs- und Artnamen bezeichnet. Die Allergennamen beginnen mit den ersten 3–4 Buchstaben des Gattungsnamens der Allergenquelle, darauf folgt ein Leerzeichen, dann die ersten 1–2 Buchstaben des Artnamens, ein Leerzeichen, eine Zahl (die meist in der chronologischen Reihenfolge der Entdeckung des Allergens in der jeweiligen Quelle vergeben wird), ein Punkt und eine vierstellige Zahl zur Bezeichnung

von Isoformen und Varianten des Allergens. Verwandte Allergensequenzen einer Spezies werden als Isoformen desselben Allergens zusammengefasst, wenn ihre Sequenzen zu mehr als 67% identisch sind. Als Varianten eines Allergens sind Sequenzen definiert, die zu >90% identisch sind. Für eine ausführlichere Darstellung der Nomenklatur sei der Leser auf Originalpublikationen zur Nomenklatur sowie einige neuere Artikel mit aktuellen Erkenntnissen zu einzelnen Aspekten der Allergennomenklatur verwiesen [7–9].

Ein Verzeichnis aller offiziell anerkannten Allergennamen ist seit 2000 im Internet verfügbar (www.allergen.org). Ursprünglich bestand es aus einer einfachen Tabelle mit dazugehörigen Literaturverweisen. Später wurde es zu einer leistungsfähigen Datenbank mit Blätter- und Suchfunktionen und Verknüpfungen zu anderen Datenbanken weiterentwickelt. Neue Einträge werden von den Mitgliedern des Allergen Nomenclature Sub-Committee in die Datenbank aufgenommen. Um eine einheitliche, den Richtlinien entsprechende Allergennomenklatur aufrechtzuerhalten, müssen Forscher Allergene, die sie neu entdeckt haben, beim Unterausschuss einreichen, bevor sie ihr Manuskript einer Fachzeitschrift zur Veröffentlichung vorlegen. Diese Einreichungen werden vertraulich behandelt; bis zur Veröffentlichung erscheinen auf der Website keine konkreten Angaben außer dem Namen des neuen Allergens, seiner biochemischen Bezeichnung und grundlegenden Angaben zu seiner Allergenität. Ein Formular zur Einreichung eines neuen Allergens ist unter www.allergen.org abrufbar. Jede Einreichung wird von mindestens 2 Mitgliedern des Unterausschusses geprüft. Um in die Datenbank aufgenommen zu werden und einen offiziellen Allergennamen zu erhalten, müssen eingereichte Allergene die folgenden Kriterien erfüllen: (i) zumindest teilweise erfolgte Sequenzierung und Eintrag in eine Nukleotid- oder Proteinsequenzdatenbank (wobei die Datenbank auch einige alte Einträge ohne entsprechende Zugangsnummer enthält), (ii) eindeutige Identifizierung der Expression in der Quelle auf

Tab. 2. Merkmale der wichtigsten Allergendatenbanken

Name	Anzahl enthaltener Allergene ^a	Erfasste Daten	Suchwerkzeuge	Aktualisierungshäufigkeit	Bemerkungen
WHO/IUIS Allergen Nomenclature Database	869 Allergene 1250 Isoformen und Varianten	Offizieller Allergenname, biochemische Bezeichnung, Quelle, Isoallergene, Literatur (Erstbeschreibung), Links zu externen Datenbanken (Literatur, Sequenz, Struktur)	Suche nach Allergennamen, biochemischen Namen und Quelle	Fortlaufend	Offizielle Referenz für Allergennamen und entsprechende Sequenzen; neue Allergene werden nach Prüfung durch das Allergen Nomenclature Sub-Committee aufgenommen
Allergome	3104 allergene Moleküle und ihre Isoformen	Allergenname, biochemische Funktion, Isoformen, Allergenquelle, Expositionsweg, Diagnostika, Links zu Sequenzen und Strukturen, Sequenzhomologe, kreuzreaktive Allergene, allergene Eigenschaften und Epidemiologie, umfangreiche Bibliographie nach Themenbereichen gruppiert	Textsuche über alle Datenbankfelder, Sequenzvergleich, leistungsstarke Literatursuche	Fortlaufend	Umfangreiche Datensammlung zu Allergenen, zusammengetragen aus anderen Datenbanken und der Literatur; alle Allergene, zu denen publizierte Daten vorliegen, sind enthalten, ohne Filterung nach Relevanz; nicht alle Allergennamen sind offiziell anerkannt
AllergenOnline	808 Allergengruppen 2035 Sequenzen	Allergenname, Typ (Inhalation, Lebensmittel etc.), Quelle, Sequenz, ausgewählte Literatur	Textsuche, Sequenzvergleich mit verschiedenen Methoden	Jährlich	Liste der Allergene wird durch Expertengremium geprüft, um eine zuverlässige Datenbank mit relevanten Allergensequenzen zu gewährleisten
Structural Database of Allergenic Proteins (SDAP)	1526 Allergene und ihre Isoformen	Allergenname und -quelle, Proteinfamilie, ausgewählte Literatur, Links zu Sequenzen und Strukturen	Sequenzvergleich, Peptidvergleich, Links zu externen Servern (z. B. BLAST)	Unbekannt ^b	Peer-Review-begutachtete Allergensequenz-Datenbank mit Sammlung von Bioinformatik-Tools; Schwerpunkt auf Allergenstrukturen und Epitopen
Immune Epitope Database (IEDB)	8288 Epitope von 322 Allergenen ^c	T- und B-Zell-Epitope von Protein- und Nichtprotein-Antigenen einschließlich Allergenen, Art des Assays, Quellantigen und -organismus, Wirtsorganismus, Art der Krankheit	Suche über alle Datenbankfelder (z. B. Organismus, Wirt, Art des Assays, Art des Epitops); Tools für die Epitopprognose	Vierteljährlich	Größte Sammlung von Epitopdaten für alle Arten von Antigenen
AllFam	1018 Allergene ohne Isoformen	Allergenname mit Links zu WHO/IUIS-Datenbank und AllergenOnline, Expositionsweg, Abstracts zu Proteinfamilien mit ausgewählter Literatur	Auflistung der Mitglieder von Allergenfamilien (gefiltert nach Quelle und Exposition), Textsuche nach Allergenen, Quellen und Proteinfamilien	Vierteljährlich	Klassifizierung von Allergenen nach Proteinfamilie

^aStand 2. März 2017. ^bKeine Aktualisierungen seit 2013. ^cOhne nichtpeptidische Epitope.

mRNA- oder Proteinebene und (iii) Bindung von IgE aus Seren von mindestens 5 Patienten, die gegen die jeweilige Quelle allergisch sind. Im März 2017 enthielt die Allergenomenklatur-Datenbank der WHO/IUIS 869 Allergene mit 1250 Isoformen und Varianten. Der Datensatz jedes offiziell anerkannten Allergens umfasst seinen Namen sowie biochemische Namen, veraltete Namen, Angaben zur

Quelle (wissenschaftlicher und allgemeinsprachlicher Name der Art, Ordnung und Gruppe) und grundlegende Angaben zu seiner Allergenität einschließlich eines Links zu der Publikation mit der Erstbeschreibung. Bei neueren Einträgen werden auch Name und Institution des Einreichers veröffentlicht. Darüber hinaus enthält jeder Eintrag Daten zu Isoformen und Varianten des Allergens wie Isoall-

ergennummer, Peptid- und Nukleinsäuresequenzen, Molekülmasse, Allergenität rekombinanter Isoallergene sowie Links zu Publikationen zur Allergenität und Sequenzierung in PubMed, zu Nukleinsäure- und Peptidsequenzen in GenBank und UniProt sowie zu Strukturen in der Protein Data Bank (PDB). Allergene und Isoformen, die nach oraler Einnahme allergische Reaktionen hervorrufen, werden als Nahrungsmittelallergene gekennzeichnet. Bei allen anderen Arten von Allergenen hingegen werden keine Angaben zum Expositionsweg gemacht. Das Allergenverzeichnis der *WHO/IUIS Allergen Nomenclature Database* ist die wichtigste Datenquelle für die meisten anderen Allergendatenbanken wie *AllergenOnline*, *SDAP* oder *AllFam*.

Allergome

Die *Allergome*-Datenbank (www.allergome.org) wurde 2003 freigeschaltet und wird von Allergen Data Laboratories betrieben, einem Unternehmen in Latina, Italien [10, 11]. Hier werden sämtliche Daten zu Allergenquellen und allergenen Molekülen aus der Literatur und aus anderen Datenbanken zusammengetragen. *Allergome* ist die wohl umfassendste Sammlung von Allergendaten. Im März 2017 enthielt sie Daten zu mehr als 3500 Allergenquellen und 3000 Allergenen (www.allergome.org/script/statistic.php). Im Gegensatz zur offiziellen Nomenklatur werden Allergene hier als «Moleküle» bezeichnet, während alle Datensätze (Allergene und Quellen) zusammenfassend als «Allergene» bezeichnet werden. Die in die Datenbank eingegebenen Daten werden nicht nach Relevanz gefiltert, sondern jeder Allergendatensatz enthält einen sogenannten Allergenitäts-Score als schnellen Überblick darüber, wie gut das Allergen im Hinblick auf Sequenz, IgE-Bindungs- und Aktivitätsassays, Hauttests, Provokationstests und epidemiologische Daten beschrieben ist. Zusätzlich zu etablierten Allergenen sind in *Allergome* auch Proteine aufgeführt, die Homologe von Allergenen sind, bisher jedoch nicht als Allergene identifiziert wurden oder aus Quellen stammen, die nicht von allergologischer Bedeutung sind. Diese Einträge sind als «in silico» gekennzeichnet.

Ein Allergendatensatz umfasst hier den Allergennamen, im Fall offiziell anerkannter Allergene einen Link zum Allergendatensatz in der WHO/IUIS-Datenbank sowie biochemische Namen. Für viele Allergene, die noch nicht vom WHO/IUIS Allergen Nomenclature Sub-Committee anerkannt sind, werden IUIS-ähnliche Allergennamen angegeben. Hierbei ist jedoch zu unterstreichen, dass diese Namen nicht vom Allergen Nomenclature Sub-Committee anerkannt sind und möglicherweise auch niemals anerkannt oder sogar geändert werden, falls dem jeweiligen Allergen später ein offizieller Name zugewiesen wird. Außerdem enthält der Datensatz noch Angaben zur Allergenquelle, darunter den wissenschaftlichen Namen, umgangssprachliche Namen in mehreren Sprachen, Bilder sowie Links zu taxonomischen Datenbanken wie der des National Center for Biotechnology Information (NCBI) oder UniProt. Die biochemischen Daten umfassen die biologische Funktion, eine Liste der Isoformen und Varianten mit Verknüpfungen zu Sequenzdaten in UniProt und zu Strukturdaten in der PDB sowie das Gewebe, in dem das Allergen exprimiert wird. An allergologischen Daten enthält die Datenbank die Expositionswege sowie eine Tabelle mit epidemiologischen Daten aus der Literatur unter Angabe der Einschlusskriterien der untersuchten Patientenkohorte, der Art des Tests, der Anzahl der Patienten sowie

des Anteils positiver Ergebnisse. Zu jedem Allergendatensatz gehört auch eine umfassende Liste mit Literaturverweisen, die nach Themenbereichen kategorisiert sind, z.B. Biochemie/Struktur/Funktion, Molekularbiologie oder Immunchemie/Allergenität.

Allergome kann entweder mit einer einfachen Textsuche oder einer erweiterten Suche durchsucht werden; bei letzterer kann der Nutzer die zu durchsuchenden Datenbankfelder eingrenzen. Die Website bietet außerdem eine Vielzahl von Data-Mining-Tools, die die Extraktion zusätzlicher Daten zu gut erforschten Allergenen ermöglichen. Die Datensätze von Allergenen mit bekannter Sequenz enthalten eine Seite mit den Proteinsequenzen aller Isoforme und Varianten. Hier sind auch Sequenzmotive aufgeführt, die von einem Algorithmus erzeugt wurden (Originalpublikation von Stadler und Stadler [12]), um Allergene in Familien einzuordnen und anhand der Sequenz Prognosen zur Allergenität abgeben zu können. Mit dem Tool *AllergomeAligner* lassen sich Sequenzen und Motive mit *Allergome* oder *UniProt* mithilfe unterschiedlicher Sequenzähnlichkeits-Suchwerkzeuge abgleichen. Des Weiteren enthält der Allergendatensatz den Allergome O-Ring, eine grafische Darstellung von Sequenzähnlichkeiten sowie IgE-Kreuzreaktivitäten innerhalb von Gruppen verwandter Allergene. Das *ReTiME-Modul* (Real Time Monitoring of IgE Sensitization) ist eine Datenbank, in der mitwirkende Allergiekliniken geographisch aufgelöste epidemiologische Daten zur Sensibilisierung gegen Allergene zusammentragen und speichern. Mit dem dem *RefArray*-Tool lässt sich in der Literatur zum Thema Allergene recherchieren. Der Nutzer kann die *Allergome*-Literaturdatenbank nach verschiedenen Kriterien durchsuchen, z.B. Allergennamen, thematische Kategorie (Biochemie, Immunologie/Allergenität usw.) oder bibliographische Standardkategorien wie Erscheinungsjahr, Zeitschrift oder Autoren.

Allergome wird in der wissenschaftlichen Fachwelt häufig genutzt. Bis 2014 ist *Allergome* laut seiner Website in 308 Publikationen zitiert worden (www.allergome.org/script/about.php#database). Im Folgenden sollen nur einige wenige Beispiele neuerer Studien genannt werden. Erstens ist der grundlegendste Anwendungszweck von *Allergome* der Abruf einer vollständigen Liste aller bekannten Allergene aus einer bestimmten Quelle. Dimitrov und Doytchinova [13] nutzten *Allergome*, um die Sequenzen von 13 Milch- und Ei-Allergenen abzurufen und deren gastrointestinale Verdauung sowie die HLA-Bindung der entstehenden Peptide zu simulieren. Sie prognostizierten die Bindung von 3 und 4 HLA-DR- und DQ-Allelen an Peptide aus Milch- und Ei-Allergenen, wobei mehrere Allele keine Bindung an Allergenpeptide zeigten und ihnen daher Schutzwirkung gegenüber der jeweiligen Allergie zugesprochen wurde. Zweitens enthält *Allergome* eine durchsuchbare Sequenzdatenbank. Diese Daten werden entsprechend verwendet, um neu entdeckte Allergene mit bekannten zu vergleichen oder auf der Grundlage von Homologien das Vorliegen putativer Allergene in sequenzierten Proteomen, Transkriptomen oder Genomen vorherzusagen. Campbell et al. [14] analysierten das gesamte Pollen-Proteom und -Transkriptom von Wilder Mohrenhirse, eines subtropischen Grases. Indem sie Transkriptom- und Proteomdaten sowie Sequenzen IgE-bindender Proteine, die durch Massenspektrometrie identifiziert wurden, mit *Allergome* verglichen, zeigten die Autoren, dass dieses Gras Vertreter sämtlicher Pollenallergenfamilien enthält,

die zuvor in anderen subtropischen Gräsern identifiziert worden waren, nicht jedoch die Allergene der Gruppe 5 und 6, die auf Pollen von Gräsern aus gemäßigten Zonen beschränkt sind. Drittens enthält *Allergome* als einzige Datenbank Daten zu Allergenquellen einschließlich solcher, bei denen die enthaltenen Allergene nicht identifiziert sind. Wir verglichen anhand dieser Daten die taxonomischen Verteilungen derjenigen Pflanzenarten, von denen bekannt ist, dass sie Pollenallergien hervorrufen, und derjenigen Arten, die identifizierte Allergene enthalten [15]. Während in *Allergome* 178 Pollenallergenquellen aufgeführt wurden, waren nur bei 52 Arten Allergene identifiziert. Insbesondere bei tropischen und subtropischen Pflanzen fehlte es an Allergendaten. Diese Ergebnisse könnten die Aufmerksamkeit zukünftiger Forschungsarbeiten verstärkt auf diese Pflanzenfamilien richten, denn ihre Bedeutung als Quelle allergener Pollen steigt infolge der zunehmenden Prävalenz von Allergien in den Entwicklungsländern und der Auswirkungen des Klimawandels auf die geographische Verteilung allergener Pflanzen [16].

AllergenOnline

Zur Sicherheitsbeurteilung von gentechnisch veränderten Organismen gehört vor der Markteinführung die Bewertung ihres allergenen Potenzials im Vergleich zum jeweiligen Wildtyp. Der prototypische Fall eines allergenen 2S-Albumins, das in gentechnisch veränderte, für Tierfutter bestimmte Sojabohnen eingebracht wurde, die aus diesem Grund nie auf den Markt kamen, führte zur Entwicklung von Richtlinien für die Allergenitätsprüfung neuartiger Lebensmittel [17]. Sie sehen Screenings mit Seren von Allergiepatienten ebenso vor wie die Testung der Proteine auf Resistenz gegen gastrointestinale Verdauung, Tierstudien und die bioinformatische Beurteilung der Ähnlichkeit neu eingebrachter Gene mit bekannten Allergenen [18].

Die *AllergenOnline*-Datenbank (www.allergenonline.org) wurde im Jahr 2005 vom Food Allergy Research and Resource Program am Institut für Lebensmittelwissenschaften und -technologie der University of Nebraska in Lincoln, NE, USA, eingerichtet. Sie soll als Peer-Review-gestützte Datenbank von Allergensequenzen dazu dienen, die Allergenität von Proteinen zu prüfen, die in gentechnisch veränderte Lebensmittel eingebracht werden [19]. Die Begründung für die Durchführung von Sequenzvergleichen mit bekannten Allergenen liegt in der Annahme, dass Proteine mit signifikanter Sequenzähnlichkeit mit Allergenen auch Kreuzreaktivität mit IgE- oder T-Zellen, die für dieses Allergen spezifisch sind, zeigen und dadurch allergische Symptome hervorrufen könnten. *AllergenOnline* sammelt putative Allergensequenzen aus der NCBI-Proteindatenbank, ergänzt durch Daten aus der *WHO/IUIS Allergen Nomenclature Database* und *Allergome*. Diese Liste wird dann auf der Grundlage von Peer-Review-Publikationen von einem Gutachtergremium gefiltert. Sequenzen werden nur dann in *AllergenOnline* aufgenommen, wenn es eindeutige Belege dafür gibt, dass das Protein IgE aus Seren von Patienten bindet, die allergisch gegen die jeweilige Quelle sind, und dass das Protein in dem relevanten Gewebe der Quelle exprimiert wird. Dem Gutachterstab gehören derzeit 8 Wissenschaftler an. Einzelheiten zum Evaluierungsprozess sind der Website von *AllergenOnline* zu entnehmen (<http://www.allergenonline.org/about.shtm>).

ml). *AllergenOnline* wird jährlich aktualisiert. Die aktuelle Version 17 aus dem Januar 2017 enthält 2035 Sequenzen.

Die Benutzeroberfläche von *AllergenOnline* enthält eine Liste aller Einträge, die mit Suchbegriffen in allen Datenbankfeldern gefiltert werden kann. Jeder Sequenz-Datensatz enthält Angaben zur Art (wissenschaftlicher und umgangssprachlicher Name), dem Allergen-namen laut WHO/IUIS (falls vorhanden), der Art des Allergens (Expositionsweg und Art der Quelle; z.B. Milben-Aeroallergen, pflanzliches Nahrungsmittelallergen), der Allergengruppe und der Sequenzlänge sowie einen Link zum Eintrag der Sequenz in der NCBI-Proteindatenbank. Allergengruppen enthalten verwandte Allergene derselben Spezies oder Gattung mit mehr als 67% Sequenzidentität. Eine in *AllergenOnline* definierte Gruppe entspricht somit in der Regel (aber nicht immer) einem Allergen im Sinne der spezi-esbasierten WHO/IUIS-Nomenklatur. Jede Gruppe ist mit einer Liste von Literaturverweisen und Sequenzeinträgen verknüpft.

AllergenOnline wurde hauptsächlich als Allergensequenzdatenbank für Sequenzvergleiche konzipiert. Die Nutzer können mit 3 Methoden ihre Sequenzen mit der Datenbank abgleichen: eine Standard-Ähnlichkeitssuche mit dem FASTA-Programm [20], eine FASTA-Suche mit allen möglichen 80 Aminosäuren langen Fragmenten der Abfragesequenz sowie die Suche nach identischen, 8 Aminosäuren langen Peptiden. Die letzteren beiden Methoden sind von Aufsichtsbehörden für die Prüfung von Transgenen empfohlen worden [21], trotz bestehender Zweifel an ihrer wissenschaftlichen Validität [18]. Mit *AllergenOnline* wurde das Allergenitätsrisiko der Cry-Proteine, die für *Bacillus thuringiensis*-Toxine codieren und häufig in Nutzpflanzen eingebracht wurden, um ihnen Resistenz gegen Schädlinge zu verleihen [22], ebenso geprüft wie das von Genen, die zur Bestäubungskontrolle in Indischen Senf eingebracht wurden [23]. In beiden Fällen wurden keine signifikanten Ähnlichkeiten mit bekannten Allergenen festgestellt. Darüber hinaus wird *AllergenOnline* von Wissenschaftlern als zuverlässige Quelle von Allergensequenzen in verschiedenen Anwendungsbereichen genutzt, z.B. in der Massenspektrometrie-basierten Quantifizierung bekannter Allergene [24] oder bei der Suche nach Homologen von neu identifizierten Allergenen [25].

Structural Database of Allergenic Proteins

Die *Structural Database of Allergenic Proteins* (SDAP; fermi.utmb.edu/SDAP/) wurde 2002 am Sealy Center for Structural Biology im Institut für Biochemie und Molekularbiologie der University of Texas, Medical Branch, in Galveston, TX, USA eingerichtet [26, 27]. Es handelt sich um eine Datenbank von Allergensequenzen, -strukturen und -epitopen sowie eine Sammlung von Bioinformatik-Tools zur Analyse von Allergenen und ihren Ähnlichkeiten. Diese Datenbank verwendet Allergendaten aus der *WHO/IUIS Allergen Nomenclature Database* und ergänzt sie durch Nicht-IUIS-Allergene aus Sequenz- (UniProt, NCBI, Protein Information Resource), Struktur- (PDB) und Literaturdatenbanken (PubMed). Neue Daten werden vom SDAP-Team eingegeben und von einem externen wissenschaftlichen Beirat geprüft, dem derzeit 3 Wissenschaftler angehören. Es wurden jedoch bisher keine formalen Kriterien veröffentlicht, nach denen Allergene aufgenommen werden. Die SDAP wurde bis 2013 regelmäßig aktualisiert. Mit Stand vom März 2017 wird jedoch auf

der Website als Zeitpunkt der letzten Aktualisierung Februar 2013 angegeben.

Zwei Gruppen von Allergenen sind in SDAP definiert: «SDAP All» enthält die Gesamtheit der Allergene, «SDAP Food» ist auf die Nahrungsmittelallergene beschränkt. Alle Suchfunktionen können bei beiden Datenbankversionen genutzt werden. Die Inhalte der Datenbanken lassen sich entweder als alphabetische Allergenlisten durchblättern oder mittels Textsuche nach Allergenamen, Quellen oder Allergenbeschreibung durchsuchen, auf Wunsch begrenzt auf WHO/IUIS- anerkannte Allergene. Jeder Allergendatensatz enthält den Allergenamen und -typ (z.B. Lebensmittel, Milben, Gräserpollen), den Anerkennungsstatus des WHO/IUIS Allergen Nomenclature Sub-Committee sowie die biochemischen Namen der Allergene (sogenannte Keywords). Darüber hinaus werden der wissenschaftliche und umgangssprachliche Name der Allergenquelle angegeben. Zudem enthält der Datensatz Links zu Sequenzdatenbanken (NCBI, UniProt, PIR) und Proteinfamiliendatenbanken (Superfamily, InterPro, Pfam) sowie die wichtigsten Literaturverweise. Bei Allergenen mit bekannter Struktur ist auch ein Link zum PDB-Eintrag enthalten. Ansonsten werden, wenn homologe Strukturen existieren, automatisch erzeugte Homologiemodelle angeboten [28]. Bei Allergenen mit bekannten Epitopen wird eine Liste der Epitope mit zugehörigen Literaturverweisen angegeben.

Jeder Allergeneintrag enthält Verknüpfungen zu verschiedenen Tools für die Sequenzanalyse, z.B. PeptideCutter zur Vorhersage von Spaltungsstellen durch Proteasen oder Chemikalien [29], Suche nach ähnlichen Sequenzen in SDAP mit FASTA und BLAST, Ähnlichkeitssuche in der NCBI- und der UniProt-Proteindatenbank [30] und Durchsuchung der Prosite-Datenbank auf Vorliegen von Sequenzmotiven, die charakteristisch für bestimmte Proteinfamilien oder -funktionen sind [31]. Von Allergenen, die durch Abgleich mit der Pfam-Datenbank einer Proteinfamilie zugeordnet wurden [32], kann der Nutzer sich eine Liste der SDAP-Allergene dieser Familie anzeigen lassen.

Die Allergensequenzen in der SDAP können mit verschiedenen Tools durchsucht werden. Nutzer können die Ähnlichkeit der von ihnen untersuchten Sequenzen mit bekannten Allergenen über Methoden testen, die von der UN-Organisation für Ernährung und Landwirtschaft (FAO) und der WHO empfohlen werden; mittels einer vollständigen FASTA-Suche, einer FASTA-Suche mit einem gleitenden 80-mer-Fenster oder einer Suche nach kurzen Abschnitten identischer Sequenzen. Darüber hinaus können die Nutzer mit dem Property Distance (PD)-Index [27] Peptidsequenzen (z.B. Epitope) mit Sequenzen in der SDAP vergleichen, um Allergene mit identischen Sequenzen oder Sequenzen mit voraussichtlich ähnlichen physikalisch-chemischen Eigenschaften und möglicher Kreuzreaktivität zu identifizieren. Der Zusammenhang zwischen PD-Index und IgE-Kreuzreaktivität ist experimentell validiert. Hierzu wurden Seren von Patienten getestet, die gegen das Hauptallergen von Wacholderpollen (Jun a 1) allergisch sind [33]. Peptide, die den 3 wichtigsten IgE-Epitopen von Jun a 1 entsprechen, wurden zusammen mit Peptiden unterschiedlicher Sequenzähnlichkeit und somit PD-Indizes im Vergleich zu den Wildtyp-Epitopen auf IgE-Bindung getestet. Die Autoren zeigten eine Assoziation der IgE-Kreuzreaktivität mit der

Sequenzähnlichkeit, ausgedrückt als PD-Index, und zogen die Schlussfolgerung, dass der PD-Index für die Vorhersage von Kreuzreaktivitäten geeignet ist.

Die SDAP wird als Quelle für Allergensequenzen im Rahmen verschiedener Anwendungen in der Bioinformatik genutzt. In einer Studie mit dem Ziel, die Allergenität von Proteinen vorherzusagen, nutzten Kumar und Shelokar [34] Methoden des maschinellen Lernens, um allergene von nichtallergenen Proteinen zu unterscheiden. Ihre Software klassifizierte Proteine anhand von mehreren Sequenzmerkmalen. Mit Allergensequenzen aus der SDAP als Testdatensätze erreichten sie eine Genauigkeit von 90% für die korrekte Vorhersage allergener und nichtallergener Proteine. Über Sequenzvergleiche mit der SDAP ist es möglich, die Allergenität von Proteinen vorherzusagen. In einer Bioinformatik-, biochemischen und immunologischen Analyse untersuchten Sharma und Kollegen die Allergenität von Osmotin, einem Pathogenese-assoziierten Protein aus Tabak, das an der Resistenz gegenüber abiotischem Stress beteiligt ist [35]. Das für Osmotin codierende Gen sollte in transgene Nutzpflanzen eingebracht werden, um deren Stresstoleranz zu erhöhen. Es stellte sich jedoch heraus, dass das Protein erhebliche Sequenzähnlichkeiten mit allergenen Mitgliedern der Familie der Thaumatin-ähnlichen Proteine aufwies und IgE-Bindung bei Seren von Patienten mit Allergien gegen Apfel oder Tomate zeigte.

Immune Epitope Database

Die Immune Epitope Database (IEDB; www.iedb.org) wurde 2004 von einem internationalen Expertenteam eingerichtet und 2006 der Öffentlichkeit zugänglich gemacht. Sie wird durch Fördermittel des National Institute of Allergy and Infectious Diseases in Bethesda, MD, USA, finanziert [36]. Die IEDB ist die wohl umfassendste Quelle von Daten zu experimentell entdeckten Epitopen. Die Datenbank enthält (sowohl lineare als auch konformationelle) Peptid- und Nichtpeptid-Epitope adaptiver Immunrezeptoren (Antikörper, T-Zell-Rezeptoren und Haupthistokompatibilitätskomplex-Proteine) von Menschen und nicht-menschlichen Tieren. Sie umfasst Epitope, die mit Infektionskrankheiten, Allergien, Autoimmunerkrankungen oder Transplantation assoziiert sind. Epitope des humanen Immundefizienzvirus (diese sind in der Los Alamos HIV Molecular Immunology Database enthalten, www.hiv.lanl.gov) und tumorassoziierte Epitope hingegen sind nicht Bestandteil der IEDB.

Die Daten, die in die IEDB Eingang finden, werden entweder von den Wissenschaftlern direkt eingereicht oder – in der Mehrzahl der Fälle – aus der Literatur entnommen, mit einer Kombination aus automatisierten PubMed-Recherchen und manueller Auswahl der resultierenden Publikationen durch ein Team kompetenter Kuratoren nach detaillierten Richtlinien aus einem ausführlichen Kurationshandbuch [37]. Der Kern der IEDB besteht aus Daten zu Experimenten, die die Erkennung eines Epitops oder Antigens durch T-Zell-Rezeptoren, Antikörper oder MHC-Moleküle beschreiben. Alle experimentellen Assays, die das Epitop oder seine Erkennung durch Immunrezeptoren charakterisieren, werden in die IEDB eingegeben, einschließlich aller negativen Ergebnisse. Diese Daten werden mit der Literaturquelle verknüpft, aus der sie entnommen wurden, mit Angabe des Antigens und des Organismus, aus dem das Epitop

stammt, dem Wirtsorganismus, der mit dem Epitop assoziierten Erkrankung und dem Expositions- oder Immunisierungsprozess. Im März 2017 enthielt die IEDB Daten zu knapp 280 000 Epitopen aus 3600 Quellorganismen, gewonnen aus mehr als 18 000 Publikationen.

Das Suchformular auf der IEDB-Website bietet dem Nutzer die am häufigsten verwendeten Suchkriterien an. Alle Epitope in der Datenbank können nach Art des Epitops (linear, konformationell, nicht-peptidisch), Assays (B-Zell-, T-Zell-, MHC-), der MHC-Restriktion, des Antigen und Herkunftsorganismus, Wirtsorganismus sowie der Erkrankung (Infektionskrankheit, Allergie, Autoimmunkrankheit, Transplantation) durchsucht werden. Alle Suchfelder haben eine automatische Vervollständigungsfunktion, um die schnelle Identifizierung potenziell zweideutiger Namen, z.B. Organismenamen oder Antigene, zu erleichtern. Die Suchergebnisse werden in 4 Tabellen als Listen von Epitopen, Antigenen, Assays und Referenzen angezeigt. Die Suche kann verfeinert werden, indem zusätzliche Kriterien hinzugefügt oder vorhandene entfernt werden. Die anfänglichen Suchkriterien können weiter eingengt werden. Der Nutzer kann beispielsweise die genaue Art des Assays (z.B. Suche nur nach ELISA-Ergebnissen oder 3D-Strukturen von Antikörper-Antigen-Komplexen) angeben, die genaue MHC-Restriktion oder eine stärker eingeschränkte Krankheitskategorie (z.B. Asthma anstelle von allergischen Erkrankungen). Die Suchfelder sind mit einem Finder-Tool verknüpft, das eine Liste hierarchisch klassifizierter Suchbegriffe zur Auswahl anzeigt (z.B. 3D-Struktur hat Unterkategorien wie Kristallographie oder NMR). Suchergebnisse für spezifische Antigene lassen sich mit dem Immunone-Browser-Tool analysieren. Das Tool erzeugt einen Plot mit der Anzahl positiver und negativer Assays sowie der Häufigkeit positiver Reaktionen jeder Aminosäure in der Antigensequenz. Zur Abbildung der Lokalisation der Epitope dient das Epitope-Viewer-Tool, eine Java-Anwendung zur Darstellung von 3D-Strukturen. Die IEDB-Website bietet außerdem noch verschiedene Tools für die Vorhersage und Analyse von Epitopen an, die auch auf Suchergebnisse angewendet werden können.

Die Anwendungsmöglichkeiten der IEDB in der Allergieforschung wurden in einer Übersichtsarbeit beschrieben [38]. Mit Stand vom März 2017 enthielt die IEDB Daten zu mehr als 8000 Peptid-Epitopen von mehr als 300 Allergenen. Epitopdaten sind von entscheidender Bedeutung für verschiedene Bereiche der Allergieforschung und der Entwicklung klinischer Anwendungen. Die epitopbasierte Diagnose könnte genauere Ergebnisse hinsichtlich der Korrelation mit der Schwere der Symptome [39], der Vorhersage klinischer Kreuzreaktivität [40] und der Prognose der Persistenz bzw. des Verschwindens von Nahrungsmittelallergien bei Kindern [41] liefern. Peptidbasierte allergenspezifische Immuntherapien können mit immundominanten T-Zell-Epitopen [42, 43], nicht-IgE-bindenden B-Zell-Epitopen [44] oder Mimotopen – Peptidprodukten, die konformationelle IgE-Epitope nachahmen [45] – als Vakzinen durchgeführt werden. Die Kenntnis der IgE-Epitope ist auch der Schlüssel zur Entwicklung hypoallergener Allergenderivate [46–48]. Die Entwicklung von IgG- und IgE-Epitop-Spezifitäten im Behandlungsverlauf zu beobachten kann außerdem helfen, das Ergebnis einer klassischen extraktbasierten Immuntherapie vorherzusagen [49, 50]. Die Charakterisierung des Epitopprofils

einer spezifischen Allergenquelle, die von T-Zellen erkannt wird, kann zudem dazu beitragen, die Dynamik der allergischen Immunantwort besser zu verstehen und Biomarker oder Therapieziele für eine spezifische Immuntherapie zu identifizieren. Mit einem immunproteomischen Forschungsansatz haben Schulten et al. [51] bei Patienten mit Gräserpollenallergie einen überraschend hohen Anteil an T_H2-Zellen gefunden, die Peptide nicht-IgE-bindender Wiesenlieschgras-Proteine erkannten. Darüber hinaus korrelierte die Erkennung einiger dieser neuartigen T-Zell-Antigene mit den Behandlungsergebnissen der gräserpollenspezifischen Immuntherapie.

AllFam – Database of Allergen Families

Nach der Identifizierung von Hunderten von Allergenen aus Dutzenden von Quellen wurde offensichtlich, dass viele Allergene evolutionär miteinander verwandt sind und Ähnlichkeiten in den Sequenzen, Strukturen, physikalisch-chemischen Eigenschaften und biochemischen Funktionen aufweisen. Der Fortschritt in der Bioinformatik ermöglichte die Klassifizierung der Allergene in Proteinfamilien. Verschiedene Wissenschaftler unternahmen dies mit unterschiedlichen Methoden, gelangten aber zu ähnlichen Ergebnissen [12, 15, 52–54]. Es stellte sich heraus, dass die meisten Allergene einer überraschend kleinen Anzahl von Proteinfamilien angehören. Dabei darf man jedoch nicht vergessen, dass auch in Proteinfamilien, die Allergene enthalten, die meisten Mitglieder nicht-allergen sind. Die Nachfrage vieler Allergieforscher und Kliniker nach einer einfachen Schnittstelle zur Proteinfamilienklassifikation der Allergene nahm der Autor des vorliegenden Artikels zum Anlass und richtete 2007 die AllFam-Datenbank der Allergenfamilien ein (www.meduniwien.ac.at/allfam/ [55]).

AllFam basierte ursprünglich auf den Allergendaten aus *Allergome*. 2016 erfolgte ein großer Relaunch der Datenbank. Seitdem bezieht AllFam seine Allergendaten aus der WHO/IUIS Allergen Nomenclature Database, ergänzt durch Nicht-IUIS-Einträge aus *AllergenOnline*. Die Allergensequenzen werden den Proteinfamilien zugeordnet, indem sie mit Familiendefinitionen aus der Pfam-Datenbank (pfam.xfam.org) abgeglichen werden. Pfam ist eine der größten Proteinfamilien-Datenbanken; Version 30.0 vom Juni 2016 enthält 16 306 Einträge [32]. Die Pfam-Familien basieren auf manuell kuratierten Sequenzvergleichen repräsentativer Familienmitglieder.

Die AllFam-Version vom 7. März 2017 ordnet die Allergene in 151 AllFam-Familien ein, die auf 216 Pfam-Domänen basieren. Die AllFam-Familien werden definiert, indem mehrere Pfam-Domänen so zusammengeführt werden, dass jedes Allergen Mitglied einer einzigen AllFam-Familie ist und dass alle verwandten Allergene derselben Familie zugeordnet sind. Dies gilt für die einzelnen Pfam-Domänen von Multidomänen-Proteinen sowie Gruppen von Pfam-Familien, die hochgradig divergente Proteinfamilien repräsentieren, die nicht von einer einzigen Pfam-Familie abgedeckt werden können.

Die Benutzeroberfläche von AllFam hat eine Browsing-Funktion zum Durchsuchen der Datenbank über die Auflistung der Allergenfamilien gefiltert nach Gruppen von Allergenquellen (Pflanzen, Tiere, Pilze, Bakterien), Expositionsweg (Inhalation, Ingestion usw.) und Anerkennung durch das WHO/IUIS Allergen Nomenclature Sub-Committee. Mit den gleichen Optionen lassen sich die allergenen Mitglieder jeder Familie auflisten und filtern. Darüber hinaus

kann man die Datenbank nach Allergennamen, Quellen, Namen von Proteinfamilien, Proteinfamilien-Zugangsnummern (sowohl aus *AllFam* als auch aus *Pfam*) und Schlüsselwörtern durchsuchen. Jedes Allergen ist mit den entsprechenden Datensätzen in der *WHO/IUIS*-Datenbank und in *AllergenOnline* verknüpft. Zu jeder *AllFam*-Familie gibt es eine Familienbeschreibung mit kurzen Zusammenfassungen der biochemischen Eigenschaften und der allergologischen Bedeutung ihrer allergenen Mitglieder, einer Liste ausgewählter wichtiger Literaturverweise und Verknüpfungen zu den entsprechenden *Pfam*-Datensätzen und Wikipedia-Einträgen.

Die Klassifikation von Proteinfamilien in *AllFam* kann zur Klärung der Frage herangezogen werden, welche Faktoren für die Allergenität bestimmter Proteine verantwortlich sind. Wir verglichen die Anzahl der *Pfam*-Domänen in Allergenen mit einer gleich großen Stichprobe zufällig ausgewählter Sequenzen und konnten zeigen, dass die Zahl der Proteinfamilien, die Allergene enthielten, signifikant geringer war als die in den zufälligen Stichproben [55]. Das sprach für die Hypothese, dass Allergene besondere Merkmale aufweisen und nicht jedes Protein zum Allergen werden kann. Darüber hinaus ergab eine biochemische Klassifikation von Allergenen, die durch Verknüpfung von *Pfam*-Familien mit der *Gene Ontology Database* erfolgte [56, 57], eine noch enger begrenzte Verteilung der Allergene. Der Großteil der Allergene waren Hydrolasen (Proteasen, Lipasen und Polysaccharidhydrolasen), metall- oder lipidbindende Proteine, Speicherproteine oder Bestandteile des Zytoskeletts [55]. *AllFam* wurde auch verwendet, um Sequenzähnlichkeiten von Allergenen und Parasitenantigenen zu analysieren. Tyagi et al. [58] verglichen die Verteilung der *Pfam*-Proteinfamilien von Allergenen und Proteinen aus Metazoan-Parasiten. Sie stellten fest, dass fast die Hälfte der mehr als 2000 analysierten Parasitenproteine Mitglieder der 10 wichtigsten Allergenfamilien laut *AllFam* waren. Die Autoren gelangten zu der Schlussfolgerung, dass die IgE-Antwort, die sich unter dem Selektionsdruck des Parasitenbefalls entwickelt hat, sich möglicherweise auf bestimmte nicht-parasitäre Proteine ausweiten kann, wenn diese einem Parasiten-Antigen ähnlich sind. In einer Folgestudie überprüfte dieselbe Arbeitsgruppe diese Hypothese, indem sie 13 rekombinante Proteine aus *Schistosoma mansoni* auf die Bindung von Antikörpern aus dem Serum von *S. mansoni*-infizierten Patienten testete [59]. Alle 9 Proteine, bei denen aufgrund ihrer *AllFam*-Allergenfamilienzugehörigkeit Antigenität prognostiziert worden war, zeigten IgE-Bindung in mindestens 10% der getesteten Seren. Hingegen zeigten nur 2 der 5 getesteten häufigen *S.-mansoni*-Proteine ohne allergene Homologe eine IgE-Bindung. Bemerkenswerterweise wurden Homologe eines dieser IgE-bindenden Antigene, Aldolasen, kürzlich als kreuzreaktive Allergene in Fisch und Hühnerfleisch beschrieben [60].

Schlussbemerkungen

Die 6 Datenbanken, die in der vorliegenden Übersichtsarbeit vorgestellt werden, sind die meistgenutzten Ressourcen für Allergendaten. Die *WHO/IUIS*-Datenbank ist die Quelle der offiziellen Allergenomenklatur. Ein großer Teil der in den anderen Datenbanken erfassten Allergene stammt aus der *WHO/IUIS*-Datenbank. *Allergome* ist die umfassendste Datenbank; sie enthält Verknüpfungen zu vielen anderen Ressourcen sowie leistungsstarke Funktionen für Literaturrecherchen. Bei *Allergome* gelten keine strengen Kriterien für die Aufnahme von Allergenen; hier ist es also am Nutzer, die allergene Bedeutung der in der Datenbank enthaltenen Proteine zu beurteilen. *AllergenOnline* ist als Peer-Review-Allergensequenz-Datenbank angelegt, anhand derer sich die Allergenität von Proteinen prognostizieren lässt. Der Fokus der *SDAP* liegt auf Allergenstrukturen und Epitopen; die Datenbank enthält Verknüpfungen zu vielen anderen Datenbanken und Bioinformatik-Tools. Allerdings scheint die *SDAP* seit 2013 nicht mehr aktualisiert zu werden. Die *IEDB* ist die größte Sammlung von Epitopdaten aus allen Bereichen der Immunologie, einschließlich der Allergologie. *Allfam* bietet eine einfache Schnittstelle zum Durchsuchen der Proteinfamilien-Klassifizierung von Allergenen.

Zusätzlich zu den hier vorgestellten Ressourcen stehen noch weitere, weniger gebräuchliche Datenbanken zur Verfügung, z.B. die *Informall Food Allergen Database* (research.bmh.manchester.ac.uk/informall/allergenic-foods/ [61]), die *Allergen Database for Food Safety* (allergen.nihs.go.jp/ADFS/) und *AllAllergy* (allallergy.net). Weitere Datenbanken wurden in der Vergangenheit eingerichtet, existieren aber heute nicht mehr. Die noch vorhandenen Datenbanken werden von der Fachwelt rege genutzt und sehr geschätzt, sind aber auch mit verschiedenen Herausforderungen konfrontiert. Angesichts der ständig wachsenden Zahl neuer Publikationen zu Allergenen wird es immer aufwendiger, die Datenbanken zeitnah auf dem aktuellen Stand zu halten und dabei eine angemessene Datenqualität zu gewährleisten. Die Betreiber der Datenbanken werden künftig höchstwahrscheinlich stärker auf automatisiertes Data-Mining und Text-Mining-Software zurückgreifen. Die manuelle Kuratierung durch Experten für molekulare Allergologie wird jedoch auch weiterhin unverzichtbar sein. Darüber hinaus müssen mehr Ressourcen für technische Fachleute aufgebracht werden, um steigenden Anforderungen an die Internetsicherheit gerecht zu werden und intuitive Benutzeroberflächen zu erstellen, die es auch Wissenschaftlern mit wenig Erfahrung in der Bioinformatik ermöglichen, aus der stetig wachsenden Datenmenge gezielt bestimmte Informationen abzurufen.

Literatur

- 1 Marsh DG, Goodfriend L, King TP, et al.: Allergen nomenclature. Bull World Health Organ 1986;64:767-774.
- 2 King TP, Hoffman D, Lowenstein H, Marsh DG, et al.: Allergen nomenclature. Int Arch Allergy Immunol 1994;105:224-233.
- 3 King TP, Hoffman D, Lowenstein H, et al.: Allergen nomenclature. Bull World Health Organ 1994;72:797-806.
- 4 King TP, Hoffman D, Lowenstein H, et al.: Allergen nomenclature. Clin Exp Allergy 1995;25:27-37.
- 5 King TP, Hoffman D, Lowenstein H, et al.: Allergen nomenclature. Allergy 1995;50:765-774.
- 6 King TP, Hoffman D, Lowenstein H, et al.: Allergen Nomenclature. J Allergy Clin Immunol 1995;96:5-14.

- 7 Chapman MD, Pomes A, Breiteneder H, Ferreira F: Nomenclature and structural biology of allergens. *J Allergy Clin Immunol* 2007;119:414-420.
- 8 Larsen JN: Allergen nomenclature: a need for the scientific community. *Arb Paul Ehrlich Inst Bundesamt Sera Impfstoffe Frankf A M* 2006;95: 5-9.
- 9 Radauer C, Nandy A, Ferreira F, et al.: Update of the WHO/IUIS Allergen Nomenclature Database based on analysis of allergen sequences. *Allergy* 2014;69:413-419.
- 10 Mari A, Rasi C, Palazzo P, Scala E: Allergen databases: current status and perspectives. *Curr Allergy Asthma Rep* 2009;9:376-383.
- 11 Mari A, Scala E, Palazzo P, et al.: Bioinformatics applied to allergy: allergen databases, from collecting sequence information to data integration. The Allergome platform as a model. *Cell Immunol* 2006;244:97-100.
- 12 Stadler MB, Stadler BM: Allergenicity prediction by protein sequence. *FASEB J* 2003;17:1141-1143.
- 13 Dimitrov I, Doytchinova I: Associations between milk and egg allergens and the HLA-DRB1/DQ polymorphism: a bioinformatics approach. *Int Arch Allergy Immunol* 2016;169:33-39.
- 14 Campbell BC, Gilding EK, Timbrell V, et al.: Total transcriptome, proteome, and allergome of Johnson grass pollen, which is important for allergic rhinitis in subtropical regions. *J Allergy Clin Immunol* 2015;135:133-142.
- 15 Radauer C, Breiteneder H: Pollen allergens are restricted to few protein families and show distinct patterns of species distribution. *J Allergy Clin Immunol* 2006;117:141-147.
- 16 D'Amato G, Holgate ST, Pawankar R, et al.: Meteorological conditions, climate change, new emerging factors, and asthma and related allergic disorders. A statement of the World Allergy Organization. *World Allergy Organ J* 2015;8:25.
- 17 Nordlee JA, Taylor SL, Townsend JA, et al.: Identification of a Brazil-nut allergen in transgenic soybeans. *N Engl J Med* 1996;334:688-692.
- 18 Goodman RE, Vieths S, Sampson HA, et al.: Allergenicity assessment of genetically modified crops – what makes sense? *Nat Biotechnol* 2008; 26:73-81.
- 19 Goodman RE, Ebisawa M, Ferreira F, et al.: AllergenOnline: A peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. *Mol Nutr Food Res* 2016;60: 1183-1198.
- 20 Pearson WR: Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol Biol* 2000;132:185-219.
- 21 Codex Alimentarius Commission: Alinorm 03/34: Appendix III. Guideline for the Conduct of Food Safety Assessment of Foods Derived from Recombinant DNA Plants. Annex IV. Annex on the Assessment of Possible Allergenicity. Rome, FAO/WHO, 2003.
- 22 Randhawa GJ, Singh M, Grover M: Bioinformatic analysis for allergenicity assessment of *Bacillus thuringiensis* Cry proteins expressed in insect-resistant food crops. *Food Chem Toxicol* 2011;49: 356-362.
- 23 Siruguri V, Bharatraj DK, Vankudavath RN, et al.: Evaluation of Bar, Barnase, and Barstar recombinant proteins expressed in genetically engineered *Brassica juncea* (Indian mustard) for potential risks of food allergy using bioinformatics and literature searches. *Food Chem Toxicol* 2015;83: 93-102.
- 24 Uvackova L, Skultety L, Bekesova S, et al.: MSE based multiplex protein analysis quantified important allergenic proteins and detected relevant peptides carrying known epitopes in wheat grain extracts. *J Proteome Res* 2013;12:4862-4869.
- 25 Zhao X, Li L, Kuang Z, et al.: Proteomic and immunological identification of two new allergens from silkworm (*Bombyx mori* L.) pupae. *Cent Eur J Immunol* 2015;40:30-34.
- 26 Ivanciuc O, Schein CH, Braun W: Data mining of sequences and 3D structures of allergenic proteins. *Bioinformatics* 2002;18:1358-1364.
- 27 Ivanciuc O, Schein CH, Braun W: SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Res* 2003;31:359-362.
- 28 Oezguen N, Zhou B, Negi SS, et al.: Comprehensive 3D-modeling of allergenic proteins and amino acid composition of potential conformational IgE epitopes. *Mol Immunol* 2008;45: 3740-3747.
- 29 Gasteiger E, Hoogland C, Gattiker A, et al.: Protein identification and analysis tools on the ExPASy server; in Walker JM (ed): *The Proteomics Protocols Handbook*. Totowa, Humana Press, 2005, pp 571-607.
- 30 Pearson WR: BLAST and FASTA similarity searching for multiple sequence alignment. *Methods Mol Biol* 2014;1079:75-101.
- 31 Sigrist CJ, de Castro E, Cerutti L, et al.: New and continuing developments at PROSITE. *Nucleic Acids Res* 2013;41:D344-D347.
- 32 Finn RD, Coghill P, Eberhardt RY, et al.: The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 2016; 44:D279-D285.
- 33 Ivanciuc O, Midoro-Horiuti T, Schein CH, et al.: The property distance index PD predicts peptides that cross-react with IgE antibodies. *Mol Immunol* 2009;46:873-883.
- 34 Kumar KK, Shelokar PS: An SVM method using evolutionary information for the identification of allergenic proteins. *Bioinformation* 2008;2: 253-256.
- 35 Sharma P, Singh AK, Singh BP, et al.: Allergenicity assessment of osmotin, a pathogenesis-related protein, used for transgenic crops. *J Agric Food Chem* 2011;59:9990-9995.
- 36 Vita R, Overton JA, Greenbaum JA, et al.: The immune epitope database (IEDB) 3.0. *Nucleic Acids Res* 2015;43:D405-D412.
- 37 Vita R, Zarebski L, Greenbaum JA, et al.: The immune epitope database 2.0. *Nucleic Acids Res* 2010;38:D854-D862.
- 38 Vaughan K, Peters B, Larche M, et al.: Strategies to query and display allergy-derived epitope data from the immune epitope database. *Int Arch Allergy Immunol* 2013;160:334-345.
- 39 Flinterman AE, Knol EF, Lencer DA, et al.: Peanut epitopes for IgE and IgG4 in peanut-sensitized children in relation to severity of peanut allergy. *J Allergy Clin Immunol* 2008;121:737-743.
- 40 Hazebrouck S, Ah-Leung S, Bidat E, et al.: Goat's milk allergy without cow's milk allergy: suppression of non-cross-reactive epitopes on caprine beta-casein. *Clin Exp Allergy* 2014;44:602-610.
- 41 Savilahti EM, Rantanen V, Lin JS, et al.: Early recovery from cow's milk allergy is associated with decreasing IgE and increasing IgG4 binding to cow's milk epitopes. *J Allergy Clin Immunol* 2010;125:1315-1321.
- 42 Couroux P, Patel D, Armstrong K, et al.: Fel d 1-derived synthetic peptide immuno-regulatory epitopes show a long-term treatment effect in cat allergic subjects. *Clin Exp Allergy* 2015;45: 974-981.
- 43 Prickett SR, Rolland JM, O'Hehir RE: Immuno-regulatory T cell epitope peptides: the new frontier in allergy therapy. *Clin Exp Allergy* 2015;45: 1015-1026.
- 44 Focke-Tejkl M, Weber M, Niespodziana K, et al.: Development and characterization of a recombinant, hypoallergenic, peptide-based vaccine for grass pollen allergy. *J Allergy Clin Immunol* 2015; 135:1207-1207.
- 45 Luzar J, Strukelj B, Lunder M: Phage display peptide libraries in molecular allergology: from epitope mapping to mimotope-based immunotherapy. *Allergy* 2016;71:1526-1532.
- 46 Zuidmeer-Jongejan L, Huber H, Swoboda I, et al.: Development of a hypoallergenic recombinant parvalbumin for first-in-man subcutaneous immunotherapy of fish allergy. *Int Arch Allergy Immunol* 2015;166:41-51.
- 47 Ueno HM, Kato T, Ohnishi H, et al.: T cell epitope-containing hypoallergenic beta-lactoglobulin for oral immunotherapy in milk allergy. *Pediatr Allergy Immunol* 2016;27:818-824.
- 48 Jutel M, Akdis CA: Novel immunotherapy vaccine development. *Curr Opin Allergy Clin Immunol* 2014;14:557-563.
- 49 Vickery BP, Lin J, Kulis M, et al.: Peanut oral immunotherapy modifies IgE and IgG4 responses to major peanut allergens. *J Allergy Clin Immunol* 2013;131:128-134.
- 50 Gepp B, Lengger N, Möbs C, et al.: Monitoring the epitope recognition profiles of IgE, IgG1, and IgG4 during birch pollen immunotherapy. *J Allergy Clin Immunol* 2016;137:1600-1603.
- 51 Schulten V, Peters B, Sette A: New strategies for allergen T cell epitope identification: going beyond IgE. *Int Arch Allergy Immunol* 2014;165: 75-82.
- 52 Jenkins JA, Griffiths-Jones S, Shewry PR, et al.: Structural relatedness of plant food allergens with specific reference to cross-reactive allergens: an in silico analysis. *J Allergy Clin Immunol* 2005;115: 163-170.
- 53 Jenkins JA, Breiteneder H, Mills EN: Evolutionary distance from human homologs reflects allergenicity of animal food proteins. *J Allergy Clin Immunol* 2007;120:1399-1405.
- 54 Ivanciuc O, Garcia T, Torres M, et al.: Characteristic motifs for families of allergenic proteins. *Mol Immunol* 2009;46:559-568.
- 55 Radauer C, Bublin M, Wagner S, et al.: Allergens are distributed into few protein families and possess a restricted number of biochemical functions. *J Allergy Clin Immunol* 2008;121:847-852.
- 56 Ashburner M, Ball CA, Blake JA, et al.: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000;25: 25-29.
- 57 Gene Ontology Consortium: Gene Ontology Consortium: going forward. *Nucleic Acids Res* 2015;43:D1049-D1056.
- 58 Tyagi N, Farnell EJ, Fitzsimmons CM, et al.: Comparisons of allergenic and metazoan parasite proteins: allergy the price of immunity. *PLoS Comput Biol* 2015;11:e1004546.
- 59 Farnell EJ, Tyagi N, Ryan S, et al.: Known allergen structures predict *Schistosoma mansoni* IgE-binding antigens in human infection. *Front Immunol* 2015;6:26.
- 60 Kuehn A, Codreanu-Morel F, Lehnert-Weber C, et al.: Cross-reactivity to fish and chicken meat - a new clinical syndrome. *Allergy* 2016;71: 1772-1781.
- 61 Mills EN, Jenkins JA, Sancho AI, et al.: Food allergy information resources for consumers, industry and regulators. *Arb Paul Ehrlich Inst Bundesamt Sera Impfstoffe Frankf A M* 2006;95:17-25.