

Arbeiten aus dem Paul-Ehrlich-Institut
11th International Paul-Ehrlich-Seminar, Bad Homburg 2005
Verlag Chmielorz

Physico-chemical Characterization of Candidate Reference Materials

Fatima Ferreira, Michael Wallner, Gabriele Gadermaier, Anja Erler,
Gerhard Fritz, Otto Glatter, Martin Himly, Peter Briza, Ronald van Ree and the
CREATE Consortium

1. Introduction

Presently, the generation of allergen products for allergy diagnosis and treatment is based on heterogeneous allergen extracts. These extracts include non-allergenic proteins and have an extremely complex and variable composition. Current standardisation protocols measure overall potencies on the basis of reactivity with human IgE antibodies from pooled allergic sera [1]. This approach does not reveal differences in composition; the major disease-related constituents, i.e. the major allergens, are not quantified [2]. Consequently, extracts with similar overall potencies can be quite different in composition. In addition, units used to express potencies are different for each allergen manufacturer.

The most important major inhalant allergens have been identified, and specific monoclonal antibodies are available. This offers the possibility to standardise allergen products on the basis of uniform mass units of major (disease-related) allergens [3]. In addition, most major allergens are now available as recombinant proteins [4].

CREATE, an EU-funded project aimed at the development of certified reference materials (CRM) based on purified natural or recombinant major allergens [5, 6]. Twenty-nine partners from industry, research laboratories, regulatory bodies and clinical centers worked together on 8 relevant allergens. To this end, the major allergens from birch pollen (Bet v 1), grass pollen (Phl p 1 and Phl p 5), olive pollen (Ole e 1), and house dust mite (Der p 1, Der p 2, Der f 1, Der f 2) were purified from their natural sources and produced as recombinant proteins.

The task of the Department of Molecular Biology at the University of Salzburg was to evaluate these natural and recombinant allergen preparations with respect to their physico-chemical properties. In general, the following physico-chemical parameters [7] are considered important for the characterization of protein preparations:

(i) Identity: must contain recombinant or natural protein with correct amino acid composition and sequence. Mass spectrometry and amino acid analysis are the methods selected for this task. Mass spectrometry data must show molecular mass according to primary structure. Amino acid analyses must show amino acid content according to primary structure.

(ii) Purity: the allergen preparations must be pure with respect to protein content. The method of choice is SDS-polyacrylamide gel electrophoresis (SDS-PAGE). More than 95% of the material must be recovered in a single band after SDS-PAGE followed by protein silver-stain. In addition, amino acid analysis can provide information on the purity of a protein sample concerning contamination by other proteins.

(iii) Homogeneity: the recombinant preparations must be homogenous with respect to molecular weight species and/or comparable to the natural counterparts. One method that is commonly used is analytical gel filtration. For monomeric proteins, the sample must contain a single peak of expected molecular weight when analyzed by analytical gel filtration using an appropriate buffer.

(iv) Structure: the allergen preparations must contain correctly folded protein. Circular dichroism (CD) spectroscopy is commonly used for this purpose. CD of candidate molecules must show typical spectra of folded protein with peak amplitudes similar to reference spectra, in this case natural allergen preparations. Small angle X-ray scattering can also provide valuable information as denatured proteins might show altered pair distance distribution function (PDDF).

Thus, in the CREATE project, the allergen preparations were analyzed concerning their identity (amino acid analysis, mass spectrometry), homogeneity (analytical gel filtration, small angle X-ray scattering), folding and thermal stability (circular dichroism spectroscopy). Table 1 gives an overview of the methods chosen for characterization of the candidate allergens and the information that can be derived concerning identity, purity, homogeneity and secondary structure.

Below, we give a brief overview on the methods employed for the physico-chemical characterization of the allergen preparations and on the conditions applied for the analyses.

2. Amino acid analysis

Although amino acid analysis lacks the speed and sensitivity of mass spectrometry methods, still it is an indispensable tool in the analysis of protein pharmaceuticals. When performed on purified proteins, the technique allows the identification of the protein. It is also an extremely useful method for quantification of proteins, giving more accurate results than colorimetric methods. The accuracy of amino acid analysis depends on the integrity of the sample. Other critical factors affecting accuracy include purity of the reagents used, presence of salts, metals or detergents, and sample handling [8].

Amino acid composition of purified natural and recombinant allergens in the CREATE project was determined using the Pico Tag method (Waters, Milford, MA,

	Identity	Purity	Homogeneity	Secondary Structure (Folding)
Amino acid analysis	Amino acid content according to primary structure.	Contamination by other proteins might change the content of individual amino acids.		
Mass spectrometry/peptide mapping	Molecular mass according to primary structure.	Contamination by other proteins or non-protein compounds gives rise to additional peaks.	Post-translation modifications and chemical modifications resulting from the recombinant production can be detected and accounted for heterogeneity.	
HPLC size exclusion chromatography (HPLC-SEC)			Homogeneity in solution: oligomeric/monomeric forms, aggregates, and degradation products can be detected.	Changes in retention times might also be attributable to denatured protein
Circular Dichroism (CD)				Folded status of proteins. Quantification of secondary structure elements is possible, but usually not reliable. Thermostability.
Small angle X-ray scattering (SAXS)			Molecular size and aggregation status of proteins in solution.	Denatured proteins might show altered PDDF
SDS-PAGE/PAS staining		Contaminating proteins with different molecular weight can be detected.	Post-translation modifications might give rise to additional bands.	

Table 1. Physico-chemical parameters for proteins and selected analytical methods used in the CREATE project

USA). Briefly, an aliquot (0.5-10 μ g) was dried under vacuum in 6 x 50 mm glass tubes. Protein samples were hydrolyzed in constant boiling 6 M HCl for 18-20 h at 116°C under vacuum. After neutralization with aqueous 20% (v/v) triethylamine and 40% (v/v) ethanol, samples and amino acid standards (Pierce, IL, USA) were dried and derivatization of the amino acids performed at room temperature for 10 min using phenylisothiocyanate in solutions of varying polarity. For determination and quantification of hydroxyproline, this amino acid was added to the amino acid stan-

dard mix and samples were processed in the same way. Cysteine and Tryptophan were not analyzed. Phenylthiocarbamyl amino acid derivatives were analyzed by reversed phase-HPLC (HP 1100, Agilent Technologies, Palo Alto, CA, USA) using a 3.9 x 150 mm, 4 μ m Nova-Pak C18 column (Waters, Milford, MA, USA). UV absorbance was monitored at 254 nm. Data obtained from amino acid analysis was also used to determine the protein concentration of all samples.

3. Peptide mapping, sequencing of peptides by MS/MS

A given protein sequence can be digested with a specific protease like trypsin to give rise to a predictable set of peptides. Accurate determination of the masses of even a few peptides generated through trypsin digestion (to an accuracy of 1 Dalton) allows the comparison of experimentally determined masses with the masses of peptides obtained through *in silico* digestion of protein sequences and matching masses can be assigned to predicted peptides. However, the unambiguous assignment of peptide masses is only possible after sequencing. Sequencing of peptides requires tandem mass spectrometry, which is usually denoted MS/MS [9, 10]. In this technique, a given parent (precursor) ion is selected in one mass spectrometer and then fragmented, usually by collisions. The m/z (mass/charge) values for the resulting daughter (product) ions are measured in a second mass spectrometer. Under favorable conditions, this procedure may yield a series of ions containing sufficient information to determine the peptide sequence. In the CREATE project, underivatized and unlabeled peptides were sequenced after ESI (Electrospray ionization) and collision-induced dissociation (CID) in a quadrupole/TOF (QTOF) instrument. Extraction of the peptide sequence from the daughter ion spectra depends on the completeness of the observed series, the accuracy with which the mass differences between the relevant ion peaks are determined and the extent to which the fragmentation spectrum can be correctly interpreted. A peptide resulting from tryptic digestion normally has a basic residue (arginine or lysine) at its C terminus and yields a prominent doubly charged ion peak when ionized by ESI. If this ion is chosen as the parent ion for an MS/MS measurement, the production of a series of y ion daughters (ions resulting from cleavage at the amide bonds and containing the C terminus) is favored and the resulting spectrum is likely to be easy to interpret. For this reason, ESI has been a popular choice for peptide sequencing. MS/MS sequencing of tryptic peptides of allergens allowed us to identify isoform sequences as well as post-translational modifications.

Sequencing of tryptic peptides from the natural and recombinant allergens were performed with the Waters capillary HPLC coupled to the Micromass (Waters) Global Ultima QTOF spectrometer. Instrument control and data acquisition were done with MassLynx 4.1 (Micromass/Waters); data analyses, sequencing and data-bank searches were performed using ProteinLynx GlobalServer 2.1 (Micromass/Waters). Tryptic peptides were obtained after in-solution digestion or after SDS-PAGE and in-gel digestion.

Analysis of tryptic peptides by LC-MS/MS involved several steps: a) loading and washing the peptides on a precolumn, b) gradient elution from the analytical column

and acquisition of MS and MS/MS data and finally c) computational data analysis. In our experiments, first the peptides were dissolved in 0.1% formic acid and loaded onto a small trap column (Waters Symmetry300 C₁₈, 5 µm, 0.18 x 23.5 mm). The column was washed for several minutes at 20 µl flow rate. During this phase, the trap column was not coupled to the analyzing column. Salts and other low molecular weight contaminations originating from the SDS gel do not bind to the column material and are flushed from the trap column. In the second phase, a switch valve connected the trap column to the analyzing column. The peptides were back-flushed to the analyzing column and eluted with an acetonitrile gradient (solvent A 0.1% v/v formic acid / 5% v/v acetonitrile, solvent B 0.1% v/v formic acid / 95% v/v acetonitrile; 5-40% B in 40 minutes, followed by 40-95% B in 5 minutes). A Waters Atlantis dC₁₈ capillary column (3 mm, 75 mm x 150 mm) was used. The required flow rate was of approximately 300 nl/min was achieved by a T-split. Peptides eluting from the column were ionized online by electrospray ionization (ESI) using the Waters nanoflow spray head or, when necessary, NewObjective fused silica needles that provided higher sensitivity.

Data acquisition was done in 2 alternating modes of the instrument. In the survey (MS) mode, the quadrupole of the spectrometer was deactivated and ions were analyzed by time-of-flight spectrometry without fragmentation. The survey mode gives an overview of all ions eluting from the column at a given time. If the software recognized signals of doubly charged ions typical for peptides, the instrument switched from the survey into the sequencing (MS/MS) mode. In the sequencing mode, the quadrupole filters and removes all ions except the peptide ions to be sequenced. The filtered ions were then fragmented in the collision cell and the fragment ions were analyzed by time-of-flight spectrometry.

In the final step, the raw data were analyzed and peptide sequences were established from the fragment patterns. A custom-made database containing published isoform sequences and sequences from the allergens analyzed in the CREATE was established. Experimental sequences were compared to this database. If sequences could not be assigned to known isoform sequences, *de novo* sequencing and BLAST sequence homology searches were performed and allowed us in some cases to identify new isoforms in the natural allergen preparations. The MS/MS experiments also provided information on the presence and possible nature of post-translational modifications.

4. High-performance Liquid Chromatography – Size Exclusion Chromatography (HPLC-SEC)

Separation of molecules using gel filtration is a simple and reliable method to analyze biomolecules on the basis of their relative size. Size exclusion chromatography provides information not only about the molecular weight but also on the aggregation status of the compounds in solution [11]. However, detection of aggregates might not be possible depending on the amounts present in the preparation or might go undetectable due to interactions with the column material. In the CREATE project, purified natural and recombinant allergen preparations were analyzed by SEC

using a HP 1100 HPLC system (Agilent Technologies, Palo Alto, CA, USA) with a TSK-Gel G2000SWXL column (Tosoh Biosep, 7.8mm ID x 30.0cm L). After equilibrating the column with sodium phosphate buffer saline pH 6.7 containing 0.05% (w/v) NaN_3 , 20-25 μg of purified protein samples were loaded at a flow rate of 0.5 ml/min and UV absorbance was monitored at 280 nm. In separate runs, the retention times of bovine serum albumin (M_r 66,000), alpha-chymotrypsin (M_r 25,000), lysozyme (M_r 14,300), and bombesin (M_r 1,600) were used as molecular weight standards. On the basis of the elution profile the approximate molecular weight and the homogeneity of the protein preparation was determined.

5. Circular Dichroism

Circular dichroism (CD) analysis of proteins is based on the differential absorption of left and right circularly polarized light in the far UV region by chiral chromophores. In proteins, mainly peptide bonds serve as chromophores. Absorption characteristics of peptide bonds are measurable depending on the environment and therefore on the secondary structure surrounding the chromophore. Compared to other structural technique such as NMR or X-ray crystallography, CD serves as a low-resolution method for the determination of secondary structure elements of proteins. Its main advantages are the little demands on both time and sample; yet CD spectra give valuable information on the folding of proteins [12, 13]. Therefore, in the CREATE project a protocol was set up to compare a series of protein pairs (natural and recombinant) in terms of their secondary structures and thermostability. Measurements were performed with a JASCO J-810 spectropolarimeter fitted with a Neslab RTE-111M temperature control system. Far UV CD spectra were recorded at 20°C, in a wavelength range was between 190-260 nm. Each spectrum represented an average of five consecutive scans. The final spectra were obtained after subtraction of the baseline recorded under identical conditions. All measurements were performed using 5mM sodium phosphate pH 7.4 as buffer system. The working protein concentrations ranged from 0.1 $\mu\text{g}/\mu\text{l}$ to 0.2 $\mu\text{g}/\mu\text{l}$ depending on the molecules and were exactly determined for each protein by amino acid analysis. All data were represented as mean residue molar ellipticity $[\theta]_{\text{MRW}}$ at a given wavelength. Further proteins were heat denatured in a temperature range of 25°C to 95°C at a temperature slope of 1°C/min and renatured under the same conditions. Folding and unfolding were analysed by monitoring CD spectra at the given wavelength of 222 nm. Several allergen pairs were identified where natural and recombinant proteins behaved similarly following the measurement parameters described above. However, our analysis also identified recombinant proteins showing significant differences in their folding when compared to natural counterparts. Thus, CD is an important technique for characterization of recombinant and natural allergens with respect to folding, thermostability, and for batch-to-batch control of the preparations.

6. Small-Angle X-ray Scattering (SAXS)

The phenomenon of small-angle X-ray scattering (SAXS) resembles the familiar situation when a beam of visible light is scattered by a colloidal suspension like milk or when the front lights of a car are scattered in the mist [14]. Clear solutions of non-

aggregated proteins, in contrast, require an electromagnetic beam of much smaller wavelength to interact with. Therefore, monochromatic X-rays are guided through the protein solution and the scattering curve [intensity x scattering vector] is recorded. Using the mathematical algorithm indirect Fourier transform, the pair distance distribution function [distance frequency x dimension] was calculated from the scattering curve [15]. The pair distance distribution function (PDDF) expresses the frequency of intramolecular distances between electrons. Thus, conclusions on the molecular size, shape, and aggregation behaviour of proteins in solution can be drawn [16]. The PDDF of a spherical molecule, for instance, shows a maximum at a distance corresponding to the radius and meets the x-axis at the maximal dimension of the molecule thus resembling a normal Gaussian curve. Extended ellipsoids or rod-like aggregates, on the other hand, give asymmetric PDDFs with maxima at rather low dimensions and less steep slopes to larger maximal dimensions. Therefore, molecular information on size and shape of proteins in solution can be derived from PDDFs. Assuming a theoretical volume of 20 nm^3 for a spherical 10 kDa protein its aggregation state in solution can be estimated from the maximal dimensions of the PDDF. A maximal dimension of 14 nm referring to a globular protein with a molecular weight of approximately 700 kDa was taken as limit to certify a non-aggregated state of the allergen preparation.

For the CREATE samples, measurements were performed with the SAXSess camera (Anton Paar KG, Graz, Austria) attached to a conventional copper X-ray generator (Philips, Eindhoven, Netherlands). The allergen solutions were measured in a thermally controlled, (25°C) SAXSess standard quartz capillary with an outer diameter of 1 mm and wall thickness of $10 \mu\text{m}$. After 60 minutes of exposure and 5 minutes fading, intensities were monitored at scattering vectors from 0.125 to 4 nm^{-1} using a BAS-1800 scanner (Fujifilm, Tokyo, Japan) and processed with the SAXSQuant software (Anton Paar KG, Graz, Austria). By recording the scattering curve between 0.125 and 4 nm^{-1} , the most significant detection range of the PDDF was set at 0.7 to 25 nm , which is the ideal size range to investigate the aggregation behavior of proteins in solution. For control, the buffers for each allergen preparation were measured accordingly and the scattering curves subtracted. Data evaluation was done by indirect Fourier transform algorithm [15] using the PCG software package (Karl-Franzens University of Graz).

7. References

1. van Ree R. Analytic aspects of the standardization of allergenic extracts. *Allergy* 1997; 52: 795-805.
2. Helm RM, Burks AW. Allergenic products standardization--the role of major allergens. *Arb Paul Ehrlich Inst Bundesamt Sera Impfstoffe Frankf A M* 1997; 91: 164-173.
3. van Ree R, Dorpema JW, Vieths S. Allergy vaccines: A need for standardisation in mass units of major allergen. *Pharmeuropa Bio* 2005; 1: 27-30.
4. Løwenstein H, Larsen JN. Recombinant allergens/allergen standardization. *Curr Allergy Asthma Rep* 2001; 1: 474-479.
5. van Ree R. The CREATE project: A new beginning of allergen standardization based on mass units of major allergens. *Arb Paul Ehrlich Inst Bundesamt Sera Impfstoffe Frankf A M* 2003; 94: 70-73; discussion 74-75.
6. van Ree R. The CREATE project: EU support for the improvement of allergen standardization in Europe. *Allergy* 2004; 59: 571-574.
7. Chirino AJ, Mire-Sluis A. Characterizing biological products and assessing comparability following manufacturing changes. *Nat Biotechnol* 2004; 22: 1383-1391.

8. Tyler MI. Amino acid analysis. An overview. *Methods Mol Biol* 2000; 159: 1-7.
9. Reinders J, Lewandrowski U, Moebius J, Wagner Y, Sickmann A. Challenges in mass spectrometry-based proteomics. *Proteomics* 2004; 4: 3686-3703.
10. Lill J. Proteomic tools for quantitation by mass spectrometry. *Mass Spectrom Rev* 2003; 22: 182-194.
11. Goetz H, Kuschel M, Wulff T, Sauber C, Miller C, Fisher S, Woodward C. Comparison of selected analytical techniques for protein sizing, quantitation and molecular weight determination. *J Biochem Biophys Methods* 2004; 60: 281-293.
12. Kelly SM, Jess TJ, Price NC. How to study proteins by circular dichroism. *Biochim Biophys Acta* 2005; 1751: 119-139.
13. Verdino P, Keller W. Circular dichroism analysis of allergens. *Methods* 2004; 32: 241-248.
14. Pilz I, Glatter O, Kratky O. Small-angle X-ray scattering. *Methods Enzymol* 1979; 61: 148-249.
15. Glatter O. A new method for the evaluation of small angle scattering data. *J Appl Cryst* 1977; 10: 415-421.
16. Svergun DI, Koch MH. Advances in structure analysis using small-angle scattering in solution. *Curr Opin Struct Biol* 2002; 12: 654-660.

8. Acknowledgments

The purified allergen preparations were supplied by:

Markus Susani, Biomay AG, Vienna Austria

Rafael Monsalve, ALK-Abelló, Madrid, Spain

Helmut Fiebig and Oliver Cromwell, Allergopharma, Reinbeck, Germany

Wolf-Meinhard Becker, Forschungszentrum Borstel, Borstel, Germany

Mayte Villalba, Universidad Complutense de Madrid, Madrid, Spain

Lisa D. Vailes and Martin D. Chapman, Indoor Biotechnologies, Cardiff, UK

This work was supported by the EU grant GRD2-2000-30032 (CREATE).

Discussions

Seppälä: I have a question regarding the primary sequence analysis by mass spectrometry. Did you reduce and alkylate those molecules and did you map the disulphide bridges from those molecules to check the folding?

Ferreira: We did the first runs without any pre-treatment, just to see what kind of peptides we were able to identify. And afterwards, except for the Bet v 1 which does not contain any cysteins, all preparations were reduced, alkylated and digested before they were subjected to the whole peptide mapping procedure.

Seppälä: Did you map the disulphide bridges?

Ferreira: In principal, we could at least be able to identify some, because we have performed the experiments before and after pre-treatment. But I have to go back to the data to analyse and specifically search for possible peptides involved in disulphide bridges.

Gavrovic-Jankulovic: I am interested in the expression system that has been used, especially for Phl p 1. Was it *E.coli*?

Ferreira: Phl p 1 was expressed in an *E.coli* T7-based expression vector.

Banik: Do you want to comment on why your recombinants are oligomers?

Ferreira: That was nothing we were expecting. It actually just came because we applied this new technology; that could give some indication. But of course under the conditions used for the SAXS analysis - high concentrations of proteins - some proteins may aggregate. But we always had the unique opportunity to compare a natural and a recombinant preparation and we did not see the same behaviour under the same conditions for the natural preparations. So, it seems to be a particular feature of the recombinant preparations. I also believe that some procedures in the manufacturing process have an influence on the outcome. We have, for example, seen differences in the Bet v 1 preparations between batch 16 and other batches. It was the same protein and the same expression vector, just the purification method was different and batch 16 oligomerized and multimerized, whereas the other batches did not. So I think we should have a close look at the manufacturing procedures. I would not want to make the statement that recombinants necessarily have the tendency to multimerize. You have seen that we had preparations where this did not happen, and we had preparations where simply by a different purification procedure we could also avoid this.