

Proteomic analysis in the identification of allergenic molecules

Expert Rev. Proteomics 7(5), 723–734 (2010)

Adriano Mari^{†1},
Maria Antonietta
Ciardiello²,
Maurizio Tamburrini²,
Chiara Rasi¹ and
Paola Palazzo¹

¹Center for Molecular Allergology,
IDI-IRCCS, Via dei Monti di Creta 104,
I-00167 Roma, Italy

²Institute of Protein Biochemistry,
C.N.R., Napoli, Italy

[†]Author for correspondence:

Tel.: +39 338 883 7352
a.mari@panservice.it

Conventional and innovative strategies can be exploited to identify and characterize new allergenic proteins. With the aim of obtaining suggestions for future improvements, this article describes our attempt to understand and describe some of the advantages and pitfalls of the methodologies and procedures often used in this field. The analysis includes the protein extract preparation, starting from the allergenic source, the separation of the proteins contained in a mixture and the detection, identification and characterization of IgE-binding molecules. Classic and emerging proteomic technologies, including mass spectrometry-based methodologies, Edman degradation procedure, microarray-based techniques and bioinformatics search strategies, have been explored. A comparative analysis of biochemistry-based proteomics and molecular biology strategies has also been given.

KEYWORDS: allergens • bioinformatics • Edman degradation • mass spectrometry • microarrayed allergens • molecular biology • protein extracts • proteomics

A high number of allergenic molecules have so far been identified in different allergenic sources. However, several indications, including clinical and immunological data, suggest that many molecules able to induce an allergic reaction have not yet been identified. During the last few decades, the traditional biochemical and immunological methodologies, which have been used since the beginning of molecular allergology, have received increasing support and improvement due to their integration with different methodologies, such as molecular biology, bioinformatics and emerging proteomic technologies. Nevertheless, whatever procedure is used to identify new allergenic molecules, each one demonstrates pitfalls at different levels, indicating that further efforts are necessary to overcome some critical points.

Current application of proteomics into the allergenic molecule characterization field

Generally, the isolation and identification of a potential allergenic molecule of a specific organism or tissue includes the extraction of the protein components from the allergenic source, the separation of the extracted proteins and the detection and identification of IgE-binding proteins.

Preparation of protein extracts

The quality of the protein extract is of utmost importance, especially if we have the aim of analyzing as many proteins/allergens as possible coming from the source under investigation. Ideally, the extract should contain all the protein components present in the starting material. However, it is well known that, depending on the characteristics of the starting material and on the protocol used, the composition of extracts may be very variable [1–3]. Very often, several extraction protocols have to be tested in order to set up the most suitable one for the aim. As an example, the search for an effective method to obtain high-quality protein extracts from *Dermatophagoides pteronyssinus* is still ongoing, despite this allergenic source having been studied for almost 50 years, because, depending on the composition of the extraction buffer, different patterns of protein components are obtained [4].

The preparation of high-quality protein extracts from some plant tissues, such as fruits, is typically more problematic than from other organisms, and is, therefore, still a challenge for researchers [5]. Plant tissues have a relatively low protein content compared with bacterial or animal tissues and, in addition, they have significant amounts of interfering substances, such as pigments, carbohydrates, polyphenols,

polysaccharides and starch, that may strongly affect, in a negative way, the extract composition and the subsequent steps of separation and analysis of the protein components [5–7]. For instance, it has been reported that the extraction of allergens from grape is hampered by low protein and high tannin and pectin concentrations. A satisfactory extraction of allergenic proteins from grape, mainly lipid transfer protein (LTP; Vit v 1), was obtained only by changing the available standard protocols and combining different factors [8].

Differences in allergen composition in olive cultivars have been reported as responsible for the important differences in the allergenic potency of the extracts [9]. Additionally, following a comparative analysis of the protein and allergen profiles, a significant influence of conditions, such as the ripening stage, postharvest handling and the extraction method, on the number and concentration of protein components in green and gold kiwi fruit extracts has been described [10]. It was necessary to deviate from standard protocols to optimize the extraction of LTP (Ara h 9) from peanut. In fact, only by choosing acidic extraction conditions was the concentration of Ara h 9 in the soluble extract increased dramatically, thus allowing its isolation and characterization [11]. The extraction of the allergens Mor n 3 and Pru p 3 from mulberry and peach, respectively, was obtained only by using high ionic strength conditions [12]. Variability in the composition of extracts from allergenic sources as a function of the extraction protocol and/or on the characteristics of the starting material has also been reported for sources such as hen's eggs [13], hazelnut [14], birch pollen [15], peanut [16], apple [2] and cherry [17]. In fact, literature reports underline that no definitive extraction protocol is actually available and suited for a large variety of allergen sources and that, instead, specific protocols need to be set up and optimized depending on the starting available material.

Separation of the extracted proteins

The protein components of the extracts can be separated using different techniques. The choice may depend on the researcher's preference, on the type of the subsequent analyses that have been planned to identify the proteins and also on the characteristics of the extract. For this aim, SDS-PAGE, 2D electrophoresis (2-DE) and liquid chromatography (LC) are widely used methods.

Protein separation by 1D electrophoresis

SDS-PAGE is a very easy and widespread technique used to obtain a fast separation of a protein mixture [18]. Furthermore, 1D electrophoresis offers the advantage that virtually all proteins are soluble in SDS and are also extremely acidic, and basic proteins are easily visualized. However, very often, the samples cannot be loaded on the gel as they are, especially if the protein concentration is low or if the solution contains buffers or other molecules that can interfere with the migration of the protein components. Therefore, the precipitation of proteins by the addition of trichloroacetic acid (TCA) is frequently performed before the electrophoretic separation as it is a very useful procedure to concentrate diluted proteins. Unfortunately, some proteins can be lost during the treatment with TCA. For instance, we have observed that LTP from peach (Pru p 3) and mulberry (Mor n 3) disappeared from the protein extract following precipitation with TCA. In line with our results,

LTP from strawberry (Fra a 3), cultivar Rosana, was not detected in the extract treated with TCA, whereas it was observed in the sample treated with urea and dithiothreitol (DTT) [19]. By contrast, the same authors observed that the treatment with TCA and a reducing agent allowed the detection of Fra a 3 in the extract obtained from a different strawberry cultivar (Elan).

Protein separation by 2D electrophoresis

2D electrophoresis is a technique displaying a higher resolution power in protein separation [20] than 1D-SDS-PAGE. Therefore, it can provide much more information on the protein composition of the extract under investigation. For instance, sets of proteins with similar structural properties, including isoforms of allergens with very similar molecular masses, but that differ in their isoelectric points (pI), can be separated by 2-DE, allowing the study of possible differential IgE-binding capacities, as reported for allergens such as fish parvalbumin [21], Bet v 1 [22], peanut allergens [23,24] and allergens from oak (Que a 1) and hornbeam (Car b 1) [25]. 2-DE is a very useful technique for comparative studies of the proteome variation in a tissue as a function of different factors. As far as allergens are concerned, it can be of significant importance to analyze, for example, different cultivars of a plant food with the aim of identifying those containing fewer (or more) allergenic components [26]. Furthermore, this technique could be useful to study the influence of environmental factors on the protein/allergen content of allergen sources [27]. Although the protein separation by 2-DE is a potential powerful tool for this kind of investigation, we can find in the available literature few studies focused on this topic. This may be due to several factors and, at least partially, to the variable reproducibility of the 2-DE protein profiles. Actually, there are still various aspects of the technique that need to be improved, including sample preparation, the resolution and detection of protein components and their identification by a correct use of bioinformatics tools [28], since the results are still dependent on the extract quality and the researcher's skill. Among the major bottlenecks is the protein solubility of the extracts, which is very often affected by the presence of a wide variety of nonprotein substances, which may also interfere with the resolution of the electrophoretic separation and/or the visualization of the results.

Sample preparation is a critical step in 2-DE, especially when the extract is derived from a plant source [29,30]. For instance, fruits are among the worst tissues for proteomic studies because it is often difficult to obtain high-quality protein extracts suitable for 2-DE analysis [5,31]. Another well-known disadvantage of the standard 2-DE technique is the loss of very hydrophobic proteins and proteins with extreme pI values. Nevertheless, during the last few years, the protocols of 2-DE with immobilized pH gradients have constantly been refined, bringing significant improvements in this direction [32,33].

Protein separation by liquid chromatography

Liquid chromatography is a very dynamic methodology that, depending on the type of chromatography selected, can separate the protein components as a function of properties such as the molecular mass, pI, hydrophobicity and capacity to bind a

ligand. LC allows good separations of acidic and basic, low- and high-molecular-weight proteins in a very short analysis time. By choosing different systems, either the separation of high amounts of proteins (preparative chromatography) or high resolution using chromatographic systems working at high pressure (fast-protein LC [FPLC] and high-pressure LC [HPLC]) can be obtained. Furthermore, a better resolution in the separation of a protein mixture can be obtained by combining different types of chromatography (multidimensional LC [MDLC]). The protein profiles are generally very reproducible, thus allowing comparative studies of the proteome variations in tissues and in allergen sources. Except for size-exclusion chromatography, generally, LC offers a high flexibility regarding the volume of the sample to be treated and usually does not require concentration/precipitation steps that are sometimes a cause of protein component loosening. For example, most of the characterized kiwi fruit allergens, including actinidin, thaumatin-like protein and kiwellin, were separated by chromatographic methods and then identified and characterized. Actinidin (Act d 1) was isolated by the combination of HPLC and anion-exchange chromatography [34]. Thaumatin-like protein (Act d 2) was isolated using a combination of gel-permeation, ion-exchange and immobilized-metal-ion affinity chromatography [35], and kiwellin was purified by ion-exchange chromatography [36]. Among the most recent literature, three major allergens were isolated from watermelon by combining several chromatographic steps, and were then identified as malate dehydrogenase, triose phosphate isomerase and profilin [37]. Mor n 3 is a novel allergenic LTP identified in black mulberry following the separation by ion-exchange and reversed-phase HPLC (RP-HPLC) [12].

Detection & identification of IgE-binding proteins

Once the components of a protein extract have been separated, whatever technique is used for the separation, IgE-binding proteins can be detected using different techniques, such as immunoblotting, dot blotting, ELISA [38] and microarray-based immunoassays [39,40]. Sometimes, the technique for IgE-binding detection is dependent on the type of protein separation used. The protein components separated by 1D or 2D electrophoresis can be tested only by immunoblotting, following the electro-transfer of molecules onto a nitrocellulose or polyvinylidene fluoride membrane. IgE-immunoblotting of proteins separated by SDS-PAGE is probably the most widespread method used for the detection of IgE-binding proteins. It is easier, faster and requires a lower amount of serum than the IgE immunoblotting of proteins separated by 2-DE. In contrast to SDS-PAGE and 2-DE, the IgE-binding properties of protein components separated by LC can be tested by any detection method. In this case, since the protein component is in solution, it is suitable for immobilization on any support, for in solution tests and for any electrophoretic separation. In addition, proteins coming from LC are usually in a native form and, therefore, they can be used as they are or denatured.

Identification of IgE-binding proteins can be obtained in different ways, including the use of monoclonal or polyclonal IgG antibodies, by testing an enzyme activity and by evaluating the

interaction with a specific ligand. However, to identify a protein without any ambiguity, it is important to have at least partial data about the amino acid sequence of the molecule. Sequence data can be obtained by mass spectrometry technologies or by Edman degradation using an automated sequencer.

A collection of mass spectrometry-based methodologies and strategies can be used to obtain information useful for the identification of a protein [41–44]. Recent advances also allow the simultaneous identification and quantification of traces of allergenic proteins in complex mixtures, such as processed food [45]. The most common approach to identify a protein, sometimes known as ‘bottom-up proteomics’, is based on the generation and analysis of a ‘peptide mass fingerprinting’ obtained by digestion of the protein with a specific protease, usually trypsin, separation of the fragments and determination of their masses by the appropriate mass spectrometer. The mass pattern is unique for each protein and is used to search in specific protein-sequence databases with appropriate software. The masses of the peptides of the unknown protein are compared with the theoretical peptide masses of each protein present in the database and the results are statistically analyzed to find the best match. A disadvantage of this method is that at least an homologous sequence of the unknown protein has to already be present in the database, otherwise it will not be found. If the information obtained with peptide mass fingerprinting does not allow unequivocal identification of the protein, its peptides can be analyzed by tandem mass spectrometry for *de novo* sequencing [46]. This methodology allows the collection of further information on the protein sequence, although it is still an emerging technology needing further improvements, including the implementation of the computational methods used for the analysis of data generated by the mass spectrometer [47]. Among the limitations of sequencing by mass spectrometry is the inability to discriminate between amino acids with identical masses, such as leucine and isoleucine. Nevertheless, significant efforts aimed at overcoming such limitations by the derivatization and fragmentation of amino acid residues are in progress [48].

Unequivocal sequence data on a protein molecule can be obtained by automated direct sequencing using the Edman degradation method. It allows the discovery of amino acid residues and their order in the protein molecule starting from the N-terminal residue [49]. Usually the elucidation of the amino acid sequence of the N-terminal region is sufficient to perform a homology search in databases and identify a protein. Furthermore, this is an easy and fast procedure. Unfortunately, when the protein has a blocked N-terminal residue, this method is not useful and it becomes necessary to perform preliminary unblocking treatments [50], or fragmentation of the molecule followed by the isolation of an internal fragment to be sequenced.

Actually, since both mass spectrometry and Edman degradation have advantages and disadvantages, the best strategy for the identification of a protein can be based on a dynamic combination of these methodologies. For instance, the unequivocal identification of a protein can be easily obtained by combining data concerning the amino acid sequence of a protein fragment, such as the sequence of the N-terminal region obtained by the Edman

degradation method, and the exact molecular mass of the entire molecule estimated by mass spectrometry. Following this strategy, many allergens have been identified, including some more that were recently described, such as Cit s 2 [51], Bra o 3 [52], Tri a 14 [53], Pla a 3 [54], watermelon allergens [37] and Mor n 3 [12]. If the sequence under investigation is (or homologs with high-percentage identity are) already present in the databases, it will be easily identified using the available softwares for homology search. On the other hand, if the homology search indicates that neither the sequence obtained by Edman degradation is present in the databases, nor homologs of that sequence are detected, then we can conclude that a new protein has been discovered. Unlike some mass spectrometry-based methodologies that provide no response regarding the analyzed protein if it is not present in databases, Edman degradation is a precious tool that always provides information on the protein under investigation, both if it is present or absent in databases. For instance, the kiwi fruit allergen Act d 5 was identified by N-terminal sequencing of the natural protein [36]. Homology searches in databases indicated the absence of known homologs, therefore it was classified as a novel allergen belonging to a new protein family.

Detection & identification of allergenic proteins by component-based allergen microarray

Among the novel methodologies useful for allergy diagnosis we find the increasing use of microarrayed purified known allergens [40,55,56]. It allows the simultaneous measurement of IgE antibodies specific for different individual allergens (i.e., multiplex analysis) with the same serum sample. This methodology can also be considered an additional tool useful to detect allergens in a mixture, such as a total protein extract coming from an allergenic source. In fact, it is possible to perform inhibition tests by pre-incubating the sera of allergic subjects with the protein extract. IgE antibodies of the sera will bind the allergenic proteins present in the extract, thus becoming unavailable for the interaction with the microarrayed molecules for which the reactivity with the sera IgE had already been established. Therefore, IgE binding to the proteins in solution is detected by measurement of decreased binding to the protein(s) on the solid phase. Unlike, for example, IgE immunoblot and ELISA, an additional advantage of this approach is that the test provides information on the identity of the IgE-binding proteins detected.

By using this approach and the Immuno Solid-Phase Allergen Chip (ISAC®) system (VBC-Genomics, Vienna, Austria), we detected the presence of LTP in the protein extract of kiwi fruit by measuring the inhibition of cross-reaction with microarrayed homologs [MARI A ET AL., UNPUBLISHED DATA]. Afterwards, the presence of LTP in the kiwi fruit extract was confirmed by the isolation of this protein and its characterization as a cross-reacting IgE-binding molecule [101].

This approach is very promising and could have an important role in designing future strategies for the identification of allergenic proteins. Hopefully, the number of allergens included in the microarray system will rapidly increase, including representative members of all the protein families containing IgE-binding molecules. In

addition, the inclusion in the microarray of as many proteins as possible, including those never described as allergens, could provide a significant improvement due to the possibility of identifying unsuspected allergens. In fact, the higher the number of microarrayed proteins, in addition to allergens, the better the value of this methodology in the identification of new potential allergens. It is of notice that the availability of proteins on the microarray with an unknown IgE-binding capacity could lead to defining some of them as not allergenic after an exhaustive and extended population testing. This will improve the bases of data to be used for defining features of allergenic and nonallergenic proteins.

Biochemistry-based proteomics & molecular biology strategies

The introduction of molecular biology techniques has provided useful additional tools and advantages to the field of protein/allergen identification and characterization. The genome sequencing of several organisms has enriched the databases with a huge number of gene sequences. Therefore, when homology searches are performed with protein sequence data obtained by proteomics methodologies, the probability of finding homologs allowing the identification of the unknown protein is much higher. A significant contribution derives also from the production and sequencing of cDNA libraries and collections of expressed sequence tags from several organisms and tissues. However, genome-derived sequences do not provide information on the proteins that are really expressed in a specific tissue and in defined conditions. Furthermore, some proteins belong to multigene families and a tissue-specific expression of the different isoforms can occur. For this reason, to investigate the proteome of a tissue, the analysis of mRNA-derived, rather than genome-derived, sequences is often preferred. Several potential allergens have been easily identified by searching in cDNA libraries and expressed sequence tag collections with the sequence information of already known allergenic proteins. However, the estimation of the protein expression directly as mRNA levels may or may not correlate with the protein level in the tissue because the accumulation patterns may be influenced by post-transcriptional processes [57–59]. Therefore, as underlined in FIGURE 1 and TABLE 1, it is still necessary to analyze the proteome directly at the protein level to know if the translated product of a mRNA is contained in a tissue and can act as an allergy trigger. It can happen that an allergen homolog detected at the DNA or mRNA level is not then detected at the protein level owing to the combined effect of several factors contributing to the modulation of protein expression and degradation. In addition, novel potential allergenic proteins, for which no homologs are known, will be overlooked when the analysis of possible allergenic molecules in a tissue is performed by using molecular biology strategies, such as a homology search at the DNA level. For instance, the Bet v 1-like IgE-binding molecules Act d 8 and Act c 8 were identified and characterized by cloning and expressing sequences contained in a kiwi fruit cDNA library [60]. These proteins had been neglected in the past, probably because of their low concentration in kiwi fruit extracts. Conversely, the analysis of the proteome of some green kiwi fruit

extracts has revealed high concentrations of a protein belonging to the Bet v 1 superfamily, but is different from Act d 8 [10]. In fact, it displays a low sequence identity with Act d 8, and has been named Act d 11 [101]. Probably, Act d 11 was overlooked when the screening of the kiwi fruit cDNA library was performed because of the low sequence identity with the Bet v 1 allergens known so far.

Therefore, a wealth of observations suggest that only biochemical characterization performed directly on the natural protein components allows the estimation of their real abundance in a tissue and the occurrence of possible protein processing [61,62] and other post-translational modifications [63–66] that can affect their structural and immunological properties.

Protein biochemistry can take some advantages from molecular biology methodologies to deal with some critical points, such as the elucidation of the complete primary structure of protein molecules (TABLE 1), especially if they have a high molecular weight. In addition, the availability of recombinant products is helpful when the purification of significant amounts of proteins represents a problem due to their very low content in the allergenic source. The combination of gene cloning and heterologous protein expression has allowed the production of a very high number of recombinant allergens. Molecular biology methodologies are considered to be the best way to obtain unlimited amounts of allergens available for diagnosis and therapy. However, their use in substitution of the natural allergens implies a preliminary validation to assess whether the recombinant forms have the same structural and immunological properties as the natural ones [67–69]. For this reason, sometimes the recombinant proteins find a limited diagnostic use due to the scarcity or lack of biochemical and immunological characterization. In fact, the number of characterized allergens available for diagnostic purposes is lower than that simply identified by molecular biology methodologies [70]. It seems that molecular biology-based studies proceed faster than the extensive biochemical and immunological characterization of allergens. Therefore, a concerted and optimized use of the different methodologies would be desirable and would provide the best results.

The usefulness of biochemistry- and molecular biology-based methodologies in the study of allergen structural and immunological characteristics, and in the preparation of purified molecules useful in allergy diagnosis, is summarized in TABLE 1.

Current strategies & new approaches to allergen identification & characterization

Classic and emerging methodologies can be combined to design new approaches or to improve those already known for the identification and characterization of new allergens. Although molecular biology is useful to identify, clone and express allergen homologs from an allergy source, the identification and biochemical and immunological characterization performed directly on the natural form of any allergen is an important step that cannot be avoided. Sometimes, a recombinant form has the same features and behavior of the natural molecule and can be considered a valid substitute. However, the availability of the natural molecule is important to make comparative studies necessary for the validation

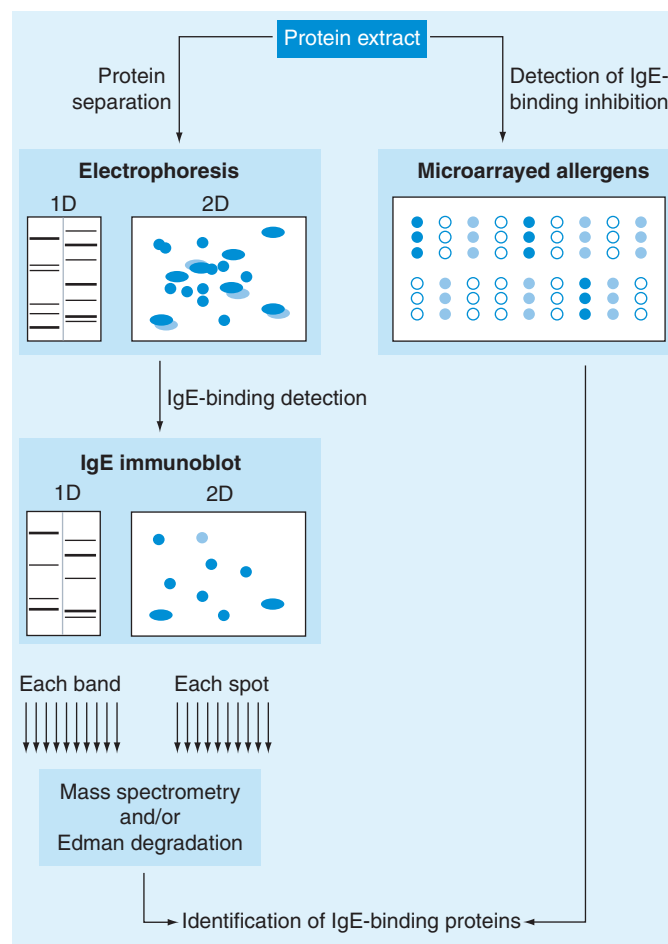


Figure 1. Identification of IgE-binding proteins.

Classic (left side) and emerging (right side) strategies useful for protein/allergen identification in a protein extract. 1D and 2D indicate the protein separation by 1D and 2D electrophoresis, respectively.

of the recombinant form. Therefore, it would be desirable in the future to address some more efforts towards the improvement of methodologies working directly on the natural molecules.

A classic procedure used for the detection of IgE-binding molecules in a protein extract includes their separation by 1D or 2D electrophoresis, followed by immunoblot to reveal the reacting molecules and their identification by Edman degradation or mass spectrometry (FIGURE 1, left side). The identification procedure has to be applied at each IgE-binding band from SDS-PAGE and at each IgE-binding spot from 2-DE.

At present, new methodologies could be exploited to make some steps of the classic procedures easier and faster. For instance, multiplex microarray-based technologies, which are increasingly being used for diagnostic purposes, could also be useful to detect allergens in a mixture by performing inhibition tests that also provide information on the identity of the various detected IgE-binding proteins (see earlier). Generally, in the absence of specific antibodies, other test systems, such as immunoblot and ELISA, just show the presence of IgE-binding proteins. Then it is necessary to investigate each of the detected molecules by Edman degradation

or mass spectrometry to acquire information on their identity. By contrast, the multiplex microarray systems provide simultaneous information on the different IgE-binding proteins present in the mixture and cross-reacting with the allergens immobilized on the microarray (FIGURE 1, right side), thus representing an advantage from a practical and economical point of view. Then, the collected information may be exploited in planning purification of these proteins.

At the end of both the procedures shown in FIGURE 1, new IgE-binding proteins could be identified, but the molecules are not yet available for further characterization. In fact, their biochemical and immunological characterization implies the availability of purified molecules. Therefore, after the identification it becomes necessary to set up a protocol to obtain the pure protein. At beginning, a relatively low amount of purified protein is required for analytical purposes. The individual IgE-binding proteins can be isolated by LC, and affinity chromatography is the preferred technique to be used for their separation (FIGURE 2). Then, the characterization of the molecule properties can be performed with the contribution of mass spectrometry, Edman degradation and molecular biology. These methodologies are complementary and the designing of future strategies should focus on the optimization of their combination in order to obtain the best results. Depending on the specific molecule under investigation, their contribution may include: the study of post-translational modifications and of the presence of isoforms in the sample by mass spectrometry; the collection of crucial information concerning the primary structure by Edman degradation; the elucidation of the complete primary structure; and the study of antigenic epitopes (e.g., by phage display technology) using molecular biology methodologies.

When the structural, immunological and clinical characterization has been completed, the use of the allergenic molecule in diagnosis and, sometimes, in immunotherapy can be evaluated. This implies the availability of high amounts of the pure allergenic protein and, therefore, a procedure for a large-scale purification is required. Sometimes it is very difficult to scale up a purification procedure using the conventional chromatographic techniques, especially when the molecule is present in very low concentrations in the starting natural source. In this case, if a comparative study has indicated that the recombinant form of the allergen is a suitable substitute for the natural molecule, then the molecular biology methodologies can be exploited to easily obtain high amounts of the recombinant allergen. Unfortunately, sometimes the recombinant molecule is not a satisfactory substitute of the natural allergen, especially if it has post-translational modifications. In this case, the scale-up of the procedure used for the analytical purification of the natural allergen can be considered the only choice to obtain high amounts of that allergen and reliable results following its use in diagnosis and therapy. The protein separation by affinity chromatography, performed

Table 1. Usefulness of the biochemistry-based and molecular biology-based methodologies in the study of allergen characteristics.

Characteristic	Biochemistry-based methodologies	Molecular biology-based methodologies
Allergen concentration <i>in vivo</i>	↑	↓
Partial primary structure elucidation	↑	↓
Complete primary structure elucidation	↑	↑
Post-translational modifications	↑	↓
Pure protein in small scale	↑	↓
Pure protein in large scale	↔	↔
Immunological characterization	↑	↑

↑: High usefulness; ↓: Low/absent usefulness; ↔: Comparable usefulness.

using antibodies specific for the allergenic protein, could allow the scale-up of the analytical procedure and the purification of suitable amounts of low-abundant natural allergens.

High-throughput data handling

All the aforementioned techniques leading to the availability of large numbers of purified natural allergens, and the advances in the recombinant DNA technologies, directly resulted in the ability to produce allergen microarrays composed of known allergens, or even parts of them [55]. Microarray technology represents a sophisticated immune-diagnostic tool as it is tailored to revealing the true molecular structures responsible for the immune reactions, while at the same time providing valuable information on an allergen panel with a single and simple assay, markedly accelerating the gathering of experimental, clinical and epidemiological data on newly identified proteins.

This approach allows the multitude of proteins spotted on a solid substrate to be used as probes for the IgE-binding detection after incubation with the allergic or even nonallergic patients' sera. The major advantages of this screening technique are the relatively small serum volume required for performing the test and its high throughput and speed in terms of results, including the detection of different immunoglobulin isotypes (IgE, IgG and its subclasses, IgA, IgM) in the same run.

Unlike in other research fields, for example, cancer, where microarray experimental results are collected in several repositories and are available through common infrastructures and platforms [71,72], such facilities are missing in the allergy/immunology field. Results originating from daily practice or the experimental work from many centers lack integration and standards.

Most of the databases available online [73], including different ranges of information on allergenic molecules, do not report real-time clinical and epidemiological data. A chance to share these resources could be achieved by networking medical institutions and research centers in a common computational infrastructure, or a grid network, providing each member with tools for securely storing, accessing and analyzing data in a distributed computational environment. This approach could be extremely beneficial for the institutions that already possess patient registers and data

but do not own an appropriate computer infrastructure for their management and do not share and compare their findings with similar institutions worldwide.

Allergome-InterAll is a platform developed in this sense. Created and developed by Allergy Data Laboratories (Latina, Italy), it has been designed as a web application to ensure a continuous correspondence between the allergome archives of data [74] and a new patients' record management system (InterAll), experimentally adopted at the Center of Molecular Allergology, IDI-IRCSS (Rome, Italy) and in a few private clinical settings. Routine tests are performed daily on patients' sera using the ISAC CRD 103 ISAC system (VBC-Genomics), to detect the IgE specificities against a panel of 103 molecules related to grass, pollen, animal dander, food, venoms and mites, among others. Once routine analysis is concluded, all results are sent in real time to a central database in an anonymized form (Allergome-ReTiME). Microarray results are interconnected with the results already available from other tests (e.g., skin test, IgE from singleplex systems and oral challenge). The more information loaded in the system the more accurate the overall view regarding single IgE-binding or not binding structures becomes.

Microarray data and tools to correctly evaluate them are therefore important in clinical practice and research applied to the study of allergens. Being valuable for explaining allergen IgE co-recognition, this technology can be further used for the development of individualized therapies: a microarray test result may provide an insight into the components that are required for an effective immunotherapy.

Bioinformatics search strategies for the identification of new allergens

In recent years, as a result of advances in recombinant DNA technology and the production of recombinant and xenogenic proteins in plant biotechnology, a compelling need for protein allergenicity prediction methods has emerged. The current availability of many databases of allergenic molecules, together with the great potential of bioinformatics and statistical tools, represent an important resource in the allergenicity risk assessment.

In this direction the WHO and the Food and Agriculture Organization (FAO) provided a guideline [75–77] with standardized procedures in addition to biological tests and introduced a decision-tree approach to the evaluation of the allergenicity of genetically modified foods. The FAO–WHO guideline is based on the hypothesis that if the structure of a query protein is sufficiently similar to any known allergen, then the protein itself is likely to be an allergen. More stringent rules have been adopted: a protein query is potentially allergenic if it shares a minimum of

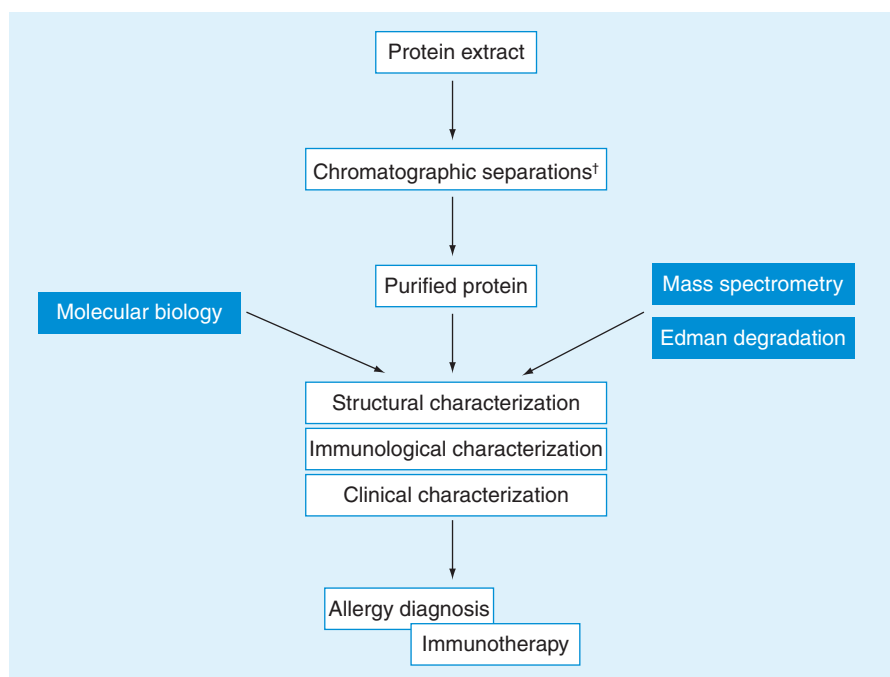


Figure 2. Purification and characterization of allergens from natural sources. The characterization of IgE-binding proteins may exploit the contribution of mass spectrometry, Edman degradation and molecular biology methodologies.

†Affinity chromatography is the best choice when specific antibodies are available.

35% sequence over an 80 amino acid window or at least six identical contiguous amino acids with a known allergen [78], the latter rule to encompass allergen epitopes. The global sequence similarity search (first criterion) is commonly performed using search algorithms such as FASTA or BLAST [79,80], and although it has been shown to be too stringent to find most true allergens [81,82], it is still accepted. Further studies have also shown that even the six contiguous amino acids criterion produces many spurious hits, and in order to achieve more effective results, the number of contiguous amino acids should be increased to eight [81,83]. Further studies have highlighted that certain protein structures and motifs appear to be particularly related to the protein allergenicity and that, among thousands of protein families, known allergens can be numbered among only a small number of families [83–85].

Another approach for the *in silico* prediction of an allergenic protein, which has not yet been tested experimentally, identified residue conservation patterns (a high proportion of surface-exposed hydrophobic residues) in the proteins Ara t 8, Act c 1, Bet v 1 and Ves v 5, indicating them as possible common structural features recognized by the immune system [86,87].

However, it is important to understand that no bioinformatics analysis alone can be used to predict whether a protein will be allergenic beyond a shadow of a doubt, and that once a novel protein has been demonstrated to be similar to any known allergen, further studies are needed, involving specific IgE binding, to assess its allergenicity.

A combined usage of the already available online allergen databases and the clinical information derived from the experimental data should also be used to enhance the *in silico* allergen

prediction. High-throughput screening technologies, such as allergen biochips for IgE detection [88], provide a huge amount of valuable experimental data that can be virtually used as a starting point to identify and collect new allergenic structures. With this aim, and to make maximum use of the data generated via an established screening method at the Center for Molecular Allergology, IDI-IRCSS, the Allergome platform has recently been provided with a procedure that merges the data of peptide microarray tests and an algorithm that searches for novel allergens in the UniprotKB sequence repository [89]. The custom peptide biochip (VBC-Genomics) has been designed for testing purposes in order to contain an array of 46 peptides (length: 15 amino acids; offset: six amino acids) spanning the entire Pen a 1.0102 tropomyosin molecule [90]. Moreover, 121 other peptides from the complete sequences of the allergens Der p 10.0101, Bla g 7.0101 and Asc s 3.0101 were computed, taking into account the 15 amino acid-long sequences with an offset of six amino acids that were not duplicated in the Pen a 1.0102 peptide panel. A total of 67 peptides were entered in the Allergome database and then spotted on the biochip to be assayed with allergic patients' sera. Since the peptide array is continuously feeding a java-based routine search in both the Allergome and the UniprotKB databases for sequences with 100% identity with the query array data, more than 2000 sequences were retrieved as putative *in silico* allergenic peptides and inserted in Allergome. In this manner, by using the Allergome-InterAll platform, we are now able to map the IgE binding for each peptide in the repository. Microarray results feed in real time the Allergome database and, as a consequence, the Allergome-InterAll platform is able to determine the peptide binding of every new sequence, which will be deposited in the UniprotKB, regardless of it being previously reported to be allergenic. Proteins identified as potentially allergenic should then be directly purified from the source organisms in order to verify *ex silico*-produced data.

Further developments of this and other *in silico*-based procedures representing valid tools to help predict the allergenicity of novel proteins could find a broader application.

Expert commentary

The allergenic proteins described so far can be included in a few protein families and possess a limited number of biochemical functions [91]. Most of them have been identified following similar procedures, generally including molecular biology and mass spectrometry methodologies. In addition, similar protocols for protein extraction from the natural source are usually used. Therefore, the possibility cannot be excluded that, depending on the methodologies used, some proteins are selectively identified, whereas others are missed. For example, only homologs of already known allergens will be found when a cDNA library is screened using as a probe either DNA or protein sequence information deriving from known allergens. A novel allergenic protein will probably not be identified if it does not have a homolog in databases and if the analysis of the molecule is performed by mass spectrometry. Furthermore, some allergens are detected as IgE-binding molecules when they are tested by dot blot rather than by 1D or 2D immunoblot [92,93].

It would be interesting to test as many proteins as possible for their IgE-binding capacity, including those described as allergens and those never described as allergens, and using a very large population. These results will be hopefully easily achievable by establishing a worldwide network of allergen/microarray/information technology users, either in research laboratories or in clinical settings. The combined use of the three technologies currently under development and testing [73,88] will allow the transfer on an internet-based platform of a wealth of data that has never before been available in the history of biomedicine. The result analysis may or may not indicate that the possible allergenic molecules belong to a few protein families. Whatever the result is, the comparative analysis of the structural features of antigenic and nonantigenic proteins would be useful to learn something more about the molecular determinants able to trigger IgE synthesis or not.

Five-year view

Most of the research groups involved in the identification and characterization of allergenic proteins have a wide expertise in molecular biology and/or mass spectrometry fields. Very often, direct protein sequencing by an automated sequencer is used as a complementary technology. During the last few years, new molecular biology-based technologies and mass spectrometry equipments have been acquired by many laboratories. Meanwhile, the company (Applied Biosystems, CA, USA), which during the last few decades has provided laboratories all around the world with automated protein sequencers based on the Edman degradation chemistry, decided to stop the production of these instruments in 2008 and to stop all technical assistance by 2013. Therefore, the contribution of automated protein sequencers to the identification and characterization of allergens is expected to decrease in the near future, and mass spectrometry will become the only way to analyze the primary structure of a natural protein/allergen. However, despite the increasing improvement of its performance, and although it is a precious tool in proteomics studies, mass spectrometry is still not ready to completely substitute the Edman degradation-based technology. Emerging methodologies, such as protein microarray [88,94], could provide a valuable contribution to the study of allergenic proteins. High-throughput technologies, such as allergen and peptide biochips, would greatly enhance the knowledge of key allergenic structures, providing facilities and processing protocols to handle a huge mass of information. It can be imagined that, in the near future, this scenario may have some effects on the field of identification and characterization of allergens.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Key issues

- Each methodology and procedure used to identify new allergenic molecules shows advantages and pitfalls, indicating that further efforts are necessary to overcome some critical points.
- The preparation of high-quality protein/allergen extracts is still a challenge for researchers.
- Depending on the techniques used for protein separation and identification, some allergens could be lost or overlooked.
- A collection of mass spectrometry-based methodologies and strategies can be used to collect information useful for the identification of a protein, but unequivocal sequence data on a natural protein molecule can only be obtained by automated direct sequencing using the Edman degradation method.
- Estimations of the protein expression directly as mRNA levels may or may not correlate with the protein level in the tissue, because the accumulation patterns may be influenced by post-transcriptional processes.
- Molecular biology is useful to identify, clone and express allergen homologs from an allergy source, but the identification and biochemical and immunological characterization performed directly on the natural form of any allergen is an important step that cannot be avoided.
- Component-based allergen microarray allows the simultaneous measurement of IgE antibodies specific for different individual allergens.
- Component-based allergen microarray is also a tool useful to detect allergens in a mixture and to provide information on the identity of the IgE-binding proteins detected.
- The adoption of microarrays containing overlapping peptides representing putative epitopes could greatly increase the precision of screening, consequently improving the immunotherapy accuracy.

References

Papers of special note have been highlighted as:

- of considerable interest

- Larsen JN, Dreborg S. Standardization of allergen extracts. *Methods Mol. Med.* 138, 133–145 (2008).
- Matthes A, Schmitz-Eiberger M. Apple (*Malus domestica* L. Borkh.) allergen Mal d 1: effect of cultivar, cultivation system, and storage conditions. *J. Agric. Food Chem.* 57, 10548–10553 (2009).
- van Kampen V, Merget R, Rabstein S *et al.* Comparison of wheat and rye flour solutions for skin prick testing: a multi-centre study (Stad 1). *Clin. Exp. Allergy* 39, 1896–1902 (2009).
- Sun JL, Zhang HY, Guo ZY, Ying WT, Qian XH, Wang JL. Comparison of three methods of protein extraction from *Dermatophagoides pteronyssinus* for two-dimensional electrophoresis. *Chin. Med. Sci. J.* 24, 64–68 (2009).
- Zheng Q, Song J, Doncaster K, Rowland E, Byers DM. Qualitative and quantitative evaluation of protein extraction protocols for apple and strawberry fruit suitable for two-dimensional electrophoresis and mass spectrometry analysis. *J. Agric. Food Chem.* 55, 1663–1673 (2007).
- Saravanan RS, Rose JK. A critical evaluation of sample extraction techniques for enhanced proteomic analysis of recalcitrant plant tissues. *Proteomics* 4, 2522–2532 (2004).
- Brunetto B, Tinghino R, Braschi MC, Antonicelli L, Pini C, Iacovacci P. Characterization and comparison of commercially available mite extracts for *in vivo* diagnosis. *Allergy* 65, 184–190 (2010).
- Vassilopoulou EV, Zuidmeer L, Akkerdaas J *et al.* Optimized techniques for the extraction of grape allergens appropriate for *in vivo* and *in vitro* testing and diagnosis. *Mol. Nutr. Food Res.* 51, 360–366 (2007).
- Alche JD, Castro AJ, Jimenez-Lopez JC *et al.* Differential characteristics of olive pollen from different cultivars: biological and clinical implications. *J. Investig. Allergol. Clin. Immunol.* 17, 17–23 (2007).
- Ciardiello MA, Giangrieco I, Tuppo L *et al.* Influence of the natural ripening stage, cold storage, and ethylene treatment on the protein and IgE-binding profiles of green and gold kiwi fruit extracts. *J. Agric. Food Chem.* 57, 1565–1571 (2009).
- **Reports a significant example concerning the influence of some factors on the number and relative abundance of allergens in the total protein extracts.**
- Krause S, Reese G, Randow S *et al.* Lipid transfer protein (Ara h 9) as a new peanut allergen relevant for a Mediterranean allergic population. *J. Allergy Clin. Immunol.* 124, 771–778 (2009).
- Ciardiello MA, Palazzo P, Bernardi ML *et al.* Biochemical, immunological and clinical characterization of a cross-reactive nonspecific lipid transfer protein 1 from mulberry. *Allergy* 65, 597–605 (2010).
- Hildebrandt S, Steinhart H, Paschke A. Comparison of different extraction solutions for the analysis of allergens in hen's egg. *Food Chem.* 108, 1088–1093 (2008).
- Dooper MM, Plassen C, Holden L, Moen LH, Namork E, Egaas E. Antibody binding to hazelnut (*Corylus avellana*) proteins: the effects of extraction procedure and hazelnut source. *Food Agric. Immunol.* 19, 229–240 (2008).
- Focke M, Marth K, Valenta R. Molecular composition and biological activity of commercial birch pollen allergen extracts. *Eur. J. Clin. Invest.* 39, 429–436 (2009).
- Kottapalli KR, Payton P, Rakwal R *et al.* Proteomic analysis of mature seed of four peanut cultivars using two-dimensional gel electrophoresis reveals distinct differential expression of storage, anti-nutritional, and allergenic proteins. *Plant Sci.* 175, 321–329 (2008).
- Primavesi L, Brenna OV, Pompei C, Pravettoni V, Farioli L, Pastorello EA. Influence of cultivar and processing on cherry (*Prunus avium*) allergenicity. *J. Agric. Food Chem.* 54, 9930–9935 (2006).
- Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680–685 (1970).
- Maghuly F, Marzban G, Laimer M. Functional genomics of allergen gene families in fruits. *Nutrients* 1, 119–132 (2009).
- O'Farrell PH. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250, 4007–4021 (1975).
- Chen L, Hefle SL, Taylor SL, Swoboda I, Goodman RE. Detecting fish parvalbumin with commercial mouse monoclonal anti-frog parvalbumin IgG. *J. Agric. Food Chem.* 54, 5577–5582 (2006).

- 22 Breitenbach M, Ferreira F, Jilek A *et al.* Biological and immunological importance of Bet v 1 isoforms. *Adv. Exp. Med. Biol.* 409, 117–126 (1996).
- 23 Guo B, Liang X, Chung SY, Holbrook CC, Maleki SJ. Proteomic analysis of peanut seed storage proteins and genetic variation in a potential peanut allergen. *Protein Pept. Lett.* 15, 567–577 (2008).
- 24 Porterfield HS, Murray KS, Schlichting DG *et al.* Effector activity of peanut allergens: a critical role for Ara h 2, Ara h 6, and their variants. *Clin. Exp. Allergy* 39, 1099–1108 (2009).
- 25 Wallner M, Erler A, Hauser M *et al.* Immunologic characterization of isoforms of Car b 1 and Que a 1, the major hornbeam and oak pollen allergens. *Allergy* 64, 452–460 (2009).
- 26 Alm R, Ekefjard A, Krogh M, Hakkinen J, Emanuelsson C. Proteomic variation is as large within as between strawberry varieties. *J. Proteome Res.* 6, 3011–3020 (2007).
- 27 Sancho AI, Gillabert M, Tapp H, Shewry PR, Skeggs PK, Mills EN. Effect of environmental stress during grain filling on the soluble proteome of wheat (*Triticum aestivum*) dough liquor. *J. Agric. Food Chem.* 56, 5386–5393 (2008).
- 28 Biron DG, Brun C, Lefevre T *et al.* The pitfalls of proteomics experiments without the correct use of bioinformatics tools. *Proteomics* 6, 5577–5596 (2006).
- Review article describing advantages and misuses of bioinformatics tools in proteomics experiments.
- 29 Carpentier SC, Witters E, Laukens K, Deckers P, Swennen R, Panis B. Preparation of protein extracts from recalcitrant plant tissues: an evaluation of different methods for two-dimensional gel electrophoresis analysis. *Proteomics* 5, 2497–2507 (2005).
- 30 Jorin JV, Maldonado AM, Castillejo MA. Plant proteome analysis: a 2006 update. *Proteomics* 7, 2947–2962 (2007).
- 31 Marzban G, Herndl A, Maghuly F, Katinger H, Laimer M. Mapping of fruit allergens by 2D electrophoresis and immunodetection. *Expert Rev. Proteomics* 5, 61–75 (2008).
- 32 Kitta K, Ohnishi-Kameyama M, Moriyama T, Ogawa T, Kawamoto S. Detection of low-molecular weight allergens resolved on two-dimensional electrophoresis with acid-urea polyacrylamide gel. *Anal. Biochem.* 351, 290–297 (2006).
- 33 Friedman DB, Hoving S, Westermeier R. Isoelectric focusing and two-dimensional gel electrophoresis. *Methods Enzymol.* 463, 515–540 (2009).
- 34 Pastorello EA, Conti A, Pravettoni V *et al.* Identification of actinidin as the major allergen of kiwi fruit. *J. Allergy Clin. Immunol.* 101, 531–537 (1998).
- 35 Gavrovic-Jankulovic M, Cirkovic T, Vuckovic O *et al.* Isolation and biochemical characterization of a thaumatin-like kiwi allergen. *J. Allergy Clin. Immunol.* 110, 805–810 (2002).
- 36 Tamburrini M, Cerasuolo I, Carratore V *et al.* Kiwellin, a novel protein from kiwi fruit. Purification, biochemical characterization and identification as an allergen*. *Protein J.* 24, 423–429 (2005).
- 37 Pastor C, Cuesta-Herranz J, Cases B *et al.* Identification of major allergens in watermelon. *Int. Arch. Allergy Immunol.* 149, 291–298 (2009).
- 38 Engvall E, Perlmann P. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunochemistry* 8, 871–874 (1971).
- 39 MacBeath G. Protein microarrays and proteomics. *Nat. Genet.* 32(Suppl.), 526–532 (2002).
- 40 Harwanegg C, Hiller R. Protein microarrays for the diagnosis of allergic diseases: state-of-the-art and future development. *Clin. Chem. Lab. Med.* 43, 1321–1326 (2005).
- 41 Gonzalez-Buitrago JM, Ferreira L, Isidoro-Garcia M, Sanz C, Lorente F, Davila I. Proteomic approaches for identifying new allergens and diagnosing allergic diseases. *Clin. Chim. Acta* 385, 21–27 (2007).
- 42 Han X, Aslanian A, Yates JR 3rd. Mass spectrometry for proteomics. *Curr. Opin. Chem. Biol.* 12, 483–490 (2008).
- 43 Monaci L, Visconti A. Mass spectrometry-based proteomics methods for analysis of food allergens. *Trends Anal. Chem.* 28, 581–591 (2009).
- Overview of mass spectrometry-based methodologies applied to allergen identification studies.
- 44 Chmelik J, Zidkova J, Rehulka P, Petry-Podgorska I, Bobalova J. Influence of different proteomic protocols on degree of high-coverage identification of nonspecific lipid transfer protein 1 modified during malting. *Electrophoresis* 30, 560–567 (2009).
- 45 Kirsch S, Fourdrilis S, Dobson R, Scippo ML, Maghuin-Rogister G, De Pauw PE. Quantitative methods for food allergens: a review. *Anal. Bioanal. Chem.* 395, 57–67 (2009).
- 46 Pevtsov S, Fedulova I, Mirzaei H, Buck C, Zhang X. Performance evaluation of existing *de novo* sequencing algorithms. *J. Proteome Res.* 5, 3018–3028 (2006).
- 47 DiMaggio PA Jr, Floudas CA. *De novo* peptide identification via tandem mass spectrometry and integer linear optimization. *Anal. Chem.* 79, 1433–1446 (2007).
- 48 Cook SL, Collin OL, Jackson GP. Metastable atom-activated dissociation mass spectrometry: leucine/isoleucine differentiation and ring cleavage of proline residues. *J. Mass Spectrom.* 44, 1211–1223 (2009).
- Describes attempts to overcome still unsolved limitations of protein sequencing by mass spectrometry.
- 49 Niall HD. Automated Edman degradation: the protein sequenator. *Methods Enzymol.* 27, 942–1010 (1973).
- 50 Ciardiello MA, D'Avino R, Amoresano A *et al.* The peculiar structural features of kiwi fruit pectin methylesterase: amino acid sequence, oligosaccharides structure, and modeling of the interaction with its natural proteinaceous inhibitor. *Proteins* 71, 195–206 (2008).
- 51 Lopez-Torrejon G, Ibanez MD, Ahrazem O *et al.* Isolation, cloning and allergenic reactivity of natural profilin Cit s 2, a major orange allergen. *Allergy* 60, 1424–1429 (2005).
- 52 Palacin A, Cumplido J, Figueroa J *et al.* Cabbage lipid transfer protein Bra o 3 is a major allergen responsible for cross-reactivity between plant foods and pollens. *J. Allergy Clin. Immunol.* 117, 1423–1429 (2006).
- 53 Palacin A, Quirce S, Armentia A *et al.* Wheat lipid transfer protein is a major allergen associated with baker's asthma. *J. Allergy Clin. Immunol.* 120, 1132–1138 (2007).
- 54 Lauer I, Miguel-Moncin MS, Abel T *et al.* Identification of a plane pollen lipid transfer protein (Pla a 3) and its immunological relation to the peach lipid-transfer protein, Pru p 3. *Clin. Exp. Allergy* 37, 261–269 (2007).
- 55 Hiller R, Laffer S, Harwanegg C *et al.* Microarrayed allergen molecules: diagnostic gatekeepers for allergy treatment. *FASEB J.* 16, 414–416 (2002).

- The first study introducing the antigen microarray technology as a valid instrument for the diagnosis of allergy.
- 56 Harwanegg C, Hutter S, Hiller R. Allergen microarrays for the diagnosis of specific IgE against components of cow's milk and hen's egg in a multiplex biochip-based immunoassay. *Methods Mol. Biol.* 385, 145–157 (2007).
- 57 Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* 19, 1720–1730 (1999).
- 58 Futch B, Latter GI, Monardo P, McLaughlin CS, Garrels JI. A sampling of the yeast proteome. *Mol. Cell. Biol.* 19, 7357–7368 (1999).
- 59 Grimplet J, Wheatley MD, Jouira HB, Deluc LG, Cramer GR, Cushman JC. Proteomic and selected metabolite analysis of grape berry tissues under well-watered and water-deficit stress conditions. *Proteomics* 9, 2503–2528 (2009).
- 60 Oberhuber C, Bulley SM, Ballmer-Weber BK *et al.* Characterization of Bet v 1-related allergens from kiwifruit relevant for patients with combined kiwifruit and birch pollen allergy. *Mol. Nutr. Food Res.* 52, 230–240 (2008).
- 61 Wopfner N, Gadermaier G, Egger M *et al.* The spectrum of allergens in ragweed and mugwort pollen. *Int. Arch. Allergy Immunol.* 138, 337–346 (2005).
- 62 Tuppo L, Giangrieco I, Palazzo P *et al.* Kiwellin, a modular protein from green and gold kiwi fruits: evidence of *in vivo* and *in vitro* processing and IgE binding. *J. Agric. Food Chem.* 56, 3812–3817 (2008).
- 63 Soldatova LN, Tsai C, Dobrovolskaia E, Markovic-Housley Z, Slater JE. Characterization of the N-glycans of recombinant bee venom hyaluronidase (Api m 2) expressed in insect cells. *Allergy Asthma Proc.* 28, 210–215 (2007).
- 64 Chevigne A, Barumandzadeh R, Gros Lambert S *et al.* Relationship between propeptide pH unfolding and inhibitory ability during ProDer p 1 activation mechanism. *J. Mol. Biol.* 374, 170–185 (2007).
- 65 Gruber P, Gadermaier G, Bauer R *et al.* Role of the polypeptide backbone and post-translational modifications in cross-reactivity of Art v 1, the major mugwort pollen allergen. *Biol. Chem.* 390, 445–451 (2009).
- 66 Fenaille F, Nony E, Chabre H *et al.* Mass spectrometric investigation of molecular variability of grass pollen group 1 allergens. *J. Proteome Res.* 8, 4014–4027 (2009).
- 67 Kurup VP, Sussman GL, Yeang HY *et al.* Specific IgE response to purified and recombinant allergens in latex allergy. *Clin. Mol. Allergy* 3, 11–19 (2005).
- 68 Fuchs HC, Bohle B, Dall'Antonia Y *et al.* Natural and recombinant molecules of the cherry allergen Pru av 2 show diverse structural and B cell characteristics but similar T cell reactivity. *Clin. Exp. Allergy* 36, 359–368 (2006).
- 69 Wallner M, Himly M, Neubauer A *et al.* The influence of recombinant production on the immunologic behavior of birch pollen isoallergens. *PLoS ONE* 4, e8457 (2009).
- 70 Asero R. Plant food allergies: a suggested approach to allergen-resolved diagnosis in the clinical practice by identifying easily available sensitization markers. *Int. Arch. Allergy Immunol.* 138, 1–11 (2005).
- 71 Barrett T, Suzek TO, Troup DB *et al.* NCBI GEO: mining millions of expression profiles – database and tools. *Nucleic Acids Res.* 33, D562–D566 (2005).
- 72 Brazma A, Parkinson H, Sarkans U *et al.* ArrayExpress – a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* 31, 68–71 (2003).
- 73 Mari A, Rasi C, Palazzo P, Scala E. Allergen databases: current status and perspectives. *Curr. Allergy Asthma Rep.* 9, 376–383 (2009).
- 74 Mari A, Scala E, Palazzo P, Ridolfi S, Zennaro D, Carabella G. Bioinformatics applied to allergy: allergen databases, from collecting sequence information to data integration. The Allergome platform as a model. *Cell. Immunol.* 244, 97–100 (2006).
- 75 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.* 36(Suppl.), S165–S186 (1996).
- 76 FAO/WHO. Evaluation of allergenicity of genetically modified foods. *Report of a Joint FAO/WHO Expert Consultation of Allergenicity of Foods Derived from Biotechnology*. Rome, Italy, 22–25 January 2001.
- 77 Codex Alimentarius Commission. *Guideline for the Conduct of Food Safety Assessment of Foods Derived from Recombinant-DNA Plants* and Appendix IV, *Annex on the Assessment of Possible Allergenicity (Appendix III)*. 25th Session, Rome, Italy, 30 June–5 July 2003.
- 78 Aalberse RC. Structural biology of allergens. *J. Allergy Clin. Immunol.* 106, 228–238 (2000).
- Key article exploring the determinants of allergenicity and the cross-reactivity phenomena.
- 79 Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA* 85, 2444–2448 (1988).
- 80 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410 (1990).
- 81 Hileman RE, Silvanovich A, Goodman RE *et al.* Bioinformatic methods for allergenicity assessment using a comprehensive allergen database. *Int. Arch. Allergy Immunol.* 128, 280–291 (2002).
- 82 Li KB, Issac P, Krishnan A. Predicting allergenic proteins using wavelet transform. *Bioinformatics* 20, 2572–2578 (2004).
- 83 Stadler MB, Stadler BM. Allergenicity prediction by protein sequence. *FASEB J.* 17, 1141–1143 (2003).
- 84 Ivanciuc O, Schein CH, Braun W. SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Res.* 31, 359–362 (2003).
- 85 Jenkins JA, Griffiths-Jones S, Shewry PR, Breiteneder H, Mills EN. Structural relatedness of plant food allergens with specific reference to cross-reactive allergens: an *in silico* analysis. *J. Allergy Clin. Immunol.* 115, 163–170 (2005).
- 86 Furmonaviciene R, Sutton BJ, Glaser F *et al.* An attempt to define allergen-specific molecular surface features: a bioinformatic approach. *Bioinformatics* 21, 4201–4204 (2005).
- 87 Thalhamer T, Dobias H, Stepanoska T *et al.* Designing hypoallergenic derivatives for allergy treatment by means of *in silico* mutation and screening. *J. Allergy Clin. Immunol.* 125, 926–934 (2010).
- 88 Scala E, Alessandri C, Bernardi ML *et al.* Cross-sectional survey on immunoglobulin E reactivity in 23 077 subjects using an allergenic molecule-based microarray detection system. *Clin. Exp. Allergy* 40, 911–921 (2010).
- First article exploring the IgE reactivity to a broad number of allergenic molecule in a huge population by means of a microarray system.
- 89 Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. UniProtKB/Swiss-Prot. *Methods Mol. Biol.* 406, 89–112 (2007).

- 90 Ayuso R, Lehrer SB, Reese G. Identification of continuous, allergenic regions of the major shrimp allergen Pen a 1 (tropomyosin). *Int. Arch. Allergy Immunol.* 127, 27–37 (2002).
- 91 Radauer C, Bublin M, Wagner S, Mari A, Breiteneder H. Allergens are distributed into few protein families and possess a restricted number of biochemical functions. *J. Allergy Clin. Immunol.* 121, 847–852 (2008).
- 92 O'Brien RM, Thomas WR, Tait BD. An immunogenetic analysis of T-cell reactive regions on the major allergen from the house dust mite, Der p I, with recombinant truncated fragments. *J. Allergy Clin. Immunol.* 93, 628–634 (1994).
- 93 Weghofer M, Thomas WR, Kronqvist M *et al.* Variability of IgE reactivity profiles among European mite allergic patients. *Eur. J. Clin. Invest.* 38, 959–965 (2008).
- 94 Stoevesandt O, Taussig MJ, He M. Protein microarrays: high-throughput tools for proteomics. *Expert Rev. Proteomics* 6, 145–157 (2009).
- Comprehensive overview of possible applications of microarray biotechnology in different fields.

Website

- 101 Allergome
www.allergome.org