HAL open science

# Guidance on allergenicity assessment of genetically modified plants

Hanspeter Naegeli, Andrew Nicholas Birch, Josep Casacuberta, Adinda de Schrijver, Mikolaj antoni Gralak, Philippe Guerche, Huw Jones, Barbara Manachini, Antoine Messéan, Elsa Ebbesen Nielsen, et al.

**To cite this version:**

Hanspeter Naegeli, Andrew Nicholas Birch, Josep Casacuberta, Adinda de Schrijver, Mikolaj antoni Gralak, et al.. Guidance on allergenicity assessment of genetically modified plants. EFSA Journal, 2017, 15 (6), 10.2903/j.efsa.2017.4862 . hal-01605253

**HAL Id: hal-01605253**

**https://hal.science/hal-01605253v1**

Submitted on 26 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

SCIENTIFIC OPINION

EFSA Journal

ADOPTED: 18 May 2017

doi: 10.2903/j.efsa.2017.4862

# Guidance on allergenicity assessment of genetically modified plants

EFSA Panel on Genetically Modified Organisms (GMO),
Hanspeter Naegeli, Andrew Nicholas Birch, Josep Casacuberta, Adinda De Schrijver, Mikolaj Antoni Gralak, Philippe Guerche, Huw Jones, Barbara Manachini, Antoine Messéan, Elsa Ebbesen Nielsen, Fabien Nogué, Christophe Robaglia, Nils Rostoks, Jeremy Sweet, Christoph Tebbe, Francesco Visioli, Jean-Michel Wal, Philippe Eigenmann, Michelle Epstein, Karin Hoffmann-Sommergruber, Frits Koning, Martinus Lovik, Clare Mills, Francisco Javier Moreno, Henk van Loveren, Regina Selb and Antonio Fernandez Dumont

## Abstract

This document provides supplementary guidance on specific topics for the allergenicity risk assessment of genetically modified plants. In particular, it supplements general recommendations outlined in previous EFSA GMO Panel guidelines and Implementing Regulation (EU) No 503/2013. The topics addressed are non-IgE-mediated adverse immune reactions to foods, *in vitro* protein digestibility tests and endogenous allergenicity. New scientific and regulatory developments regarding these three topics are described in this document. Considerations on the practical implementation of those developments in the risk assessment of genetically modified plants are discussed and recommended, where appropriate.

© 2017 European Food Safety Authority. *EFSA Journal* published by John Wiley and Sons Ltd on behalf of European Food Safety Authority.

**Keywords:** guidance, allergenicity assessment, newly expressed proteins, endogenous allergenicity, GMO

**Requestor:** EFSA

**Question number:** EFSA-Q-2014-00547

**Correspondence:** gmo@efsa.europa.eu

**Panel members:** Andrew Nicholas Birch, Josep Casacuberta, Adinda De Schrijver, Mikolaj Antoni Gralak, Philippe Guerche, Huw Jones, Barbara Manachini, Antoine Messéan, Hanspeter Naegeli, Elsa Ebbesen Nielsen, Fabien Nogué, Christophe Robaglia, Nils Rostoks, Jeremy Sweet, Christoph Tebbe, Francesco Visioli and Jean-Michel Wal.

**Acknowledgements:** The Panel wishes to thank the hearing expert John Mclaughlin, and the EFSA staff members Claudia Paoletti and Elisabeth Waigmann for the support provided to this scientific opinion.

**Suggested citation:** EFSA GMO Panel (EFSA Panel on Genetically Modified Organisms), Naegeli H, Birch AN, Casacuberta J, De Schrijver A, Gralak MA, Guerche P, Jones H, Manachini B, Messéan A, Nielsen EE, Nogué F, Robaglia C, Rostoks N, Sweet J, Tebbe C, Visioli F, Wal J-M, Eigenmann P, Epstein M, Hoffmann-Sommergruber K, Koning F, Lovik M, Mills C, Moreno FJ, van Loveren H, Selb R and Fernandez Dumont A, 2017. Guidance on allergenicity assessment of genetically modified plants. EFSA Journal 2017;15(5):4862, 49 pp. https://doi.org/10.2903/j.efsa.2017.4862

**ISSN:** 1831-4732

© 2017 European Food Safety Authority. *EFSA Journal* published by John Wiley and Sons Ltd on behalf of European Food Safety Authority.

This is an open access article under the terms of the Creative Commons Attribution-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited and no modifications or adaptations are made.

The EFSA Journal is a publication of the European Food Safety Authority, an agency of the European Union.

## Summary

Following a request from the European Food Safety Authority (EFSA) to the Panel on Genetically Modified Organisms (GMO Panel), a Working Group was established to develop supplementary guidance for the allergenicity assessment of genetically modified (GM) plants.

This EFSA GMO Panel document provides supplementary guidance for the risk assessment of GM plants and derived food and feed, submitted within the framework of Regulation (EC) No 1829/2003. It supplements the EFSA GMO Panel guidance document on risk assessment of food and feed from GM plants published in 2011 and Implementing Regulation EU (No) 503/2013. The purpose of this document is to provide detailed guidance to assist the applicant in the preparation and presentation of an application according to such Regulation.

In particular, this document addresses three main topics: (i) non-IgE-mediated adverse immune reactions to food; (ii) *in vitro* protein digestibility tests; and (iii) endogenous allergenicity.

New scientific and regulatory developments on these three topics are described. Considerations regarding the practical implementation of those developments in the risk assessment of GM plants are discussed and, where appropriate recommendations made to supplement previous guidance documents.

Briefly, for non-IgE-mediated adverse immune reactions to food detailed risk assessment considerations are provided to determine the safety profile of the protein or peptide under assessment with regard to its potential to cause celiac disease. This assessment will include available information on the source of the transgene and on the protein itself as well as on data from *in silico* and *in vitro* testing, as and when appropriate.

For *in vitro* protein digestibility tests, the EFSA GMO Panel considers that additional investigations are needed before any additional recommendation in the form of guidance for applicants can be provided. To this end, an interim phase is considered necessary to evaluate the revisions to the *in vitro* gastrointestinal digestion test, proposed by EFSA, which are presented in an Annex to this document.

For assessing endogenous allergenicity of GM plants and to support the practical implementation of mandatory requirements in Implementing Regulation EU (No) 503/2013, this guidance document provides further information on: (i) relevant crops subjected to such analysis; (ii) relevant allergens that should be quantified; (iii) methodology to be used for quantification; and (iv) principles to be followed for data interpretation and risk assessment considerations.

During the development of this document, EFSA involved stakeholders and the general public at different stages, strengthening new means of engagement in its scientific process.

# Table of contents

Abstract ..... 1
Summary ..... 3
1. Introduction ..... 5
1.1. Background as provided by EFSA ..... 5
1.2. Terms of Reference as provided by EFSA ..... 5
1.3. Objectives ..... 5
1.4. Scope ..... 5
1.5. Transition period ..... 6
2. Allergenicity assessment ..... 6
2.1. Non-IgE-mediated adverse immune reactions to foods ..... 7
2.1.1. Celiac disease ..... 7
2.1.2. Risk assessment considerations ..... 8
2.2. *In vitro* protein digestibility tests ..... 11
2.2.1. Risk assessment considerations ..... 11
2.3. Endogenous allergenicity ..... 12
2.3.1. Relevant crops for analysis ..... 13
2.3.2. Relevant allergens for quantification ..... 13
2.3.3. Methodology for quantification ..... 14
2.3.4. Data interpretation and risk assessment ..... 14
Documentation provided to EFSA ..... 15
References ..... 15
Abbreviations ..... 16
Annex A – Non-IgE-mediated adverse immune reactions to foods ..... 18
Annex B – *In vitro* protein digestibility tests ..... 29
Annex C – Endogenous allergenicity ..... 46

# 1. Introduction

## 1.1. Background as provided by EFSA

Allergenicity assessment of genetically modified (GM) plants is performed following the recommendations laid down in the EFSA Guidance Document (2011). These recommendations are mainly based on considerations from the EFSA GMO Panel (2010) Scientific Opinion on allergenicity assessment of GM plants and microorganisms, and derived food and feed.

In 2012, the European Food Safety Authority (EFSA) launched a procurement call entitled: 'Literature reviews on: (i) non-IgE-mediated adverse immune reactions to foods, and (ii) *in vitro* digestibility tests for allergenicity assessment'. The aim of the project was to obtain relevant information related to these two topics to be used as background information for further discussion within the EFSA Panel on Genetically Modified Organisms (GMO Panel). The review on non-IgE-mediated adverse immune reactions to food identified relevant methodology (i.e. *in silico* and *in vitro*) that could be applied in the allergenicity assessment process (Mills et al., 2013a). The review dealing with *in vitro* digestibility testing for allergenicity assessment highlighted the need for better standardisation and harmonisation of the conditions used (e.g. pHs, enzyme:substrate ratios, controls) when performing *in vitro* digestibility studies (Mills et al., 2013b).

In addition, the new Implementing Regulation (EU) No 503/2013[1] (IR503/2013) on applications for authorisation of GM food and feed has been in place since December 2013. This most recent regulation includes certain allergens (as defined in OECD Consensus documents) in the compositional analysis, and consequently, the requirement for quantitative measurement of individual allergens. The development of supplementary guidelines on this topic would be useful to assist both applicants and risk assessors in the practical implementation of this requirement.

Therefore, the EFSA GMO Panel was of the opinion that supplementary guidelines on allergenicity assessment are needed to incorporate new developments in the area into the risk assessment process.

## 1.2. Terms of Reference as provided by EFSA

The tasks of the Working Group of the GMO Panel are (i) to develop supplementary guidelines for the allergenicity assessment of GM plants; (ii) to participate in a workshop with stakeholders organised by EFSA; (iii) to consult the public on the draft Scientific Opinion; and (iv) to review and revise the draft Scientific Opinion accordingly.

## 1.3. Objectives

This guidance document is designed to assist applicants in the preparation and presentation of a well-structured application to demonstrate the safety of the GM plant under assessment, with respect to the allergenicity risks. Recommendations are also provided for the correct interpretation of the data in the risk assessment process.

EFSA will continue to review the state-of-the-art in science and in the light of experience gained from the evaluation of GM plant applications, updating the guidance document, as and when appropriate.

## 1.4. Scope

This document provides supplementary guidance for the risk assessment of GM plants and derived food and feed, submitted within the framework of Regulation (EC) No 1829/2003[2]. It supplements the Guidance Document on risk assessment of food and feed from GM plants (EFSA GMO Panel, 2011) and IR503/2013.

The supplementary Guidance Document addresses three main topics: (i) non-IgE-mediated adverse immune reactions to food; (ii) *in vitro* protein digestibility tests; and (iii) endogenous allergenicity.

[1] Commission Implementing Regulation (EU) No 503/2013 of 3 April 2013 on applications for authorisation of genetically modified food and feed in accordance with Regulation (EC) No 1829/2003 of the European Parliament and of the Council and amending Commission Regulations (EC) No 641/2004 and (EC) No 1981/2006. Official Journal of the European Union L157, p. 1–48.

[2] Regulation (EC) No 1829/2003 of the European Parliament and of the Council of 22 September 2003 on genetically modified food and feed. OJ L 268, 18.10.2003, 1–23.

The purpose of this document is to provide detailed guidance to assist the applicant in the preparation and presentation of an application, according to Articles 5(8) and 17(8) of Regulation (EC) No 1829/2003. Specific guidance on the submission of an application will be revised by EFSA, accordingly.

### 1.5. Transition period

The transition period to allow for the implementation of the EFSA recommendations – the precise time point from the time of adoption of this document when the requirements laid down in this guidance document will be fully applicable to newly submitted applications – is as follows.

For non-IgE-mediated adverse immune reactions to food, a 6-month transition period is considered appropriate. While a period of 3 months would have been in line with the indicative timelines for the applicant to submit updated bioinformatic analysis (EFSA, 2014), an overall 6-month period is considered adequate because a new algorithm, even if simple, will be needed.

For endogenous allergenicity, if plant material needs to be generated for testing, a 24-month transition period is considered appropriate (in line with indicative timelines for the applicant to submit similar requests, EFSA 2014). If plant material does not need to be generated, a 12-month period is considered appropriate for the applicants to fully align the strategy for the selection and measurement of relevant allergens as recommended in this EFSA guidance document. It is noted that assessment of endogenous allergenicity is currently required in IR503/2013 and that the present document provides additional considerations regarding the practical implementation of such a requirement. Based on the experience gained reviewing applications submitted under IR503/2013, many relevant allergens have already been selected and measured by applicants.

For the remaining topic on *in vitro* protein digestibility testing, the EFSA GMO Panel considers that additional investigation is needed before any further recommendation in the form of guidance for applicants can be provided. To this end, an interim phase is considered necessary to evaluate the revisions proposed by EFSA to the *in vitro* gastrointestinal digestion test, which is presented in Annex B. During this interim phase, the laboratory(ies) involved will further detail and apply the refined digestion test methodology proposed by the EFSA GMO Panel. After this period, EFSA will evaluate whether the test adds value and, if so, what further steps are needed for its final implementation in the form of guidance for applicants, which will focus only on *in vitro* protein digestibility testing. During such interim phase and until the evaluation of the new approach is completed, the EFSA GMO Panel will continue to follow the weight-of-evidence approach for allergenicity assessment as described by the EFSA (EFSA GMO Panel, 2011) and Codex Alimentarius (Codex Alimentarius, 2003, 2009).

## 2. Allergenicity assessment

Food allergies represent an important public health problem affecting approximately 2–4% of the adult population and up to 8–9% of children (medically diagnosed), the prevalence rate for self-reported food allergy being several times higher (EFSA NDA Panel, 2014). Essentially, the only way to avoid triggering reactions in individuals who are already allergic is avoidance of the relevant food(s). Nevertheless, it has been previously noted that in everyday life strict avoidance of specific foods is difficult to achieve (Crevel et al., 2008) and not completely effective in preventing allergic reactions (Madsen et al., 2010; Fernandez et al., 2013).

Concerning the potential allergenicity of (novel) proteins, there is in practice no possibility to ensure full certainty, as to the absence of allergenic risk. There is no single test or parameter that, on its own, can provide sufficient evidence to predict the allergenicity of a protein or peptide. This is because our understanding of what causes a protein or a peptide to become allergenic in susceptible individuals is incomplete. Furthermore, the development of allergic disease depends not only on the allergen but also on genetic predisposition of the individual and other environmental factors. Nevertheless, a high degree of confidence in genetically modified organisms (GMOs) safety can be reached using a weight-of-evidence approach (EFSA, 2006; Codex Alimentarius 2009; EFSA GMO Panel, 2011). Importantly, this approach must be based on the best and most up-to-date scientific knowledge and methodologies. The field of molecular biology is rapidly developing and consequently regulations and guidance documents need to be updated, as and when appropriate, to take scientific advances on board and to reduce remaining uncertainty after a weight-of-evidence evaluation.

To this end, and following the outcome of an EFSA procurement regarding literature reviews on (i) non-IgE-mediated adverse immune reactions to food and (ii) *in vitro* digestibility tests for allergenicity assessment of the newly expressed protein, the EFSA GMO Panel identified new scientific

information to be considered in the allergenicity assessment of GMOs. In addition, the more recent IR503/2013 on applications for authorisation of GM food and feed included the mandatory measurement of certain allergens in the compositional analysis of GM plants. The development of a supplementary guidance document on the topic of allergenicity was considered necessary to assist both applicants as well as risk assessors.

Within this context, in 2014 a Working Group of the EFSA GMO Panel was established to develop a supplementary guidance document on allergenicity assessment of GMOs focusing on three topics: (i) non-IgE-mediated adverse immune reactions to food; (ii) *in vitro* protein digestibility testing; and (iii) endogenous allergenicity in the recipient plant. A stakeholder meeting to provide early input to the guidance development was held in Brussels in June 2015.[3] To further secure timely feedback to the EFSA GMO Panel Working Group during the guidance development, a 'Focus group' was established.[4] Additional engagement with stakeholders took place during a public consultation,[5] and a subsequent EFSA info session to further address comments received.[6]

## 2.1. Non-IgE-mediated adverse immune reactions to foods

Non-IgE-mediated adverse immune reactions to antigenic food components comprise a large group of diseases, mostly occurring during childhood. Of these, the best characterised diseases include food protein-induced enterocolitis (FPIES), as well as eosinophilic diseases of the gastrointestinal tract, where food products play a role in the pathogenesis (eosinophilic oesophagitis, proctocolitis) (see Annex A). However, the exact pathogenic mechanisms of these diseases are insufficiently understood, and the diagnosis mostly relies on positive food challenges. Thus, insights about the food components involved on the molecular level and knowledge on clearly recognised immune mechanisms for these diseases are currently lacking.

In contrast, celiac disease (CD) is a well characterised non-IgE-mediated adverse immune reaction to food, and the food proteins involved were described (Koning et al., 2015; van Bergen et al., 2015; Annex A). Here, gluten has been identified as the environmental trigger initiating the immune reaction. The involvement of the immune system in the disease is well established, as pro-inflammatory T cells specific for gluten fragments bound to the disease-predisposing HLA-DQ2 or HLA-DQ8 molecules are typically present in the inflamed intestine of patients. This T-cell response also leads to the production of IgA autoantibodies specific for tissue transglutaminase. CD diagnosis is based on the examination of characteristic histopathological changes in small intestinal biopsies in combination with serological tests with positive IgA against tissue transglutaminase being the most reliable.

Consequently, at the present time, assessment of newly expressed proteins with regard to non-IgE-mediated adverse immune reactions should focus only on CD. For other non-IgE-mediated adverse immune reactions to foods than CD, additional knowledge on the pathogenic mechanisms is necessary before they can be considered into the allergenicity assessment.

### 2.1.1. Celiac disease

CD is a disease of the small intestine characterised by flattening of the intestinal mucosa, resulting in a variety of clinical symptoms including malabsorption, failure to thrive, diarrhoea and stomach ache. A detailed description of the pathogenesis of the disease can be found in Annex A.

Briefly, the disease is caused by an uncontrolled intestinal immune response to gluten proteins in wheat (*Triticum* spp), gluten-like hordeins in barley (*Hordeum vulgare)* and secalins in rye (*Secale cereale)* (Green and Cellier, 2007). Oat (*Avena sativa)* is generally considered safe for patients (Garsed and Scott, 2007), although exceptions were reported (Lundin et al., 2003; Comino et al., 2015). The only available treatment is a lifelong gluten-free diet implying the exclusion of all food products that contain wheat, barley and rye or gluten and gluten-like proteins from these cereals. CD affects approximately 1% of the world population (Abadie et al., 2011).

[3] http://www.efsa.europa.eu/en/supporting/pub/899e
[4] http://www.efsa.europa.eu/sites/default/files/assets/shp_dg_guidance_document_allergenicity.pdf
[5] http://www.efsa.europa.eu/en/consultations/call/160726
[6] http://www.efsa.europa.eu/en/events/event/161123

### 2.1.2. Risk assessment considerations

A large number of methods and tests can be used to investigate the potential (detrimental) properties of proteins and peptides under assessment with regard to CD, but in practice it will not be necessary to apply this full array to safeguard CD patients.

Rather, an integrated, stepwise, case-by-case approach should be used in the assessment of the newly expressed protein(s) in relation to its(their) potential to cause CD. This overall strategy is in line with the general principles followed for the allergenicity assessment of newly expressed protein(s) as defined by EFSA (EFSA GMO Panel, 2011) and Codex Alimentarius (2009). In this context and briefly, the first step in the assessment should consider the available information on the source of the protein and on the human exposure to the protein itself. This knowledge on the protein can be used to calibrate the risk assessment strategy to follow on a case-by-case basis. If the available knowledge on the protein under assessment is insufficient to support its safety, additional considerations on its properties are necessary to investigate the potential to cause CD. To this end, *in silico* approaches can be employed, starting with searches for sequence identity (e.g. searches with known CD peptide sequences and motif searches). In a second step, if concerns from the sequence identity search are present, *in silico* peptide modelling can be applied. When potentially CD relevant sequences which cannot be disregarded by *in silico* testing are identified, in a third step other *in vitro* tests such as HLA-DQ-peptide binding assays and/or testing with T-cell clones derived from patients with CD (Figures 1 and 2) can be performed to determine the safety profile of the protein/peptide under assessment. Further details on the approach to follow can be found below.

**A) Step 1 – knowledge on the protein/searches for sequence identity**

Knowledge on the protein

Knowledge on the protein should include detailed information on the source of the transgene as regard its use as food and solid documented information that the protein itself is consumed by individuals with CD without causing the disease. An exposure assessment might be necessary to support this consideration.

If the knowledge on the protein is unavailable or insufficient to support its safety, additional considerations are needed as described below (see also Figure 1).

Searches for sequence identity

*a) Searches involving a perfect sequence match with known CD peptide sequences*

It is well established that prolamins and closely related proteins (Shewry et al., 2003) harbour the sequences that cause CD (Tye-Din et al., 2010). Therefore, an initial consideration is to determine if the protein of interest belongs to these families of proteins. Thus, identity searches with known CD peptide sequences (see Annex A-1) should be performed. If this search results in a perfect match with a peptide sequence known to cause CD, a hazard has been identified.

If this is not the case and there is insufficient knowledge on the protein under assessment, further investigations to identify/dismiss proteins/peptides that could potentially cause CD should be provided (Figures 1 and 2).

*b) Searches involving a partial sequence match with known CD peptide sequences*

T cells respond to antigenic peptides bound to an HLA-molecule. In the case of CD, it concerns gluten-derived peptides bound to either HLA-DQ2 or -DQ8. However, it is well known that such T cells can also respond to peptides in which one or more amino acids are replaced. This is the basis, for example, for T-cell cross-reactivity towards gluten peptides and homologous, but not identical, peptides derived from barley and rye. For this reason, antigenicity of a protein or fragments thereof for patients with CD cannot be excluded only on the basis of a lack of a perfect match with gluten sequences.

– *Searches for sequence identity with the Q/E-X1-P-X2 motif (Figure 3):*

Examination of the list of epitopes currently identified (see Annex A-1) reveals that a characteristic Q-X1-P-X2 motif is present in the large majority of HLA-DQ2 epitopes. This motif is a target for the enzyme tissue transglutaminase 2 (TG2) which yields E-X1-P-X2 (X1 = L, Q, F, S or E; X2 = Y, F, A, V or Q; for further details see Figure 3 and Annex A).

Additional considerations on the position and nature of adjacent amino acid sequences to the Q/E-X1-P-x2 motif should also be taken into account when performing the assessment (examples are provided in Annex A-2). If concerns are raised, additional tests will be required (Figures 1 and 2).

– *Searches for sequence identity without the Q/E-x1-P-x2 motif:*

Few known CD peptide sequences do not contain the Q/E-X1-P-X2 motif. Therefore, an identity search with known CD peptide sequences, which do not contain a Q/E-X1-P-X2 motif, should be performed allowing one or more amino acid mismatches. The position and nature of the mismatched and identical amino acids determine if the peptide sequence has the potential to be a T-cell stimulatory epitope (additional considerations and examples are provided in Annex A-2). If concerns are raised, additional tests will be required (Figures 1 and 2). Known CD peptide sequences are listed in Table A.1 as well as in publicly available celiac peptide databases (see Annex A-1).

Following these searches for sequence identity, four outcomes are possible:

- if the Q/E-X1-P-X2 motif is not present and no concerns are raised during the identity search with specific known CD peptide sequences, the probability of a T-cell epitope is unlikely;
- if the Q/E-X1-P-X2 motif is not present but concerns are raised during the identity search with specific known CD peptide sequences, further investigation is required;
- if the Q/E-X1-P-X2 motif is present and considerations on adjacent sequences can be used to eliminate concerns, the probability of a T-cell epitope is unlikely;
- if the Q/E-X1-P-X2 motif is present and considerations on adjacent sequences cannot be used to eliminate concerns, a potential T-cell epitope is detected and further investigation is required.

Further details on the search for sequence identity are listed in Annex A-2.

**B) Step 2 – HLA-DQ-peptide modelling**

Several HLA-DQ2-gliadin and HLA-DQ8-gliadin structures are publicly available. These structures can be used to model a peptide of interest into HLA-DQ2 or HLA-DQ8. This HLA-DQ-peptide modelling can then allow for a comparison, which can indicate the likelihood that such a peptide will bind to HLA-DQ2 or HLA-DQ8 and will provide insights into the position and orientation of the T-cell receptor contact residues in the HLA-DQ-bound peptide.

Two outcomes are possible:

- if no relevant HLA-DQ binding and/or similarity with the available HLA-DQ-gliadin structures are predicted, the probability of a T-cell epitope is unlikely;
- if HLA-DQ binding and a high degree of similarity with the available HLA-DQ-gliadin structures are predicted, this indicates potential cross-reactivity of the investigated peptide. Consequently, the potential capacity of the peptide to trigger CD should be determined by additional *in vitro* approach(es) described below.

Further details on this step are listed in Annex A-3.

**C) Step 3 – *In vitro* approaches**

Proposals of *in vitro* approaches that can be used to further investigate the potential of the newly expressed protein to cause CD are found below.

*HLA-DQ peptide binding assays*:

For peptides to evoke T-cell responses, they must bind to HLA-molecules. HLA-DQ2- and HLA-DQ8-specific peptide binding assays were developed and can be exploited to determine the likelihood that peptides under investigation might be immunogenic.

Two outcomes are possible:

- if high affinity binding is detected (see Annex A-4), further testing is required;
- if low or no affinity binding is detected, the probability that the peptide is immunogenic is low. Therefore, no further testing is required.

For further details, please see Annex A-4 for an overview of publications that have reported HLA-DQ-peptide binding assays.

*T-cell testing:*

Recognition of gluten peptides by CD4+ T cells from one or more CD patients has been a prerequisite for defining toxic CD peptides (Sollid et al., 2012).

Such T cells were isolated in a number of laboratories where the necessary expertise and appropriate infrastructure are available. These T cells were used to provide conclusive evidence on the capacity of a specific peptide sequence to stimulate CD-causative T-cell responses. If a T-cell response with the protein/peptide under assessment is observed, then a hazard has been identified.

For further details, please see Annex A-5 for an overview of publications that have reported T cells specific for HLA-DQ-gluten complexes.

In vitro *protein digestibility:*

Due to the proline-rich nature, gluten proteins are highly resistant to proteolytic degradation. This results in relatively long peptides that harbour one or more T-cell stimulatory epitopes. Further details are listed in Annex A-6 and in the chapter on *in vitro* protein digestibility testing (Annex B).

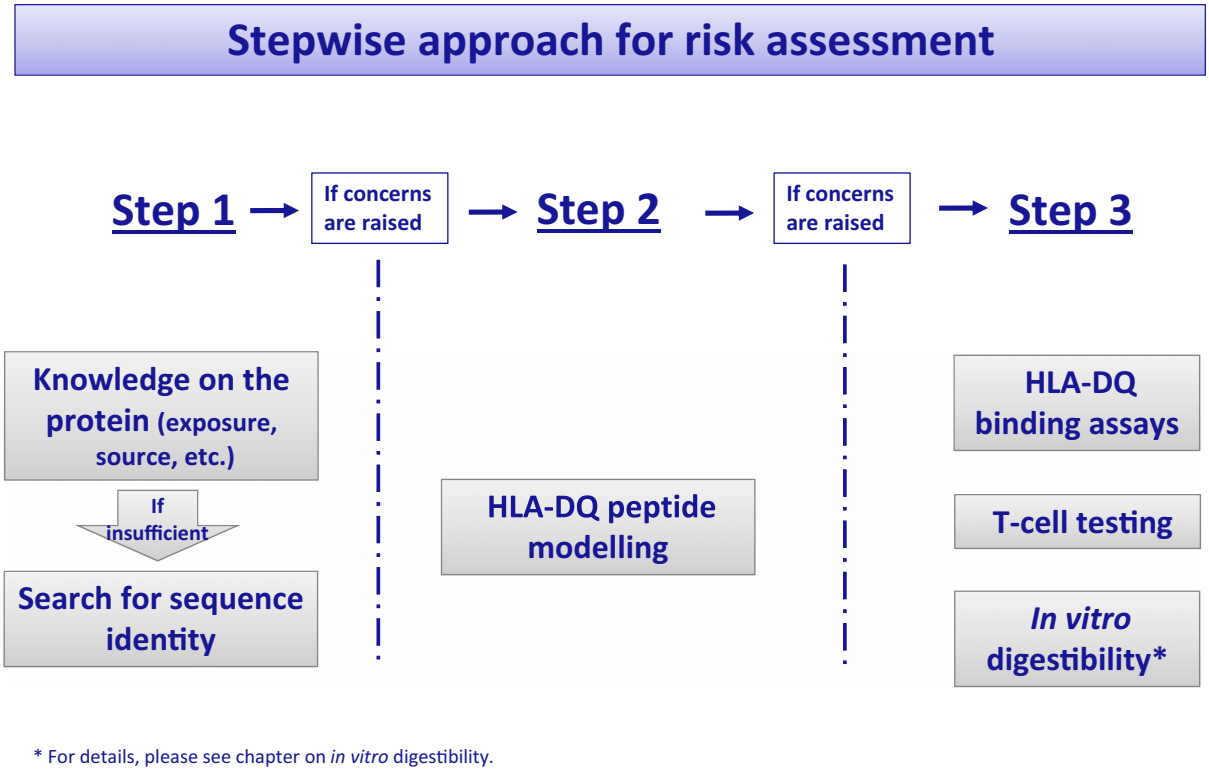



**Figure 1:** Stepwise approach for risk assessment

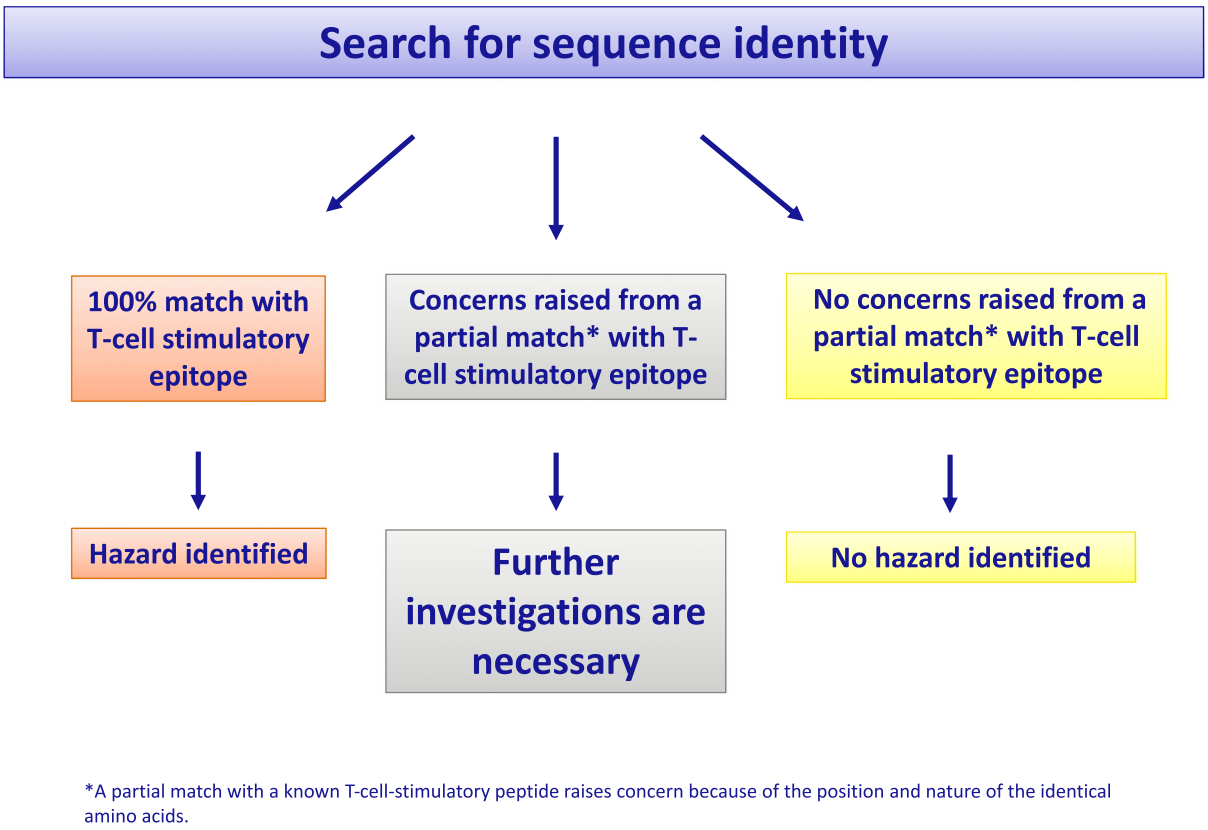



**Figure 2:** Search for sequence identity

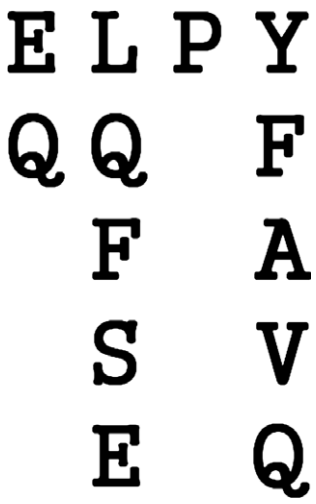



**Figure 3:** Q/E-X1-P-X2 motif: possible combinations for the Q/E-X1-P-X2 motif found in the large majority of identified immunogenic gluten-derived epitopes. It was noted that, while position 1 is always either glutamic acid (E) or glutamine (Q) and position 3 always consists of a proline (P), also positions 2 (X1) and 4 (X2) are restricted to certain amino acids

## 2.2. *In vitro* protein digestibility tests

### 2.2.1. Risk assessment considerations

*In vitro* digestibility tests can provide useful data on the susceptibility of a protein to digestion which can reflect its digestibility in the human gastrointestinal tract and subsequently provide

information on its immunogenicity. There is evidence that gastrointestinal digestion can affect the immunogenicity of dietary proteins related to both IgE and non-IgE-mediated adverse reactions to foods (see Annex B). Therefore, *in vitro* protein digestion can be used as an additional piece of information in the weight-of-evidence approach followed for the allergenicity assessment of newly expressed proteins, because no single test is fully predictive of the allergenic potential of a protein (Codex Alimentarius, 2003, 2009; EFSA GMO Panel, 2011).

The pepsin resistance test is the most commonly used digestion test for this assessment, in line with international guidelines (Codex Alimentarius 2003, 2009), the EFSA Guidance Document (2011) and Implementing Regulation (EU) No 503/2013 (IR503/2013). EFSA previously highlighted the limitations of the classical pepsin resistance test for allergenicity risk assessment and recommended that resistance to digestion of (novel) proteins should be evaluated using other *in vitro* digestibility methods designed to more closely simulate the conditions of the human digestion process (EFSA GMO Panel, 2010, 2011).

In Annex B of the present document and based on state-of-the-art in science, the EFSA GMO Panel proposes a refined *in vitro* digestion test that extends the conditions currently used in the classical pepsin resistance test in order to better reflect the range of conditions found *in vivo*. This elaborated test includes additional conditions more representative of the gastric environment with regard to pH and pepsin levels, together with an intestinal digestion phase. In addition, more informative read-outs of the test are laid out which define the extent to which either the intact protein or resistant fragments remain after *in vitro* digestion.

The EFSA GMO Panel considers that additional investigation is needed before any additional recommendation in the form of guidance for applicants can be provided on the proposed *in vitro* protein digestibility tests.

To this end, an interim phase (~ 2 years duration) is considered necessary to evaluate the proposed revisions to the *in vitro* gastrointestinal digestion test (see Annex B). During this interim phase, the laboratory(ies) involved, working with EFSA, will further detail and apply the refined digestion test methodology. After this period, EFSA will assess whether the test adds value to the allergenicity risk assessment and, if so, what further steps are needed for its final implementation in the form of guidance for applicants. An outline proposal for such an interim phase is provided in Annex B-5.

During the interim phase, and until the evaluation of the new approach is completed, the EFSA GMO Panel will continue to follow the weight-of-evidence approach for allergenicity assessment as described by EFSA (EFSA GMO Panel, 2011) and Codex Alimentarius (Codex Alimentarius, 2003, 2009).

## 2.3. Endogenous allergenicity

According to the EFSA Guidance Document (EFSA GMO Panel, 2011) and in line with principles of Codex Alimentarius (2003, 2009), endogenous allergenicity assessment of GM plants is a relevant element to be considered. The purpose of the assessment of endogenous allergenicity is to investigate that no unintended effect of the genetic modification changes the levels of endogenous allergens in a manner that would adversely impact on human and animal health (EFSA GMO Panel, 2011; König et al., 2004; Metcalfe et al., 1996; Thomas et al., 2008). This follows the same principles as those used to support the measurement of any other compound/endpoint in the compositional analysis of GM plants.

EFSA (EFSA GMO Panel, 2011) and Codex Alimentarius (2003, 2009) foresee the assessment of endogenous allergenicity only when the plant receiving the new gene(s) is recognised to be allergenic. In these cases, any potential change in the overall allergenicity of the GM plant compared with that of its non-GM comparator(s) should be analysed. Historically, this analysis was performed using sera from allergic individuals, but limitations of this assay for routine risk assessment purposes have been previously described (Fernandez et al., 2013; Selb et al., 2017).

EFSA has previously recommended the inclusion of relevant endogenous allergens in the comparative compositional analysis, implying the quantitative measurement of individual allergens (EFSA GMO Panel, 2010, 2011). This recommendation became a mandatory requirement when Implementing Regulation (EU) No 503/2013 (IR503/2013) on applications for authorisation of GM food and feed came into force. To assist applicants and risk assessors in its practical implementation, this supplementary guidance document provides further information on: (i) relevant crops for analysis; (ii) relevant allergens for quantification; (iii) methodology for quantification; and (iv) principles to be followed for data interpretation and risk assessment considerations.

The EFSA GMO Panel Guidance Documents are continuously updated to consider new scientific and regulatory developments in the field. To this end, further revisions and updates to this supplementary

EFSA Guidance Document might be needed once experience has been gained with the assessment of endogenous allergenicity and the new strategies/tools being deployed to support them.

### 2.3.1. Relevant crops for analysis

According to IR503/2013 and in line with EFSA Guidance Document (EFSA GMO Panel, 2011), an assessment of endogenous allergenicity should be performed on a case-by-case basis. When the recipient plant is recognised to be allergenic, the applicant should test any potential change in the allergenicity of the GM food or feed by comparison of the allergen repertoire with that of its appropriate comparator(s).

To date, EFSA has performed endogenous allergenicity risk assessments based on experimental data for foods recognised to be common food allergens and of public health importance as listed in Annex II of the European Regulation on food information to consumers.[7,8] In this context, soybean is recognised to be a common allergenic food and EFSA GMO Panel Scientific Opinions on GM soybean applications, which included an endogenous allergenicity assessment, were published previously (Annex C-1). To date, EFSA has not received any application involving a common allergenic food other than soybean. For crops not recognised as being commonly allergenic, specific experimental data on endogenous allergenicity are not requested by EFSA. In these cases, the assessment is carried out considering potential effects of the genetic modification on the general composition and molecular characteristics of the GM plant. However, this does not preclude EFSA requesting experimental data on endogenous allergenicity for these, if considered necessary, e.g. if the allergenic status of these foods changes. In addition, other plant-derived foods not currently listed in Annex II of such European Regulation (e.g. fruits), which might be genetically engineered in the future, should be subjected to such assessments if considered necessary. For such considerations, risk assessors, risk managers, health professionals and stakeholders can provide valuable input.

### 2.3.2. Relevant allergens for quantification

**Soybean**

Soybean is recognised as a common allergenic food by European Regulation,[7,8] and suggested as one of the eight foods accounting for approximately 90% of food allergies (FDA, 2004; OECD, 2012).

The quantitative measurement of soybean allergens as part of the compositional analysis is now a mandatory requirement in the IR503/2013 where reference to allergens in OECD consensus document is provided.

In the OECD consensus document on soybean, proteins termed as 'potential allergens' are described in Table 20, Section III-C (OECD, 2012). Nevertheless, in accordance with Article 5(2) and 5 (3) of the IR503/2013: (i) EFSA may accept derogations from specific requirements, if they are demonstrated not to be scientifically necessary for food/feed safety assessment or technically not possible to perform; and/or (ii) EFSA may request data not foreseen in OECD consensus documents anytime, if considered necessary based on new scientific findings.

In line with IR503/2013, the OECD allergen list should be taken as the starting point for the collection of 'potential allergens' (Figure 4). In addition, this list should be complemented with searches in the scientific and medical literature, and in various updated databases (EFSA GMO Panel, 2010; Appendix 3.13, Table I for a list of relevant databases). It is noted that on the one hand the OECD list of allergens might not be complete at a given point in time, e.g. it might be out dated and/or miss relevant (more recent) entries. On the other hand, not all 'potential allergens' listed in this OECD consensus document can currently be measured due to technical reasons (e.g. amino acid sequence not available) and/or their clinical relevance might not have been demonstrated. Once a comprehensive search of 'potential allergens' in the literature and databases is conducted, the relevant allergens selected for quantification and subsequent comparative analysis should be justified. As a complementary and/or alternative approach, a systematic review could be performed, aiming to

[7] EC, 2003. Directive 2003/89/EC of the European Parliament and of the Council of 10 November 2003 amending Directive 2000/13/EC as regards indication of the ingredients present in foodstuffs.

[8] EC, 2011. Regulation (EU) No 1169/2011 of the European Parliament and of the Council of 25 October 2011 on the provision of food information to consumers, amending Regulations (EC) No 1924/2006 and (EC) No 1925/2006 of the European Parliament and of the Council, and repealing Commission Directive 87/250/EEC, Council Directive 90/496/EEC, Commission Directive 1999/10/EC, Directive 2000/13/EC of the European Parliament and of the Council, Commission Directives 2002/67/EC and 2008/5/EC and Commission Regulation (EC) No 608/2004.

identify clinically relevant allergens. A scientific rational, explaining why an allergen is not considered relevant should be provided. A representation of these considerations can be found in Figure 4.

A possible approach how to identify proteins relevant for the endogenous allergenicity assessment of soybean is described in Annex C-2.

**Other GM plants**

For foods other than soybean which are recognised allergenic (risk assessors, risk managers, health professionals and stakeholders can provide valuable input), a similar approach/strategy for the identification of relevant allergens as the one followed for soybean (see Annex C-2) should be applied, whenever considered necessary. To date, EFSA has not received any application involving a common allergenic food other than soybean.

### 2.3.3. Methodology for quantification

Either enzyme-linked immunosorbent assay (ELISA) or mass spectrometry (MS) approaches are appropriate methods for the quantification of endogenous allergens, both allowing the specific detection and quantification of single known allergens. Further considerations on these methodologies are described in Annex C-3.

### 2.3.4. Data interpretation and risk assessment

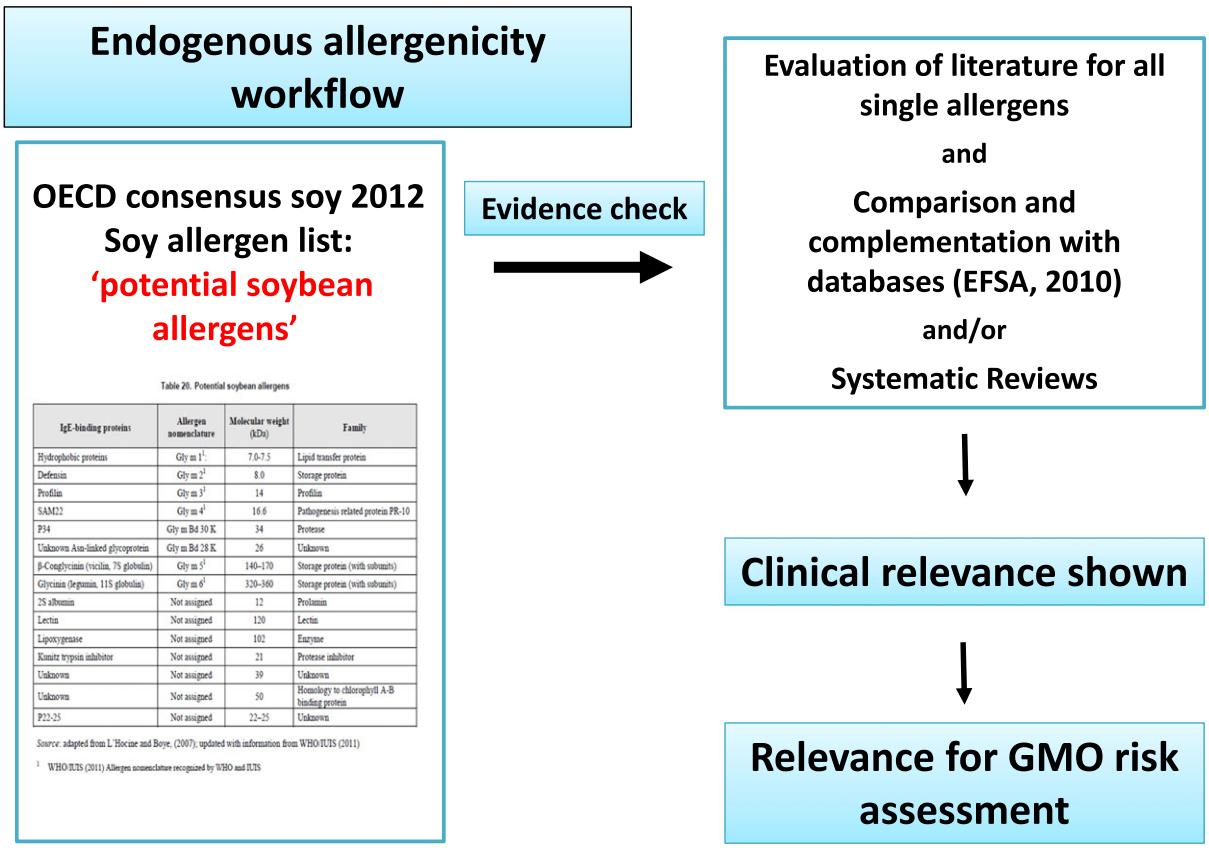
According to IR503/2013, conclusions of the allergenicity assessment should indicate 'whether the genetically modified food or feed is likely to be more allergenic than its conventional counterpart'. The starting point in the assessment is the identification of any potential change in the allergenicity of the GM food or feed by comparison of the allergen repertoire of the GM plant and its conventional counterpart, taking into account natural variability.

Allergens included in the compositional analysis should be measured and analysed according to the principles of the comparative assessment performed for any other compositional compounds (see section 1.3.2 of IR503/2013). To this end, the starting point of the assessment should be the identification of statistically significant differences between the GM plant and its conventional counterpart. A further evaluation should investigate whether or not the differences observed fall within or outside the range of natural variation estimated from the reference varieties included in the field trial, i.e. the equivalence test (IR503/2013). In case the levels of a specific allergen in a GM plant increases significantly from the levels observed in the appropriate comparator(s) and falls outside the estimated range of natural variation, the biological relevance in relation to human and animal health should be assessed.

Additional considerations and/or experimental data might be needed on a case-by-case basis. As for other compounds included in the compositional analysis, the nature of these additional considerations and/or experimental data depend on the number and magnitude of the changes identified, as well as on the clinical/safety relevance of the specific allergen(s)/compound(s) involved.

Ultimately, when a potential increase in allergenicity due to the genetic modification cannot be excluded, the GM food or feed should be further characterised in the light of its anticipated intake, as requested by IR503/2013. Occupational allergy should also be considered with respect to inhalation or contact with potential allergens. In all cases, an exposure assessment should focus on the European population aiming at identifying particular groups at high risk, which might be affected by a specific change of the allergen content.

Possible approaches for data interpretation and risk assessment of soybean endogenous allergenicity are summarised in Annex C-4.




| IgE-binding proteins | Allergen nomenclature | Molecular weight (kDa) | Family |
|---|---|---|---|
| Hydrophobic proteins | Gly m 1[1] | 7.0-7.5 | Lipid transfer protein |
| Defensin | Gly m 2[1] | 8.0 | Storage protein |
| Profilin | Gly m 3[1] | 14 | Profilin |
| SAM22 | Gly m 4[1] | 16.6 | Pathogenesis related protein PR-10 |
| P34 | Gly m Bd 30 K | 34 | Protease |
| Unknown Asn-linked glycoprotein | Gly m Bd 28 K | 26 | Unknown |
| β-Conglycinin (vicilin, 7S globulin) | Gly m 5[1] | 140–170 | Storage protein (with subunits) |
| Glycinin (legumin, 11S globulin) | Gly m 6[1] | 320–360 | Storage protein (with subunits) |
| 2S albumin | Not assigned | 12 | Prolamin |
| Lectin | Not assigned | 120 | Lectin |
| Lipoxygenase | Not assigned | 102 | Enzyme |
| Kunitz trypsin inhibitor | Not assigned | 21 | Protease inhibitor |
| Unknown | Not assigned | 39 | Unknown |
| Unknown | Not assigned | 50 | Homology to chlorophyll A-B binding protein |
| P22-25 | Not assigned | 22–25 | Unknown |

**Figure 4:** Endogenous allergenicity workflow for the selection of relevant allergens in soybean

## Documentation provided to EFSA

1) Proposal for a self-task mandate of the EFSA GMO Panel to establish a Working Group to develop supplementary guidelines for the allergenicity assessment of GM plants to incorporate new developments. May 2014. Submitted by the Chair of the ESA GMO Panel.
2) Acceptance of the self-task mandate of the EFSA GMO Panel to establish a Working Group to develop supplementary guidelines for the allergenicity assessment of GM plants to incorporate new developments. July 2014. Submitted by EFSA Executive Director.

## References

Abadie V, Sollid LM, Barreiro LB and Jabri B, 2011. Integration of genetic and immunological insights into a model of celiac disease pathogenesis. Annual Review of Immunology, 29, 493–525.

van Bergen J, Mulder CJ, Mearin ML and Koning F, 2015. Local communication among mucosal immune cells in patients with celiac disease. Gastroenterol, 148, 1187–1194.

Codex Alimentarius, 2003. Foods derived from modern biotechnology. Codex Alimentarius Commission, Joint FAO/WHO Food Standards Programme, Rome.

Codex Alimentarius, 2009. Foods derived from modern biotechnology. Codex Alimentarius Commission, Joint FAO/WHO Food Standards Programme, Rome.

Comino I, de Lourdes Moreno M and Sousa C, 2015. Role of oats in celiac disease. World Journal of Gastroenterology, 21, 11825–11831.

Crevel RWR, Ballmer-Weber BK, Holzhauser T, Hourihane JO'B, Knulst AC, Mackie AR, Timmermans F and Taylor SL, 2008. Thresholds for food allergens and their value to different stakeholders. Allergy 63:597–609.

EFSA (European Food Safety Authority), 2006. Guidance document for the risk assessment of geneticallymodified plants and derived food and feed by the Scientific Panel on Genetically Modified Organisms (GMO) - including draft document updated in 2008 (reference EFSA-Q-2003-005A). EFSA Journal 2006;4(4):99, 105 pp. https://doi.org/10.2903/j.efsa.2006.99

EFSA (European Food Safety Authority), 2014. Indicative timelines for submitting additional or supplementary information to EFSA during the risk assessment process of regulated products. EFSA Journal 2014;12(1):3553, 37 pp. https://doi.org/10.2903/j.efsa.2014.3553

EFSA GMO Panel (EFSA Panel on Genetically Modified Organisms), 2010. Scientific Opinion on the assessment of allergenicity of GM plants and microorganisms and derived food and feed. EFSA Journal 2010;8(7):1700, 168 pp. https://doi.org/10.2903/j.efsa.2010.1700

EFSA GMO Panel (EFSA Panel on Genetically Modified Organisms), 2011. Scientific Opinion on Guidance for risk assessment of food and feed from genetically modified plants. EFSA Journal 2011;9(5): 2150, 37 pp. https://doi.org/10.2903/j.efsa.2011.2150

EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2014. Scientific Opinion on the evaluation of allergenic foods and food ingredients for labelling purposes. EFSA Journal 2014;12(11):3894, 286 pp. https://doi.org/10.2903/j.efsa.2014.3894

FDA, 2004. Food allergen labeling and consumer protection act (FALCPA). Congressional record v. 150. Available online: http://www.fda.gov/downloads/Food/LabelingNutrition/FoodAllergensLabeling/GuidanceComplianceREgulatoryInformation/UCM179394.pdf

Fernandez A, Mills EN, Lovik M, Spoek A, Germini A, Mikalsen A and Wal JM, 2013. Endogenous allergens and compositional analysis in the allergenicity assessment of genetically modified plants. Food and Chemical Toxicology, 62, 1–6.

Garsed K and Scott BB, 2007. Can oats be taken in a gluten-free diet? A systematic review. Scandinavian Journal of Gastroenterology, 42, 171–178.

Green PH and Cellier C, 2007. Celiac disease. New England Journal of Medicine, 357, 1731–1743.

König A, Cockburn A, Crevel RWR, Debruyne E, Grafstroem R, Hammerling U, Kimber I, Knudsen I, Kuiper HA, Peijnenburg AACM, Penninks AH, Poulsen M, Schauzu M and Wal JM, 2004. Assessment of the safety of foods derived from genetically modified (GM) crops. Food and Chemical Toxicology, 42, 1047–1088.

Koning F, Thomas R, Rossjohn J and Toes RE, 2015. Coeliac disease and rheumatoid arthritis: similar mechanisms, different antigens. Nature Reviews Rheumatology, 11, 450–461.

Lundin KEA, Nilsen EM, Scott HG, Løberg EM, Gjøen A, Bratlie J, Skar V, Mendez E, Løvik A and Kett K, 2003. Oats induced villous atrophy in coeliac disease. Gut, 52, 1649–1652.

Madsen CB, Crevel RWR, Chan C-H, Dubois AEJ, DunnGalvin A, Flokstra-de Blok BMJ, Gowland MH, Hattersley S, Hourihane JO'B, Nørhede P, Pfaff S, Rowe G, Schnadt S and Vlieg-Boerstra BJ, 2010. Food allergy: stakeholder perspectives on acceptable risk. Regulatory Toxicology and Pharmacology 57, 256–265.

Metcalfe DD, Astwood JD, Townsend R, Sampson HS, Taylor SL and Fuchs RL, 1996. Assessment of the allergenic potential of foods derived from genetically engineered crop plants. Critical Reviews in Food Science and Nutrition, 36, S165–S186.

Mills ENC, Marsh JT, Boyle R, Hoffmann-Sommergruber K, DuPont D, Bartara J, Bakalis S, McLaughlin J and Shewry PR; The University of Manchester, 2013a. Literature review: 'non-IgE-mediated immune adverse reactions to foods'. EFSA supporting publication 2013:EN-527, 40 pp.

Mills ENC, Marsh JT, Johnson PE, Boyle R, Hoffmann-Sommergruber K, DuPont D, Bartras J, Bakalis S, McLaughlin J and Shewry PR; The University of Manchester, 2013b. Literature review: 'in vitro digestibility tests for allergenicity assessment'. EFSA supporting publication 2013:EN-529, 52 pp.

OECD, 2012. Revised consensus document on compositional considerations for new 452 varieties of soybean [*Glycine max* (L.) Merr.]: key food and feed nutrients, 453 antinutrients, toxicants and allergens. Series on the Safety of Novel Foods and 454 Feeds No. 25.

Selb R, Wal JM, Moreno FJ, Lovik M, Mills C, Hoffmann-Sommergruber K and Fernandez A, 2017. Assessment of endogenous allergenicity of genetically modified plants exemplified by soybean - where do we stand? Food and Chemical Toxicology, 101, 139–148.

Shewry PR, Halford NG and Lafiandra D, 2003. Genetics of wheat gluten proteins. Advances in Genetics, 49, 111–184.

Sollid LM, Qiao S, Anderson RP, Gianfrani C and Koning F, 2012. Nomenclature and listing of celiac disease relevant gluten T-cell-epitopes restricted by HLA-DQ molecules. Immunogenetics, 64, 455–460.

Thomas K, Herouet-Guicheney C, Ladics G, McClain S, MacIntosh S, Privalle L and Woolhiser M, 2008. Current and future methods for evaluating the allergenic potential of proteins: international workshop report 23–25 October 2007. Food and Chemical Toxicology, 46, 3219–3225.

Tye-Din JA, Stewart JA, Dromey JA, Beissbarth T, van Heel DA, Tatham A, Henderson K, Mannering SI, Gianfrani C, Jewell DP, Hill AV, McCluskey J, Rossjohn J and Anderson RP, 2010. Comprehensive, quantitative mapping of T cell epitopes in gluten in celiac disease. Science Translational Medicine, 2, 41–51.

## Abbreviations

| | |
|---|---|
| CD | celiac disease |
| DBPCFC | double-blind placebo-controlled food challenge |
| ELISA | enzyme-linked immunosorbent assay |
| FPIES | food-protein induced enterocolitis |
| GANA | *N*-carbobenzoxy-diglycyl-L-arginyl-2-naphthylamide hydrochloride |
| GGPNA | γ-glutamyl-*p*-nitroanilide |
| GM | genetically modified |

| | |
|---|---|
| GMO | genetically modified organism |
| HMW | high molecular weight |
| IgA | immunoglobulin A |
| IgE | immunoglobulin E |
| LMW | low molecular weight |
| MHC | major histocompatibility complex |
| MS | mass spectrometry |
| MS-MS | tandem mass spectrometry |
| OECD | Organisation for Economic Co-operation and Development |
| PPI | proton pump inhibitor |
| SDS-PAGE | sodium dodecyl sulfate-polyacrylamide gel electrophoresis |
| TG2 | transglutaminase 2 |

# Annex A – Non-IgE-mediated adverse immune reactions to foods

*Background*

Celiac disease (CD) has a strong genetic component. It is associated with particular immune response genes, i.e. those responsible for the class II major histocompatibility complex (MHC) molecules, called HLA in humans. Most CD patients express particular HLA-DQ-molecules. HLA-DQ molecules are dimers of an alpha-(DQA1) and a beta-(DQB1) chain. Like all HLA-molecules, HLA-DQ molecules bind short peptides and present these to T cells of the immune system. While T cells ignore HLA-bound peptides derived from harmless ('self') proteins, HLA-bound peptides derived from pathogens are specifically detected and this recognition leads to the generation of a protective T-cell response and eradication of the pathogen. The large majority of CD patients (approximately 95%) express HLA-DQ2.5 (DQA1*05:01, DQB1*02:01) (Sollid et al., 2012) while the remainder are usually HLA-DQ8 positive (DQA1*03, DQB1*03:02). The few patients that express neither DQ2.5 nor DQ8, often express HLA-DQ molecules that contain only one of the DQ2.5-chains, e.g. DQ2.2 (DQA1*02:01, DQB1*02:01) or DQ7.5 (DQA1*05, DQB1*03:01) (Karell et al., 2003). In affected people, but not in healthy individuals, pro-inflammatory gluten-specific CD4+ T cells are present in the lamina propria of the affected duodenum. Importantly, these CD4+ T cells recognise gluten peptides only when presented by the disease associated HLA-DQ molecules (Lundin et al., 1993, 1994; Tye-Din et al., 2010; Vader et al., 2002b; van de Wal et al., 1998b). In essence, in patients with CD the immune system displays an aberrant response: the harmless gluten proteins in the food are recognised as hazardous, leading to a pro-inflammatory response as long as gluten is consumed. Elimination of gluten from the diet constitutes an effective treatment because the T-cell stimulatory gluten peptides are no longer present. Unfortunately, once a gluten-specific T-cell response has developed, this leads to immunological memory and every subsequent exposure to gluten will reactivate the gluten-reactive T cells and consequently lead to inflammation. A lifelong gluten-free diet is thus required.

Gluten is the cohesive mass that remains when starch has been removed from wheat dough (Shewry et al., 1992). Gluten consists of gliadin and glutenin subcomponents. The gliadins are subdivided into α-, γ- and ω-gliadins, and dozens of variants of each type are typically present in a single wheat variety. The glutenins are subdivided into high molecular weight (HMW) and low molecular weight (LMW) subunits. The most commonly used wheat varieties are bread wheats (*Triticum aestivum*), which are hexaploid species, and pasta wheats (*Triticum durum*), which are tetraploid species. Thus, in any single wheat variety up to a hundred different gluten proteins are listed, many of which are highly similar and only differ by a few amino acids from each other.

T-cell epitopes derived from the α-, γ- and ω-gliadins as well as from the HMW- and LMW-glutenins were reported (Arentz-Hansen et al., 2000; Shan et al., 2002; Sjöström et al., 1998; Vader et al., 2002b; van de Wal et al., 1998b, 1999). In addition, T-cell epitopes in both hordeins and secalins were identified that are highly homologous or identical to those found in wheat (Tye-Din et al., 2010; Vader et al., 2003). The gluten-like avenins of oat are more distinct; however, avenin-specific as well as cross-reactive T-cell responses were described (Arentz-Hansen et al., 2004; Vader et al., 2003).

High affinity binding of peptides to either HLA-DQ2.5 or -DQ8 depends on the presence of one or more negatively charged amino acids. As gluten proteins are virtually devoid of negatively charged amino acids, native gluten-derived peptides bind poorly to these HLA-DQ molecules. Due to the activity of the enzyme tissue transglutaminase 2 (TG2) in the gastrointestinal tract, the required negative charge(s) are introduced when this enzyme converts glutamine residues within gluten peptides into negatively charged glutamic acid (Molberg et al., 1998; Vader et al., 2002a; van de Wal et al., 1998a). These deamidated gluten peptides then bind with increased affinity to HLA-DQ2.5 or -DQ8, and this stronger binding enhances or causes immunogenicity (van de Wal et al., 1998; Arentz-Hansen et al., 2000; Henderson et al., 2007; Kim et al., 2004; Moustakas et al., 2000; Quarsten et al., 1999).

The specificity of TG2 for particular target sequences in gluten proteins plays a crucial role in the generation of a relatively large number of gluten-derived peptides that bind to HLA-DQ2.5. Glutamine and proline are abundantly present in gluten proteins, together they comprise over 50% of the amino acids in gluten. Therefore, Q-X-P and Q-P sequences (where Q is glutamine; P is proline; X is any amino acid except P) are often found in gluten proteins and TG2 typically deamidates glutamine residues in Q-X-P sequences, but not in QP sequences (Vader et al., 2002a). Therefore, immunogenic gluten-derived peptides are typically found in the proline rich-regions of gluten proteins and usually contain a Q-X-P motif. 'Classic' examples of such peptides are the immunodominant T-cell epitopes present in the N-terminal part of the α-gliadins: PFPQPQLPY and PQPQLPYPQ. In fact, these sequences

have a 7-amino acid overlap and in both sequences only one Q-residue is a target for TG2, the Q in the QLPY sequence which allows the introduction of a negative charge at either position 6 or position 4, respectively. In both instances, this generates a gluten peptide that binds with high affinity to HLA-DQ2.5. The available crystal structures of HLA-DQ2.5-gliadin and the bound cognate T-cell receptor demonstrate that the negatively charged glutamic acid serves as an anchor residue for peptide binding to HLA-DQ2.5 and does not contact the T-cell receptor (Petersen et al., 2014). Additionally, the proline-rich nature of gluten renders these proteins resistant to degradation by enzymes in the gastrointestinal tract. Relatively long gluten fragments are, therefore, present in the small intestine. This likely contributes to the immunogenic nature of these peptides (Shan et al., 2002). Thus, at least three factors contribute to the immunogenicity of gluten: (a) resistance to proteolytic degradation, (b) specific recognition by TG2, and (c) peptide binding properties of HLA-DQ2.5 and HLA-DQ8.

The glutamic acid introduced by TG2 is usually in position 4 (p4) or p6 in HLA-DQ2.5 restricted epitopes and at position p1 and/or p9 in HLA-DQ8 restricted epitopes (Sollid et al., 2012). As a consequence, HLA-DQ2.5-restricted gluten epitopes carry a proline at either p6 or p8. This positioning of proline residues is less strict in the case of the DQ8 epitopes. In all cases, the glutamic acid residues serve as anchors important for binding of the peptides to either HLA-DQ2.5 or -DQ8.

It is important to note that, while polyclonal T-cell responses to multiple T-cell epitopes are usually detected in CD patients, responses to the DQ2.5-glia-$\alpha$1, DQ2.5-glia-$\alpha$2 epitopes and homologues thereof in the $\omega$-gliadins, hordeins and secalins are dominant in DQ2.5-positive patients (Arentz-Hansen et al., 2000; Tollefsen et al., 2006; Tye-Din et al., 2010). In DQ8-positive patients, responses to the DQ8-glia-$\alpha$1 epitope are most frequently found (Tollefsen et al., 2006; van de Wal et al., 1998b).

The following criteria were used to define CD reactive epitopes in Sollid et al. (2012):

- reactivity against the epitope must have been defined by at least one specific T-cell clone;
- the HLA-restriction element involved must have been unequivocally defined;
- the 9-amino acid core of the epitope must have been defined either by an analysis with truncated peptides and/or HLA-binding with lysine scan of the epitope, or a comparable approach. In a lysine scan, all amino acids in the sequence of interest are individually replaced by a lysine and the impact of these single amino acids substitutions on HLA-binding is determined, information which usually reveals which amino acids in the sequence are required for binding to HLA.

Because CD is caused by an immune response to a foreign protein and all symptoms disappear upon withdrawal of gluten from the diet, the condition should not be regarded as a true autoimmune disease. Autoantibodies specific for tissue transglutaminase appear to be secondary to the T-cell driven immune response to gluten and disappear if gluten is eliminated from the diet (Rossjohn and Koning, 2016).

Other conditions linked to wheat or gluten, summarised as 'non-celiac gluten sensitivity', are not part of this document because there are no known definite underlying pathomechanisms (Aziz et al., 2015). In the last years, the diagnosis of non-celiac gluten sensitivity has emerged and refers to clinically diagnosed gluten-related symptoms without criteria for gluten allergy or celiac disease (Czaja-Bulsa, 2015). The unambiguous diagnosis of this condition is hampered by several facts. Symptoms are pleomorphic and often vague and limited in time. They mostly concern the gastrointestinal tract, but can also involve among many others: mood disorders, chronic recurrent headache, chronic fatigue or muscle disorders. Nevertheless, avoidance of gluten-containing foods leading to symptom relief, and re-exposure provoking relapse are suggestive of a gluten-related origin of the disease. It can, however, not be excluded that the symptoms are due to other components present in wheat and related cereals. The cause of this condition remains undefined. Immune mechanisms have not been identified and the character of the symptoms mostly suggests a gluten-intolerance (difficult digestion). Non-celiac gluten sensitivity seems more prominent in individuals with irritable bowel disease, and in those suffering of FODMAP intolerance (Biesiekierski et al., 2013).

The current chapter of the guidance document dealing with non-IgE-mediated food sensitivity is at the current stage of knowledge not addressing non-celiac gluten sensitivity. This disease is currently only marginally understood, in particular with regard to its mechanism, and is increasingly (mostly self) diagnosed. Future studies might unravel the exact mechanism of the disease and thus indicate whether additional risk assessment considerations will be needed – aspects of which we currently have no evidence/sufficient knowledge.

To sum up, CD, a condition in which immunological mechanisms were extensively studied, can be considered as a pathognomonic, gluten-related non-IgE mediated food sensitivity. CD is within the field of this guidance document and has been extensively addressed accordingly.

*Annex A-1. T-cell epitopes known in celiac disease*

Several listings/databases of the known T-cell stimulatory sequences identified in gluten, hordeins, secalins and avenins are available including the https://propepper.net database and the http://www.allergenonline.org/ database. Furthermore, an overview of the best characterised epitopes along with a unified nomenclature is presented in Sollid et al. (2012). Table A.1 lists the epitopes described in the latter manuscript. This compilation lists both the HLA-DQ2- and HLA-DQ8-restricted epitopes and includes the known immunodominant epitopes present in the α- and ω-gliadins as well as the less commonly recognised epitopes in the γ-gliadins and the LMW- and HMW-glutenins. The list contains a convenient overview of the most important and well defined epitopes. It is, nevertheless, important to note that due to the extreme variability of the gluten and gluten-like proteins in barley and rye, (single) amino acid variants of these epitopes do exist, some of which may also exhibit T-cell stimulatory activity. Also, it cannot be fully excluded that additional gluten epitopes may be identified in the future. However, this is unlikely because large numbers of patients have already been extensively tested for gluten reactivity, on the basis of gluten peptide libraries (Tye-Din et al., 2010). Thus, any protein containing one or more sequences that display high sequence identity to the epitope sequences present in this list will likely have the capacity to trigger gluten-specific T cells.

The development of a comprehensive database that is publicly available, curated regularly and appropriately built and designed for risk assessment purposes is an important aspect to be further investigated.

References provided in background section of Annex A and Annex A-1:

Arentz-Hansen H, Körner R, Molberg, Quarsten H, Vader W, Kooy YM, Lundin KEA, Koning F, Roepstorff P, Sollid LM and McAdam SN, 2000. The intestinal T cell response to α-gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase. The Journal of Experimental Medicine 191, 603–612.

Arentz-Hansen H, Fleckenstein B, Molberg, Scott H, Koning F, Jung G, Roepstorff P, Lundin KE and Sollid LM, 2004. The molecular basis for oat intolerance in patients with celiac disease. Plos Med, 1, e1.

Aziz I, Hadjivassilou M and Sanders DS, 2015. The Spectrum of Noncoeliac Gluten Sensitivity. Nature Reviews. Gastroenterology and Hepatology, 12, 516–526.

Biesiekierski JR, Peters S, Newnham ED, Rosella O, Muir JG and Gibson PR, 2013. No effects of gluten in patients with self-reported non-celiac gluten sensitivity after dietary reduction of fermentable, poorly absorbed, short-chain carbohydrates. Gastroenterology, 145, 320–328.

Czaja-Bulsa G, 2015. Non coeliac gluten sensitivity. A new disease with gluten intolerance. Clinical Nutrition, 34, 189–194.

Ellis HJ, Pollock EL, Engel W, Fraser JS, Rosen-Bonson S, Wieser H and Ciclitira PJ, 2003. Investigation of the putative immunodominant T cell epitopes in coeliac disease. Gut, 52, 212–217.

Henderson KN, Tye-Din JA, Reid HH, Chen Z, Borg NA, Beissbarth T, Tatham A, Mannering SI, Purcell AW, Dudek NL, van Heel DA, McCluskey J, Rossjohn J and Anderson RP, 2007. A structural and immunological basis for the role of human leukocyte antigen DQ8 in celiac disease. Immunity, 27, 23–34.

Karell K, Louka AS, Moodie SJ, Ascher H, Clot F, Greco L, Ciclitira PJ, Sollid LM and Partanen J, 2003. HLA types in celiac disease patients not carrying the DQA1*05-DQB1*02 (DQ2) heterodimer: results from the European Genetics Cluster on Celiac Disease. Human Immunology, 64, 469–477.

Kim CY, Quarsten H, Bergseng E, Khosla C and Sollid LM, 2004. Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. Proceedings of the National Academy of Sciences of the United States of America, 101, 4175–4179.

Lundin KEA, Scott H, Fausa O, Thorsby E and Sollid LM, 1994. T cells from the small intestinal mucosa of a DR4, DQ7/DR4, DQ8 celiac disease patient preferentially recognize gliadin when presented by DQ8. Human Immunology, 41, 285–291

Lundin KEA, Scott H, Hansen T, Paulsen G, Halstensen TS, Fausa O, Thorsby E and Sollid LM, 1993. Gliadin-specific, HLA-DQ(α*0501,β1*0201) restricted T cells isolated from the small intestinal mucosa of celiac disease patients. Journal of Experimental Medicine, 178, 187–196.

Mitea C, Salentijn EM, van Veelen P, Goryunova SV, van der Meer IM, van den Broeck HC, Mujico JR, Monserrat V, Gilissen LJ, Drijfhout JW, Dekking L, Koning F and Smulders MJ, 2010. A universal approach to eliminate antigenic properties of alpha-gliadin peptides in celiac disease. PLoS One, 5, e15637.

Molberg, McAdam SN, Körner R, Quarsten H, Kristiansen C, Madsen L, Fugger L, Scott H, Roepstorff P, Lundin KEA, Sjöström H and Sollid LM, 1998. Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells. Nature Medicine, 4, 713–717.

Moustakas AK, van de WY, Routsias J, Kooy YM, Van VP, Drijfhout JW, Koning F and Papadopoulos GK, 2000. Structure of celiac disease-associated HLA-DQ8 and non-associated HLA-DQ9 alleles in complex with two disease-specific epitopes. International of Immunology, 12, 1157–1166.

Petersen J, Montserrat V, Mujico JR, Loh KL, Beringer DX, van Lummel M, Thompson A, Mearin ML, Schweizer J, Kooy-Winkelaar Y, van Bergen J, Drijfhout JW, Kan WT, La Gruta NL, Anderson RP, Reid HH, Koning F and Rossjohn J, 2014. T-cell receptor recognition of HLA-DQ2-gliadin complexes associated with celiac disease. Nature Structural and Molecular Biology, 21, 480–488.

Quarsten H, Molberg, Fugger L, McAdam SN and Sollid LM, 1999. HLA binding and T cell recognition of a tissue transglutaminase-modified gliadin epitope. European Journal of Immunology, 29, 2506–2514.

Rossjohn J and Koning F, 2016. A biased view toward celiac disease. Mucosal Immunology, 9, 583–586.

Shan L, Molberg, Parrot I, Hausch F, Filiz F, Gray GM, Sollid LM and Khosla C, 2002. Structural basis for gluten intolerance in celiac sprue. Science, 297, 2275–2279.

Shewry PR, Tatham AS and Kasarda DD, 1992. Cereal proteins and coeliac disease. In Marsh MN (ed.) Coeliac disease. Blackwell Scientific Publications, Oxford.

Sjöström H, Lundin KEA, Molberg, Körner R, McAdam SN, Anthonsen D, Quarsten H, Noren O, Roepstorff P, Thorsby E and Sollid LM, 1998. Identification of a gliadin T-cell epitope in coeliac disease: general importance of gliadin deamidation for intestinal T-cell recognition. Scandinavian Journal of Immunology, 48, 111–115.

Sollid LM, Qiao S, Anderson RP, Gianfrani C and Koning F, 2012. Nomenclature and listing of celiac disease relevant gluten T-cell-epitopes restricted by HLA-DQ molecules. Immunogenetics, 64, 455–460.

Tollefsen S, Arentz-Hansen H, Fleckenstein B, Molberg, Raki M, Kwok WW, Jung G, Lundin KE and Sollid LM, 2006. HLA-DQ2 and -DQ8 signatures of gluten T cell epitopes in celiac disease. Journal of Clinical Investigation, 116, 2226–2236.

Tye-Din JA, Stewart JA, Dromey JA, Beissbarth T, van Heel DA, Tatham A, Henderson K, Mannering SI, Gianfrani C, Jewell DP, Hill AV, McCluskey J, Rossjohn J and Anderson RP, 2010. Comprehensive, quantitative mapping of T cell epitopes in gluten in celiac disease. Science Translational Medicine, 21, 41–51.

Vader LW, de Ru A, van Der WY, Kooy YM, Benckhuijsen W, Mearin ML, Drijfhout JW, van Veelen P and Koning F, 2002a. Specificity of tissue transglutaminase explains cereal toxicity in celiac disease. Journal of Experimental Medicine, 195, 643–649.

Vader W, Kooy Y, van Veelen P, de Ru A, Harris D, Benckhuijsen W, Pena S, Mearin L, Drijfhout JW and Koning F, 2002b. The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides. Gastroenterology, 122, 1729–1737.

Vader LW, Stepniak DT, Bunnik EM, Kooy YM, de Haan W, Drijfhout JW, van Veelen PA and Koning F, 2003. Characterization of cereal toxicity for celiac disease patients based on protein homology in grains. Gastroenterology, 125, 1105–1113.

van de Wal Y, Kooy Y, van Veelen P, Pena S, Mearin L, Papadopoulos G and Koning F, 1998a. Selective deamidation by tissue transglutaminase strongly enhances gliadin-specific T cell reactivity. Journal of Immunology 161, 1585–1588.

van de Wal Y, Kooy YM, van Veelen PA, Pena SA, Mearin LM, Molberg, Lundin KEA, Sollid LM, Mutis T, Benckhuijsen WE, Drijfhout JW and Koning F, 1998b. Small intestinal T cells of celiac disease patients recognize a natural pepsin fragment of gliadin. Proceedings of the National Academy of Sciences USA, 95, 10050–10054.

van de Wal Y, Kooy YM, van Veelen P, Vader W, August SA, Drijfhout JW, Pena SA and Koning F, 1999. Glutenin is involved in the gluten-driven mucosal T cell response. European Journal of Immunology, 29, 3133–3139.

**Table A.1:** List of celiac disease relevant DQ2 and DQ8 restricted T-cell epitopes recognised by CD4+ T cells (taken from Sollid et al., 2012)

| **DQ2 restricted epitopes** | | |
|---|---|---|
| **Epitope** | **Motif** | **Reference** |
| DQ2.5-glia-α1a | P F P Q P **Q L P Y** | Arentz-Hansen et al. (2000) |
| DQ2.5-glia-α1b | P Y P Q P **Q L P Y** | Arentz-Hansen et al. (2002) |
| DQ2.5-glia-α2 | P Q P **Q L P Y** P Q | Arentz-Hansen et al. (2000) |
| DQ2.5-glia-α3 | F R P **Q Q P Y** P Q | Vader et al. (2002b) |
| DQ2.5-glia-γ1 | P Q Q S F P Q Q Q | Sjöström et al. (1998) |
| DQ2.5-glia-γ2 | I Q P **Q Q P A** Q L | Qiao et al. (2005), Vader et al. (2002b) |
| DQ2.5-glia-γ3 | Q Q P **Q Q P Y** P Q | Arentz-Hansen et al. (2002) |
| DQ2.5-glia-γ4a | S Q P Q Q **Q F P Q** | Arentz-Hansen et al. (2002) |
| DQ2.5-glia-γ4b | P Q P Q Q **Q F P Q** | Qiao et al. (2005) |
| DQ2.5-glia-γ4c | Q Q P **Q Q P F** P Q | Arentz-Hansen et al. (2002) |
| DQ2.5-glia-γ4d | P Q P **Q Q P F** C Q | Qiao (unpublished) |

| DQ2 restricted epitopes | | |
|---|---|---|
| **Epitope** | **Motif** | **Reference** |
| DQ2.5-glia-γ5 | Q Q P F P **Q Q P Q** | Arentz-Hansen et al. (2002) |
| DQ2.5-glia-ω1 | P F P Q P **Q Q P F** | Tye-Din et al. (2010) |
| DQ2.5-glia-ω2 | P Q P **Q Q P F** P W | Tye-Din et al. (2010) |
| DQ2.2-glut-L1 | P F S Q Q **Q Q P V** | Vader et al. (2002b) |
| DQ2.5-glut-L2 | F S Q Q Q **Q S P F** | Stepniak et al. (2005), Vader et al. (2002b) |
| DQ2.5-hor-1 | P F P Q P **Q Q P F** | Tye-Din et al. (2010), Vader et al. (2003) |
| DQ2.5-hor-2 | P Q P **Q Q P F** P Q | Vader et al. (2003) |
| DQ2.5-sec-1 | P F P Q P **Q Q P F** | Tye-Din et al. (2010), Vader et al. (2003) |
| DQ2.5-sec-2 | P Q P **Q Q P F** P Q | Vader et al. (2003) |
| DQ2.5-ave-1 | P Y P E Q **Q E P F** | Arentz-Hansen et al. (2004), Vader et al. (2003) |
| DQ2.5-ave-1b | P Y P E Q **Q Q P F** | Arentz-Hansen et al. (2004), Vader et al. (2003) |

| DQ8 restricted epitopes | | |
|---|---|---|
| **Epitope** | **Motif** | **Reference** |
| DQ8-glia-α1 | **Q** G S F Q P S Q **Q** | van de Wal et al. (1998b) |
| DQ8-glia-γ1a | **Q** Q P Q Q P F P **Q** | Tollefsen et al. (2006) |
| DQ8-glia-γ1b | **Q** Q P Q Q P Y P **Q** | Tollefsen et al. (2006) |
| DQ8-glut-H1 | **Q** G Y Y P T S P **Q** | van de Wal et al. (1999) |

The single letter code for amino acids is used. A characteristic Q-X1-P-X2 motif is present in the large majority of HLA-DQ2 epitopes (in bold). This sequence is a target sequence for TG2, which yields E-X1-P-X2. Due to the introduction of the negatively charged amino acid glutamate, the peptides become high affinity binders for HLA-DQ2. glia-α = α-gliadin; glia-γ = γ-gliadin; hor = hordein; sec = secalin; ave = avenin; glut-L = LWM-glutenin, glut-H = HMW-glutenin.

*Annex A-2 Considerations for sequence identity searches*

Closer inspection of the Q/E-X1-P-X2 sequences demonstrates that only a limited number of amino acids are present at the X-positions: L, Q, F, S or E at position X1 and Y, F, A, V or Q at position X2 (Figure 3). Thus, a search motif that incorporates these amino acids should identify peptide sequences with the potential to bind to HLA-DQ2.5 and stimulate gluten-specific T cells. Additional considerations on position and nature of adjacent amino acid sequences to the Q/E-X1-P-X2 motif should also be taken into account. For example, positively charged amino acids in general diminish the likelihood of DQ-binding and T-cell recognition, in particular positive charge at positions p1, p4, p6, p7 and p9 are detrimental. It is noted that P-P is not identified in T-cell epitopes. In contrast, P-X-P in addition to Q/E-X1-P-X2 is associated with the most immunogenic epitopes.

As discussed in the document, any 9 amino acid-residue peptide which shows identity to a known T-cell epitope might be able to induce an immune response in CD patients. However, the necessary number of amino acids identical in a peptide to trigger a response is challenging to define, because the ability to bind to CD-specific MHC molecules and the interaction with T cells is highly dependent on the nature and position of certain amino acids. Therefore, a definite size cut-off in respect to identity to a known epitope indicating potential hazardous peptides, for which further assessment would be needed, is demanding. Single amino acids substitutions were described to abolish T-cell reactivity in epitopes of α-2 gliadin (Ellis et al., 2003, Mitea et al., 2010). On the other hand, a major alpha gliadin peptide of wheat (glia-α1b, Table A.1) shares five out of nine amino acids with peptides in oat (DQ2.5-ave-1(b)), which are highly suspected to induce immune responses in some celiac disease patients (Vader et al., 2003, Arentz-Hansen et al., 2004). In this case, however, the motif Q/E-X1-P-X2 is present in both sequences.

Considering the search for sequence identity without the motif, based on current knowledge and given that the vast majority of sequences that do not contain the motif Q/E-X1-P-X2 are DQ8 epitopes, it is reasonable to consider allowing three amino acid mismatches. It is highly unlikely that more than three amino acid mismatches in sequences lacking the motif Q/E-X1-P-X2 would result in a safety concern. For DQ8 epitopes, the three amino acids mismatch consideration excludes a potential exchange of glutamines (Q) at position 1 and/or position 9 by glutamic acid (E) where E is known to

enhance binding to HLA-DQ8. It is noted that these considerations might require adaptations based on experience gained in the assessment and on additional knowledge acquired.

Few examples illustrating in practical terms such considerations are provided below.

Two immunodominant T-cell epitopes are present in the α-gliadins with partially overlapping sequences, sequence 1 PFPQPQLPY and sequence 2 PQPQLPYPQ. In these cases, the sequence QLPY and thus the Q-X1-P-X2 motif is present.

– **Example 1, the motif is present.**

The protein of interest contains the following sequence ALPLTQLPASR that we called sequence 3. The comparison of this sequence with sequences 1 (PFPQPQLPY) and 2 (PQPQLPYPQ) is the following.

Comparison of sequences 3 (ALPLTQLPASR) and 1 (PFPQPQLPY):

At position 1 (p1), the P is replaced by A, both being uncharged small amino acids.
At p2, F is replaced by L, both uncharged large amino acids.
At p3, an identical amino acid is present.
At p4, Q is replaced by L, both uncharged large amino acids.
At p5, P is replaced by T, different properties but both small amino acids.
At p6, an identical amino acid is present.
At p7, an identical amino acid is present.
At p8, an identical amino acid is present.
At P9, Y is replaced by A. Here, a large amino acid is replaced by a small amino acid but this does not prohibit binding to HLA-DQ2 (Petersen et al., 2016).
Overall: there is the potential for cross-reactivity of the 9-amino acid sequence ALPLTQLPA, additional investigations are needed.

Comparison of sequence 3 (ALPLTQLPASR) but considering first P as p1 and sequence 2 (PQPQLPYPQ):

At p1, an identical amino acid is present.
At p2, Q is replaced by L, both uncharged large amino acids.
At p3, P is replaced by T, different properties but both small amino acids.
At p4, an identical amino acid is present.
At p5, an identical amino acid is present.
At p6, an identical amino acid is present.
At P7, Y is replaced by A. Here, a large amino acid is replaced by a small amino acid and at this position this is known to eliminate T-cell reactivity (Petersen et al., 2016).
At p8, P is replaced by S and this is known to eliminate T-cell reactivity (Mitea et al., 2012).
At p9, Q is replaced by R, a positively charged amino acid and this is known to have a negative effect of binding to HLA-DQ2.
Overall: there are three significant differences, two of which are independently known to eliminate T-cell reactivity. Altogether, it is highly unlikely that the nine amino acid sequence PLTQLPASR will bind to HLA-DQ2 and stimulate T cells. No additional investigations for this nine amino acid sequence are needed.

– **Example 2, the motif is present.**

The protein of interest contains the following sequence KARGVQSPAEI that we called sequence 4. The comparison of this sequence with sequences 1 (PFPQPQLPY) and 2 (PQPQLPYPQ) is the following.

Comparison of sequences 4 (KARGVQSPAEI) and 1 (PFPQPQLPY):

At p1, the P is replaced by K, a positively charged amino acid and this is known to have a negative effect of binding to HLA-DQ2.
At p2, F is replaced by A. Here a large amino acid is replaced by a small amino acid.
At p3, P is replaced by R, a positively charged amino acid.
At p4, Q is replaced by G. Here a large amino acid is replaced by a small amino acid.
At p5, P is replaced by V.
At p6, an identical amino acid is present.
At p7, L is replaced by S. Here, a large amino acid is replaced with a small amino acid that has different properties.
At p8, an identical amino acid is present.

At P9, Y is replaced by A. Here, a large amino acid is replaced by a small amino acid but this does not prohibit binding to HLA-DQ2 (Petersen et al., 2016).
Overall: there are six significant differences. Altogether, it is highly unlikely that this sequence will bind to HLA-DQ2 and stimulate T cells. No additional investigations are needed.

Comparison of sequences 4 (KARGVQSPAEI) but considering R as p1, and sequence 2 (PQPQLPYPQ):

At p1, the P is replaced by R, a positively charged amino acid and this is known to have a negative effect of binding to HLA-DQ2.
At p2, Q is replaced by G. Here, a large amino acid is replaced by a small amino acid.
At p3, P is replaced by V.
At p4, an identical amino acid is present.
At p5, L is replaced by S. Here, a large amino acid is replaced by a small amino acid that has different properties.
At p6, an identical amino acid is present.
At P7, Y is replaced by A. Here, a large amino acid is replaced by a small amino acid and at this position this is known to eliminate T-cell reactivity (Petersen et al., 2016).
At p8, P is replaced by E, a negatively charged amino acid.
At p9, Q is replaced by I, both large non-charged amino acids.
Overall: there are five significant differences, one of which is known to eliminate T-cell reactivity by its own. Altogether, it is highly unlikely that this sequence will bind to HLA-DQ2 and stimulate T cells. No additional investigations are needed.

– **Example 3, the motif is not present.**

The protein of interest contains the following sequence: EGSIQAGQQ. This sequence has some identity to the HLA-DQ8 epitope QGSFQPSQQ where four mismatches (including an exchange of a Q by an E at position 1) can be identified.

Comparison of sequence EGSIQAGQQ with sequence QGSFQPSQQ:

At p1, Q has been replaced by E, as previously mentioned; this change is known to enhance binding to HLA-DQ8.
At p2, an identical amino acid is present. This is an important T-cell receptor contact residue (Broughton et al., 2012).
At p3, an identical amino acid is present. This is an important T-cell receptor contact residue (Broughton et al., 2012).
At p4, F is replaced by I, both large uncharged amino acids.
At p5, an identical amino acid is present. This is an important T-cell receptor contact residue (Broughton et al., 2012).
At p6, P is replaced by A, both small uncharged amino acids.
At p7, S is replaced by G, both small amino acids but with different properties.
At p8, an identical amino acid is present. This is an important T-cell receptor contact residue (Broughton et al., 2012).
At p9, an identical amino acid is present. This is an HLA-DQ8 anchor residue.
Overall: despite the four amino acid mismatch critical amino acids for HLA-binding and T-cell recognition are conserved (including a replacement at p1 by E). There is the potential for cross-reactivity, additional investigations are needed.

*Annex A-3. Considerations for HLA-DQ peptide modelling*

Recent studies have determined the T-cell receptor repertoire used by CD4+ T cells specific for immunodominant gluten epitopes and crystal structures of such T-cell receptors bound to the HLA-DQ-gluten complexes were determined. As a consequence, detailed knowledge is available indicating which amino acids in the gluten peptides are responsible for the high-affinity binding to HLA-DQ and which amino acids mediate the specific interaction with the T-cell receptor. Several publications describing the binding of CD peptides to HLA-DQ molecules are available. The coordinates of all these structures are available through public databases (Broughton et al., 2012; Henderson et al., 2007; Kim et al., 2004; Petersen et al., 2014; Petersen et al., 2015, 2016; Rossjohn and Koning, 2016). As an example, the x-ray coordinates of HLA-DQ8 bound to a gliadin peptide were deposited in the Protein Data Bank (www.rcsb.org) with the accession number 2NNA while the coordinates and structure factors for the HLA-DQ2-gliadin complexes have been deposited in the Protein Data Bank under the

following accession codes: S2 TCR–DQ2.5-glia-α1a, 4OZI; S16 TCR–HLA-DQ2α2, 4OZH; D2 TCR–HLA-DQ2α2, 4OZG; JR5.1 TCR–HLA-DQ2α2, 4OZF. This extensive information allows modelling studies in which the gluten peptide can be replaced by any peptide sequence of choice (Moustakis et al., 2000; Wiesner et al., 2008; van Heemst et al., 2015). It can be anticipated that, if the selected peptide is likely to bind to HLA-DQ in a way that it resembles the known structure of gluten peptides bound to HLA-DQ, it might have the capacity to stimulate gluten-specific T cells (Figure A.1, Wiesner et al., 2008). Thus, molecular modelling can be employed to aid the determination of potential T-cell stimulatory properties of peptide sequences.

Considering the great relevance of this HLA-DQ peptide modelling step, the future development of a publicly available software tool specifically designed for risk assessment purposes is desirable.

A non-exhaustive list of selected publications reporting on HLA-DQ-modelling assays:

Broughton SE, Petersen J, Theodossis A, Scally SW, Loh KL, Thompson A, van Bergen J, Kooy-Winkelaar Y, Henderson KN, Beddoe T, Tye-Din JA, Mannering SI, Purcell AW, McCluskey J, Anderson RP, Koning F, Reid H and Rossjohn J, 2012. Biased T cell receptor usage directed against human leukocyte antigen DQ8-restricted gliadin peptides is associated with Celiac Disease. Immunity, 37, 611–621.

Henderson KN, Tye-Din JA, Reid HH, Chen Z, Borg NA, Beissbarth T, Tatham A, Mannering SI, Purcell AW, Dudek NL, van Heel DA, McCluskey J, Rossjohn J and Anderson RP, 2007. A structural and immunological basis for the role of human leukocyte antigen DQ8 in celiac disease. Immunity, 27, 23–34.

Kim CY, Quarsten H, Bergseng E, Khosla C and Sollid LM, 2004. Structural basis for HLA-DQ2-mediated presentation of gluten epitopes in celiac disease. Proceedings of the National Academy of Sciences of the United States of America, 101, 4175–4179.

Moustakas AK, van de WY, Routsias J, Kooy YM, Van VP, Drijfhout JW, Koning F and Papadopoulos GK, 2000. Structure of celiac disease-associated HLA-DQ8 and non-associated HLA-DQ9 alleles in complex with two disease-specific epitopes. International Immunology, 12, 1157–1166.

Petersen J, Montserrat V, Mujico JR, Loh KL, Beringer DX, van Lummel M, Thompson A, Mearin ML, Schweizer J, Kooy-Winkelaar Y, van Bergen J, Drijfhout JW, Kan WT, La Gruta NL, Anderson RP, Reid HH, Koning F and Rossjohn J, 2014. T-cell receptor recognition of HLA-DQ2-gliadin complexes associated with celiac disease. Nature Structural and Molecular Biology, 21, 480–488.

Petersen J, Kooy-Winkelaar Y, Loh KL, Tran M, van Bergen J, Koning F, Jamie Rossjohn J and Reid HH, 2016. Diverse T cell receptor gene usage in HLA-DQ8-associated celiac disease converges into a consensus binding solution. Structure, 24, 1643–1657.

Petersen J, van Bergen J, Loh K, Kooy-Winkelaar Y, Beringer DX, Thompson A, Bakker SF, Mulder CJ, Ladell K, McLaren JE, Price DA, Rossjohn J, Reid HH and Koning F, 2015. Determinants of gliadin-specific T cell selection in celiac disease. Journal of Immunology, 194, 6112–6122.

Rossjohn J and Koning F, 2016. A biased view toward celiac disease. Mucosal Immunology, 9, 583–586.

van Heemst J, Jansen DT, Polydorides S, Moustakas AK, Bax M, Feitsma AL, Bontrop-Elferink DG, Baarse M, van der Woude D, Wolbink GJ, Rispens T, Koning F, de Vries RR, Papadopoulos GK, Archontis G, Huizinga TW and Toes RE, 2015. Crossreactivity to vinculin and microbes provides a molecular basis for HLA-based protection against rheumatoid arthritis. Nature Communications, 6, 6681.

Wiesner M, Stepniak D, de Ru AH, Moustakis AK, Drijfhout JW, Papadopoulos GK, van Veelen PA and Koning F, 2008. Dominance of an alternative CLIP-sequence in the celiac disease associated HLA-DQ2 molecule. Immunogenetics, 60, 551–555.

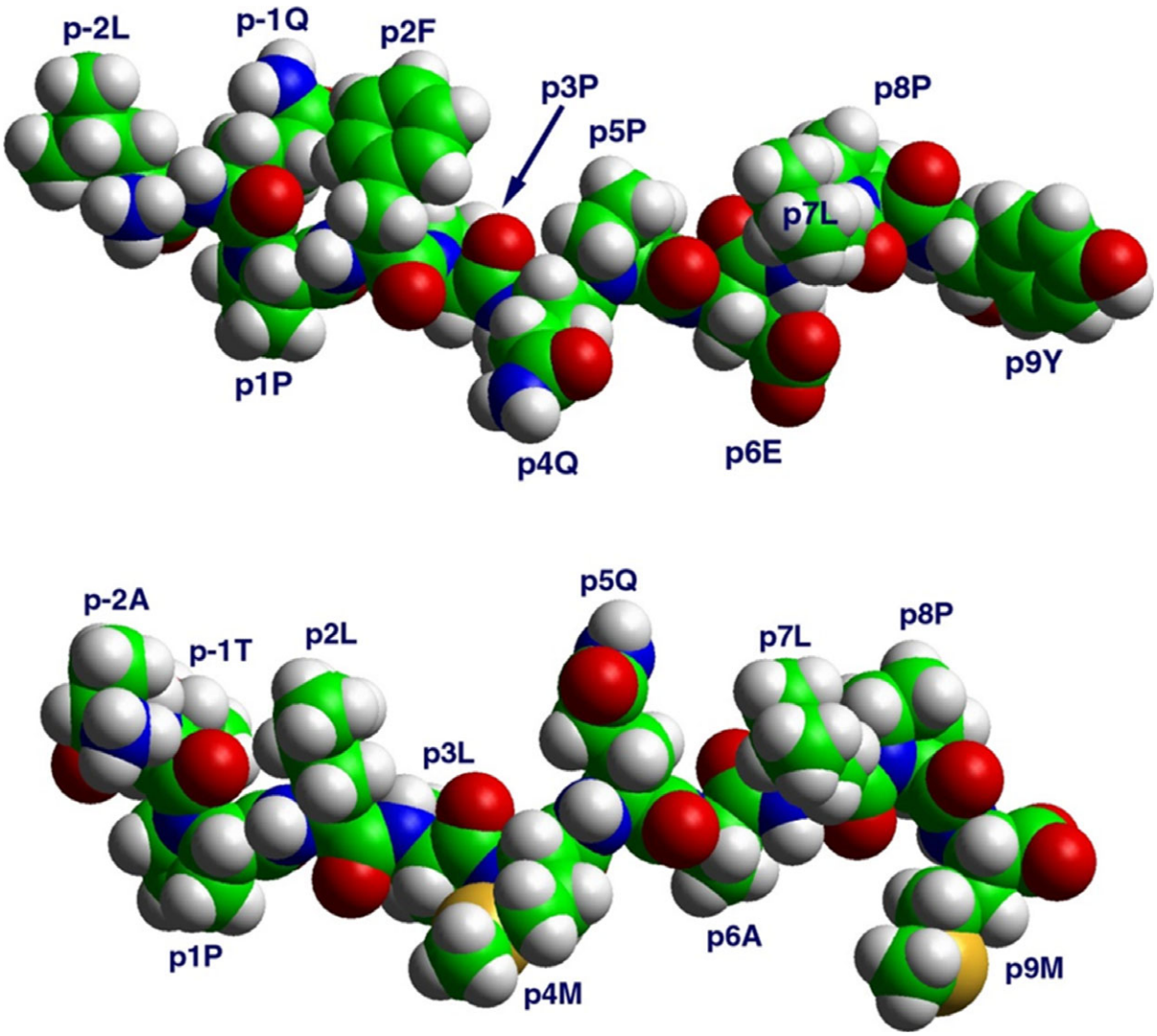



**Figure A.1:** Shown is a side view of two peptides, one a gliadin T-cell epitope (top panel) while the other is a peptide derived from a self-antigen (lower panel). The side chains of the amino acids that point downwards anchor the peptide to the HLA-DQ-molecule, the upward-pointing amino acids can be contacted by the T-cell receptor. Even though the peptide share some sequence similarity (P at p1, L at P7, P at p8, large amino acid at p2), there are several features that are predicted to prohibit a functional interaction between a gliadin specific T-cell receptor and the self-peptide, like an overall different conformation and the presence of a large amino acids at p5 (Q) instead of the much smaller proline (P) in the gliadin peptide (taken from Wiesner et al., 2008)

*Annex A-4. Considerations for HLA-DQ peptide binding assays*

Binding affinities in HLA-DQ-peptide binding assays have been previously reported in the literature (e.g. Vader et al., 2003). A non-exhaustive list of publications that have reported on HLA-DQ-peptide binding assays:

Ettinger RA and Kwok WW, 1998. A peptide binding motif for HLA-DQA1*0102/DQB1*0602, the class II MHC molecule associated with dominant protection in insulin-dependent diabetes mellitus. Journal of Immunology, 160, 2365–2373.

Johansen BH, Gjertsen HA, Vartdal F, Buus S, Thorsby E, Lundin KE and Sollid LM, 1996. Binding of peptides from the N-terminal region of alpha-gliadin to the celiac disease-associated HLA-DQ2 molecule assessed in biochemical and T cell assays. Clinical Immunology and Immunopathology, 79, 288–293.

Johansen BH, Vartdal F, Eriksen JA, Thorsby E and Sollid LM, 1996. Identification of a putative motif for binding of peptides to HLA-DQ2. International Immunology, 8, 177–182.

Johansen BH, Jensen T, Thorpe CJ, Vartdal F, Thorsby E and Sollid LM, 1996. Both alpha and beta chain polymorphisms determine the specificity of the disease-associated HLA-DQ2 molecules, with beta chain residues being most influential. Immunogenetics, 45, 142–150.

Kwok WW, Nepom GT and Raymond FC, 1995. HLA-DQ polymorphisms are highly selective for peptide binding interactions. Journal of Immunology, 155, 2468–2476.

Nepom BS, Nepom GT, Coleman M and Kwok WW, 1996. Critical contribution of beta chain residue 57 in peptide binding ability of both HLA-DR and -DQ molecules. Proceedings of the National Academy of Sciences of the United States of America, 93, 7202–7206.

Stepniak D, Wiesner M, de Ru AH, Moustakas AK, Drijfhout JW, Papadopoulos GK, van Veelen PA and Koning F, 2008. Large-scale characterization of natural ligands explains the unique gluten-binding properties of HLA-DQ2. Journal of Immunology, 180, 3268–3278.

Straumfors A, Johansen BH, Vartdal F, Sollid LM, Thorsby E and Buus S, 1998. A peptide-binding assay for the disease-associated HLA-DQ8 molecule. Scandinavian Journal of Immunology, 47, 561–567.

Terreaux C, Walk T, van de Wal Y, Koning F, Jung G and Fleckenstein B, 1998. Increased HLA-DQ2-affinity of a synthetic gliadin peptide by acid-induced deamidation of glutamine residues. Bioorganic and Medicinal Chemistry Letters, 8, 2039–2044.

Vader W, Stepniak D, Kooy Y, Mearin L, Thompson A, van Rood JJ, Spaenij L and Koning F, 2003. The HLA-DQ2 gene dose effect in celiac disease is directly related to the magnitude and breadth of gluten-specific T cell responses. Proceedings of the National Academy of Sciences of the United States of America, 100, 12390–12395.

van de Wal Y, Kooy YM, Drijfhout JW, Amons R, Papadopoulos GK and Koning F, 1997. Unique peptide binding characteristics of the disease-associated DQ(alpha 1*0501, beta 1*0201) vs the non-disease-associated DQ (alpha 1*0201, beta 1*0202) molecule. Immunogenetics, 46, 484–492.

van de Wal Y, Kooy YM, Drijfhout JW, Amons R and Koning F, 1996. Peptide binding characteristics of the coeliac disease-associated DQ(alpha1*0501, beta1*0201) molecule. Immunogenetics, 44, 246–253.

van de Wal Y, Amons R and Koning F, 2000. Characterization of HLA-DQ-specific peptide-binding motifs. Methods in Molecular Medicine, 41, 97–103.

Vartdal F, Johansen BH, Friede T, Thorpe CJ, Stevanović S, Eriksen JE, Sletten K, Thorsby E, Rammensee HG and Sollid LM, 1996. The peptide binding motif of the disease associated HLA-DQ (alpha 1* 0501, beta 1* 0201) molecule. European Journal of Immunology, 26, 2764–2772.

Wiesner M, Stepniak D, de Ru AH, Moustakis AK, Drijfhout JW, Papadopoulos GK, van Veelen PA and Koning F, 2008. Dominance of an alternative CLIP sequence in the celiac disease associated HLA-DQ2 molecule. Immunogenetics, 60, 551–555.

*Annex A-5. Non-exhaustive list of publications that have reported on gluten-specific T cells*

Anderson RP, Degano P, Godkin AJ, Jewell DP and Hill AV, 2000. In vivo antigen challenge in celiac disease identifies a single transglutaminase-modified peptide as the dominant A-gliadin T-cell epitope. Nature Medicine, 6, 337–342

Arentz-Hansen H, Fleckenstein B, Molberg, Scott H, Koning F, Jung G, Roepstorff P, Lundin KE and Sollid LM, 2004. The molecular basis for oat intolerance in patients with celiac disease. Plos Medicine, 1, e1.

Arentz-Hansen H, Körner R, Molberg, Quarsten H, Vader W, Kooy YM, Lundin KEA, Koning F, Roepstorff P, Sollid LM and McAdam SN, 2000. The intestinal T cell response to α-gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase. Journal of Experimental Medicine, 191, 603–612.

Molberg, McAdam SN, Körner R, Quarsten H, Kristiansen C, Madsen L, Fugger L, Scott H, Roepstorff P, Lundin KEA, Sjöström H and Sollid LM, 1998. Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells. Nature Medicine, 4, 713–717.

Moustakas AK, van de WY, Routsias J, Kooy YM, Van VP, Drijfhout JW, Koning F and Papadopoulos GK, 2000. Structure of celiac disease-associated HLA-DQ8 and non-associated HLA-DQ9 alleles in complex with two disease-specific epitopes. International Immunology, 12, 1157–1166.

Petersen J, Montserrat V, Mujico JR, Loh KL, Beringer DX, van Lummel M, Thompson A, Mearin ML, Schweizer J, Kooy-Winkelaar Y, van Bergen J, Drijfhout JW, Kan WT, La Gruta NL, Anderson RP, Reid HH, Koning F and Rossjohn J, 2014. T-cell receptor recognition of HLA-DQ2-gliadin complexes associated with celiac disease. Nature Structural and Molecular Biology, 21, 480–488.

Quarsten H, Molberg, Fugger L, McAdam SN and Sollid LM, 1999. HLA binding and T cell recognition of a tissue transglutaminase-modified gliadin epitope. European Journal of Immunology, 29, 2506–2514.

Shan L, Molberg, Parrot I, Hausch F, Filiz F, Gray GM, Sollid LM and Khosla C, 2002. Structural basis for gluten intolerance in celiac sprue. Science, 297, 2275–2279.

Sjöström H, Lundin KEA, Molberg, K"rner R, McAdam SN, Anthonsen D, Quarsten H, Noren O, Roepstorff P, Thorsby E and Sollid LM, 1998. Identification of a gliadin T-cell epitope in coeliac disease: general importance of gliadin deamidation for intestinal T-cell recognition. Scandinavian Journal of Immunology, 48, 111–115.

Tollefsen S, Arentz-Hansen H, Fleckenstein B, Molberg, Raki M, Kwok WW, Jung G, Lundin KE and Sollid LM, 2006. HLA-DQ2 and -DQ8 signatures of gluten T cell epitopes in celiac disease. The Journal of Clinical Investigation, 116, 2226–2236.

Tye-Din JA, Stewart JA, Dromey JA, Beissbarth T, van Heel DA, Tatham A, Henderson K, Mannering SI, Gianfrani C, Jewell DP, Hill AV, McCluskey J, Rossjohn J and Anderson RP, 2010. Comprehensive, quantitative mapping of T cell epitopes in gluten in celiac disease. Science Translational Medicine, 2, 41–51.

Vader LW, de Ru A, van Der WY, Kooy YM, Benckhuijsen W, Mearin ML, Drijfhout JW, van Veelen P and Koning F, 2002a. Specificity of tissue transglutaminase explains cereal toxicity in celiac disease. Journal of Experimental Medicine, 195, 643–649.

Vader LW, Stepniak DT, Bunnik EM, Kooy YM, de Haan W, Drijfhout JW, van Veelen PA and Koning F, 2003. Characterization of cereal toxicity for celiac disease patients based on protein homology in grains. Gastroenterology, 125, 1105–1113.

Vader W, Kooy Y, van Veelen P, de Ru A, Harris D, Benckhuijsen W, Pena S, Mearin L, Drijfhout JW and Koning F, 2002b. The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides. Gastroenterology, 122, 1729–1737.

van de Wal Y, Kooy Y, van Veelen P, Pena S, Mearin L, Papadopoulos G and Koning F, 1998a. Selective deamidation by tissue transglutaminase strongly enhances gliadin-specific T cell reactivity. The Journal of Immunology, 161, 1585–1588.

van de Wal Y, Kooy YM, van Veelen P, Vader W, August SA, Drijfhout JW, Pena SA and Koning F, 1999. Glutenin is involved in the gluten-driven mucosal T cell response. European Journal of Immunology, 29, 3133–3139.

van de Wal Y, Kooy YM, van Veelen PA, Pena SA, Mearin LM, Molberg, Lundin KEA, Sollid LM, Mutis T, Benckhuijsen WE, Drijfhout JW and Koning F, 1998b. Small intestinal T cells of celiac disease patients recognise a natural pepsin fragment of gliadin. Proceedings of the Nationall Academy of Sciences United States of America, 95, 10050–10054.

### *Annex A-6. Considerations for in vitro* digestibility tests

As specified above and because a 9-amino acid core is almost invariably required for efficient peptide binding to HLA-DQ, proteins that are easily degraded into fragments shorter than 9 amino acids are unlikely to harbour T-cell stimulatory epitopes. Thus, determination of proteolytic resistance of proteins can aid in the identification of proteins with potential harmful potential. For more details, please see chapter on *in vitro* protein digestibility in this document (Annex B).

# Annex B – *In vitro* protein digestibility tests

As specified in Section 2.2 of this document, an interim phase (~2 years duration) is proposed to assess the utility of an enhanced *in vitro* digestion test, with a review undertaken in the final year. During this phase, the laboratory(ies) involved will detail and apply the refined digestion test methodology. After this period, EFSA will discuss whether the test adds value and, if so, what further steps are needed for its final implementation.

## Annex B-1. Background

*In vitro* digestibility tests can provide useful data on the susceptibility of a protein to digestion which may reflect its digestibility in the human gastrointestinal system and subsequent presentation to the host's immune system (Foster et al., 2013). There is evidence that gastrointestinal digestion can affect the immunogenicity of dietary proteins in relation to both IgE and non-IgE-mediated adverse reactions to foods as discussed below. However, notably, the ability of digestion-resistant dietary proteins or derived fragments to initiate diseases, such as IgE-mediated allergies or celiac disease (CD), also depends on the predisposition of the host. For CD, expression of certain HLA genotypes is a well described risk factor for developing the condition (See Section 2.1). However, while a genetic predisposition towards atopy is also thought to play a role in the development of IgE-mediated allergies to a variety of environmental agents including foods, specific genetic and other risk factors involved in the development of specific IgE have yet to be defined. This makes the risk assessment process less certain for IgE-mediated allergies compared to CD.

*IgE-mediated adverse reactions to foods:* The events leading to the breakdown of oral tolerance, allergic sensitisation and development of food allergies in susceptible individuals are poorly understood. It is likely that multiple pathways could ultimately lead to a failure to develop or loss of oral tolerance (Chinthrajah et al., 2016). Notwithstanding the role of route of exposure in sensitisation, understanding how proteins are presented to the gastrointestinal mucosal immune system may provide insight into the mechanisms controlling the balance between tolerisation and sensitisation. Furthermore, oral exposure is also central to non-IgE assessments where gastrointestinal digestion has been demonstrated to be important in delivery of immunologically active fragments to gastrointestinal mucosal sites. Impaired gastric digestion of food allergens has been associated with both development of IgE responses to foods (sensitisation) and modulating the severity of IgE-mediated reactions to foods (elicitation). A recent study of patients undergoing gastric-bypass surgery, a procedure known to reduce post-prandial gastric acidity and where the bulk of food ingested reaches the small intestine without prior gastric digestion, resulted in a significant increase in sensitisation to food (Shakeri-Leidenmuhler et al., 2015). Impaired digestion of cod fish proteins may also be a risk factor for severe reactions in fish allergic individuals (Untersmayr et al., 2007). Studies on animal models have shown that suppression of gastric acid secretion using widely prescribed antacid medication increases propensity to sensitisation to proteins from cod fish (Untersmayr et al., 2003), celery (Untersmayr et al., 2008), hazelnut (Scholl et al., 2005) and egg (Diesner et al., 2008). Furthermore, an association between the use of antiulcer drugs and the induction of IgE-mediated allergy to a variety of foods such as milk, potato, celery, carrots, apple, orange, wheat, and rye flour has been reported (Untersmayr et al., 2005). Such studies suggest that elevated gastric pH, and the resulting reduction in peptic digestion, may enhance the potential of foods to cause allergies. Uptake of allergens into the circulation may play an important role in eliciting allergic reactions (Strait et al., 2011) but data in humans are sparse and conflicting as to the role of allergen uptake in triggering reactions, although uptake appears much greater in allergic than healthy individuals, as indicated in studies of wheat allergy (Brockow et al., 2015; Matsuo et al., 2005).

*Non-IgE-mediated adverse reactions to foods:* Studies seeking to define the structural basis of the toxicity of gluten in CD have made extensive use of *in vitro* gastrointestinal digestion, including the action of brush border proteases, to generate physiologically relevant fragments of gluten. These studies notably identified a 33mer peptide that is especially resistant to digestion and which is likely to persist in the brush border with a half-life estimated to be ~20 h. This peptide is also an excellent substrate for tissue transglutaminase TG2 and was a potent stimulator of T cells from celiac patients (Shan et al., 2002).

Such data support the premise that immunologically active fragments of food proteins persisting in the gastrointestinal lumen can play a role in driving immune-mediated adverse reactions to foods. Persistent, soluble intact proteins and fragments are more likely to be sampled by the gastrointestinal

epithelium and are hence exposed to cells of the immune system as potentially immunogenic fragments, with particulates being sampled by M cells within the Peyer's patches. In animal models, there is evidence that uptake of particulate peanut protein bodies by M cells may promote sensitisation by virtue of their size, with soluble material being taken up by the same route (Chambers et al., 2004). Although the exact mechanism(s) of antigen uptake in the intestine are still to be clearly elucidated (Pabst and Mowat, 2012), it is generally accepted that most food allergens should retain sufficient structural integrity as they pass through the human gastrointestinal tract to sensitise or elicit an allergic response (Metcalfe et al., 1996). After uptake by antigen presenting cells, peptides generated through endosomal proteolysis bind to MHC class II molecules (Rudensky et al., 1991; Chicz et al., 1993). Peptides shorter than 9 amino acid residues are probably unable to bind to MHC class II molecules and to activate T cells because there is a minimum length requirement – the so called 'peptide binding register' (Mohan and Unanue, 2012). Thus, peptides shorter than 9 amino acid residues are highly unlikely to stimulate an immune response. However, unlike celiac disease, IgE-mediated allergies involve development of humoral (i.e. antibody) responses which requires activation of both B cells and T cells. The B-cell receptors involved in B-cell activation have a requirement for multivalent antigen and it has emerged that the way in which B cells encounter antigen *in vivo* depends on its properties, such as size (Harwood et al., 2010). As a consequence, larger peptide sizes are required for development of IgE responses. Thus, digestion of the allergen β-lactoglobulin into fragments of less than 3,000 Da abolished the proteins immunogenicity (Bogh et al., 2013), while peptides of less than 1,500 Da resulting from pepsinolysis of peanut Ara h 1, had markedly reduced immunogenicity, which was completely lost when the digest was fractionated (Bogh et al., 2012). With regard to elicitation of reactions, synthetic antigens have been used to investigate the mechanism of degranulation (Handlogten et al., 2013). These studies have shown that a minimum of two distinct epitopes is required to trigger degranulation, with the epitopes separated by a maximum of 6.4 nm on the synthetic antigen. On this basis peptides of at least 3–5 kDa would be required to cross-link IgE bound to the surface of effector cells, such as mast cells and basophils. However, peptides as small as 791 Da and 1413 Da resulting from the pepsinolysis of the avocado pear allergen Prs s 1, have been found to inhibit IgE binding to crude avocado extract in ELISA-inhibition assays and to elicit reactions *in vivo* using skin testing, respectively (Diaz-Perales et al., 2003). Such peptides are unlikely to carry multiple IgE epitopes, suggesting that peptide aggregation may play a role in triggering degranulation.

To either sensitise or trigger an allergic reaction in an already sensitised individual, a food protein needs to be 'bioaccessible' by the hosts' immune system. 'Bioaccessibility' describes the ability of a chemical entity (such as a protein) to be released from food during the digestive process, which can consequently interact with and/or be absorbed by the gastrointestinal epithelium (Holst and Williamson, 2008). The EFSA Guidance Document for risk assessment of food and feed from genetically modified plants published in 2011 states that 'the impact of the possible interaction between the protein and other components of the matrix as well as the effects of the processing should be taken into account in *in vitro* digestibility tests' (EFSA GMO Panel, 2011). *In vitro* digestibility tests have been applied to investigate the effects of digestion on the food matrix, and how processing conditions (including thermal treatment) may affect susceptibility to simulated gastrointestinal proteolysis (e.g. Minekus et al., 2014, Smith et al., 2015). However, given the diversity of food matrices and food processing procedures, our knowledge of their effects on susceptibility of proteins to digestion is limited. As a consequence, the effects of processing and of the food matrix on the susceptibility of a particular protein to digestion are difficult to predict. Because there is no effective animal model for food allergy, studies are often limited to investigating the impact of the food matrix on elicitation of allergic reactions in allergic human subjects while data on allergic sensitisation is very limited regarding this aspect (Ballmer-Weber et al., 2002; Bartnikas et al., 2012; Brenna et al., 2000; Grimshaw et al., 2003; Mackie et al., 2012; Netting et al., 2013; Worm et al., 2009).

## Annex B-2. Types of *in vitro* digestibility tests

The pepsin resistance test, which is embedded in EFSA Guidance Document (EFSA GMO Panel, 2011) and Codex Alimentarius (2003, 2009), is currently used for allergenicity risk assessment as an additional piece of information within the weight-of-evidence approach, because no single test is fully predictive of the allergenic potential of a protein. The classical pepsin resistance test has several limitations (EFSA GMO Panel, 2010) including:

- the pH value usually employed in the assay is extremely acidic. Because pepsin activity is pH dependent, the pattern of proteolysis may not reflect the one likely to be found *in vivo* other than at the end of gastric emptying or in the fasted state;
- pepsin is added in a gross excess to the protein substrate, affecting the kinetics of the digestion;
- the correlation with allergenicity of proteins has been questioned. Studies comparing the digestibility of allergens with that of non-allergenic dietary proteins showed that food allergens were not always inherently more stable to pepsin digestion than non-allergenic proteins (Fu et al., 2002; Thomas et al., 2004; Herman et al., 2007).

While this test may contribute to understanding the biochemical properties of newly expressed proteins (information on the physicochemical stability of a protein, such as the rigidity of the polypeptide backbone at low pH, determine susceptibility to pepsinolysis), it does not allow to characterise how these proteins behave under the overall physiological conditions encountered in the digestive tract. Furthermore, the test does not reflect changes in the digestive process that take place across the life course (Rémond et al., 2015). These limitations were previously highlighted by EFSA (EFSA GMO Panel, 2010). In addition, the EFSA Guidance Document for risk assessment of food and feed from genetically modified plants (EFSA GMO Panel, 2011) indicates that 'the digestibility of the newly expressed proteins in specific segments of the population such as infants and individuals with impaired digestive functions may be assessed employing *in vitro* digestibility tests using different conditions'. Given these considerations, it is proposed to embed the classical pepsin resistance test within a suite of tests aimed to characterise how a newly expressed protein may behave during the digestive process (see Annex B-5).

Gastrointestinal digestion assays usually aim to simulate 'normal' digestive function with regard to digestive enzymes and acid secretion (Macierzanka et al., 2009; Moreno, 2007; Minekus 2014). However, alterations in the digestive milieu are often observed in patients with various gastrointestinal conditions (Kay and Jorgensen, 1994), in young infants with underdeveloped digestive system (Armand et al., 1996) and the elderly with a weakened digestive function (Hosking et al., 1975). In an *in vitro* study assessing the impact of digestion on celery allergenicity in an aged population, decreased gastric proteolysis was identified (Untersmayr et al., 2008). Simulated infant digestion models (Dupont et al., 2010; Ménard et al., 2014) were used to study the gastrointestinal tolerance of casein and α-lactalbumin and to optimise the milk processing and formula production. Experiments using simulated gastric fluid where the pH was increased were used to assess the stability of allergens in fish, milk and hazelnut in patients taking antacids where intragastric pH is increased (Scholl et al., 2005; Untersmayr et al., 2005; Untersmayr et al., 2007). Another study developed a gastrointestinal model simulating the physicochemical conditions of the elderly's gastrointestinal tract which was applied to investigate the fate of bovine whey proteins (Levi and Lesmes, 2014). Finally, the COST Infogest network has proposed a standardised batch gastrointestinal digestion method based on physiologically relevant conditions that could be applied for various endpoints (Minekus et al., 2014).

*In vitro* models simulating the physiological conditions of gastrointestinal digestion (in either healthy or diseased individuals) by sequential addition of digestive enzymes, biosurfactants and fluids have been designed to understand the degradation of proteins and other constituents during digestion. Some *in vitro* models have also included biosurfactants, such as phosphatidylcholine, which can be found at low levels in gastric juice and which is also a component of bile (Moreno et al., 2005a; Mandalari et al., 2009a). Biosurfactants can have complex effects on the phase behaviour of lipid rich foods, because emulsification is an important aspect of lipid digestion (Macierzanka et al., 2009), but they may also have effects on digestion of proteins associated with lipids, and there is some evidence, although equivocal, that bile salts may affect the activity of pancreatic proteases. For instance, the activity of trypsin and chymotrypsin is increased in the presence of bile salts (Robic et al., 2011).

The time course of simulated digestion tests can be based on the residence time of food in the stomach. This is dependent on the type of meal ingested. For example, liquid and solid meals display different gastric emptying rates after ingestion (Rémond et al., 2015). The half-time ($t_{1/2}$), which indicates when 50% of an ingested meal is emptied, ranges from 10 to 60 min for liquid meals, whereas $t_{1/2}$ values reported for solid foods ranges from 50 to 115 min. Other factors, such as other meal components, meal volume, caloric content, ratio between liquid and solid in the meal or the type of dietary fibres also have an influence on the gastric emptying rate.

Digestibility studies using gastrointestinal conditions can provide useful data regarding persistency of newly expressed proteins and/or of digestion derived immunologically active fragments in the

gastrointestinal lumen which may pose a risk of causing an immune-mediated adverse reaction in a susceptible individual. The presence of digestion resistant fragments only provides an indication of exposure of the gastrointestinal mucosal surface and is not on its own predictive of allergenicity, because this property is a function of the way in which the fragments interact with the individual. Consequently, resistance to gastrointestinal digestion of a newly expressed protein should be considered as part of the assessment for its potential to cause allergic reactions via the oral route. However, it should also be noted that other non-oral routes of exposure such as respiratory or cutaneous have to be considered (EFSA GMO Panel, 2010).

There is no internationally accepted model/protocol available to perform gastrointestinal *in vitro* digestibility tests for purified proteins although this has been developed for whole foods (Minekus, 2014). This is a consensus model applied to several foods and based on available *in vivo* physiological data resulting from the COST Infogest network. This batch *in vitro* digestion assay includes an oral phase, as well as subsequent gastric and intestinal phases. Interlaboratory trials were performed at the European level to assess digestion of skim milk powder (Egger et al., 2016). Such *in vitro* digestibility methods will require adaptation to make them useful for analysis of purified proteins, taking into account the need for further standardisation and validation.

There are only two validated (ring-trialled) studies of digestibility tests for purified proteins which were published up to date:

- Thomas et al. (2004). Modification of the seminal paper of Astwood et al., (1996) by lowering the pepsin:protein ratio and looking at the effect of very acidic pH values (1.2 and 2.0). *In vitro* pepsinolysis of 10 proteins (allergens and non-allergens) at pH 1.2 and 2.0 were evaluated by nine laboratories. The authors observed that pH did not have an influence on the time of digestion of protein large-fragments but the detection by gel electrophoresis of low molecular weight proteolytic fragments was less consistent because the different fixation and staining methods used;
- Mandalari et al. (2009b). These authors evaluated the *in vitro* gastrointestinal digestibility of β-casein and β-lactoglobulin by using a low-protease assay with and without the addition of phospholipids (based on Moreno et al., 2005a,b and Mandalari et al., 2009a) and the high-protease assay (based on Astwood et al., 1996 and Fu et al., 2002). Twelve laboratories tested the method without the addition of phospholipids and five labs studied the effect of the addition of surfactants. This study demonstrated that the low-protease assay was robust and reproducible although further validation should be undertaken using a more extensive panel of proteins. In addition, this interlaboratory trial showed that the largest factor governing irreproducibility was the sampling and electrophoresis methods used to analyse digestion products.

# Annex B-3. *In vivo* pH range and fluids compositions determined in the gastrointestinal tract

## B-3.1. pH conditions in the gastrointestinal tract

There is an increasing body of data on the *in vivo* fluctuations in pH along the gastrointestinal tract coming from telemetry studies using single use capsules which house pH and temperature sensors and a wireless transceiver capable of reporting pH measurements and temperature values with high precision. Nevertheless, much of the data on infants comes from intubation studies where there are concerns that the placement of probes in the stomach, may have resulted in an underestimation of intragastric pH and an over-estimation of the buffering effect of food in pre-term infants (Omari and Davidson, 2003). Since the presence of food affects luminal conditions, both by virtue of its own buffering capacity and the stimulatory effect food has on gastric and pancreatic secretions, conditions are summarised below for both the fasted and the fed state.

### B-3.1.1. Stomach

Fasted state: Telemetry studies have shown wide fluctuations in intragastric pH in adults, ranging from pH 1 to 8 with median values of 1.4–4.6 and an overall mean pH of 2.7 (±0.8) (n = 20) (Koziolek et al., 2015). Other studies also showed fasting intragastric pH to be ~ 2.0 in a group of 12 subjects (Banerjee et al., 2010), or ranges of 1.5–3 with seven subjects (Ono et al., 2009) and 1.0–2.5 (n = 66) (Evans et al., 1988). These study groups were dominated by male subjects so no findings were

available to show any gender variations. Concerning infants, a recent meta-analysis of data from both pre- and full-term infants indicated the mean intragastric pH in the fasted infant was 2.8 (±0.9) (n = 102) (Kamstrup et al., 2017).

Fed state: After consuming a meal, the intragastric pH increases due to the buffering capacity of the food. Thus, intragastric pH in adults is generally around 5.5 and slowly drops to around 1.5 at the end of gastric emptying (Michalek et al., 2011). In one group of healthy full-term newborns (n = 25), the pH increased to ~ 6.4 after feeding and gradually decreased to ~pH 3.5 (Mason, 1962). A second group of 15 healthy preterm infants found that the intragastric pH rose to around 7 after feeding dropping to 1.5–2.0 at the end of gastric emptying (Omari and Davidson, 2003). A meta-analysis of a number of studies on the changes in intragastric pH of pre- and full-term infants shows the same patterns as adults with a higher pH (~ 6.5) immediately after feeding, which reduces over time to between pH 2.0–4.5 depending on the study (Bourlieu et al., 2014). A second meta-analysis of data collected in infants (pre- and full-term) showed that the mean intragastric pH in the fed infant is 6.4 (±0.6) (n = 102) (Kamstrup et al., 2017).

Drug interventions with proton pump inhibitors (PPI), such as omeprazole or lansoprazole, are used to treat peptic ulcers and gastroesophageal reflux disease in adults and infants (Omari et al., 2007). Their use results in an increase in fasting gastric pH to 6.0 although the length of treatment required depends on the dose and form in which the drugs are given (Ono et al., 2009; Banerjee et al., 2010). Similar effects are observed in infants (Faure et al., 2001). The fed-state intragastric pH of individuals taking PPIs remains around 6.0 and does reduce but only to around 4.5 at the end of gastric emptying (n = 10 and n = 7) (Ono et al., 2009; Banerjee et al., 2010).

#### B-3.1.2. Small intestine

An early study showed duodenal pH to range from 5.0 to 7.2 with a median value of 6.4 which rose to 7.3 in the distal small intestine (n = 39) (Fallingborg et al., 1989). In a more recent study, fasted measurements of pH in the proximal segment of the small bowel gave values of 5.9–6.3 with a mean value of 6.0 (± 0.2) while in the distal intestine the pH was higher at around 7.4–7.8 with a mean value of 7.7 (± 0.15) (n = 20) (Koziolek et al., 2015). In another study, duodenal pH in the fed state was found to be 5.2–6.1 (mean of 5.6, n = 17) and did not seem to alter greatly in either the fasted state or in patients with dyspepsia (Bratten and Jones, 2009). The pH in the distal ileum has also been found to be ~ 8.1 in both the fed and fasted state. In another study, the mean pH in the proximal small intestine measured in 55 normal subjects was 6.63 (± 0.53), whereas in the terminal ileum was 7.49 (±0.46) (n = 58) (Evans et al., 1988).

There is very little data available on the pH of the intestine in infants (either pre- or full-term) with only two studies reported. These data indicate that the intestinal pH of infants is 6.6–7.0 and only drops slightly to about 6.4 after feeding (Kamstrup et al., 2017).

### B.3.2. Composition and concentration of digestive enzymes in gastric and intestinal fluids

The current state-of-the-art reveals that gastrointestinal digestion of proteins is a multistage process, starting with pepsin cleavage in stomach and proceeding through to the small intestine with digestion taking place through the action of a mixture of pancreatic endoproteases (such as trypsin, chymotrypsin and elastase) together with amino- and carboxy-peptidases. Finally, at the intestinal epithelial surface further digestion takes place through the action of brush border aminopeptidases (Akimov and Bezuglov, 2012). Collectively, these enzymes reduce proteins down to free amino acids and the simple di- and tripeptides which can be taken up by transporters on epithelial cell surfaces.

Data on the composition of gastric and intestinal fluids are generally obtained from analysis of aspirates taken through invasive intubation procedures and consequently data from healthy individuals are sparse. Furthermore, analysis of fluid composition has generally been undertaken to identify disease biomarkers. Thus, in contrast to gastric and intestinal pH which is directly measured, the levels of digestive enzymes are normally reported either in activity levels (directly measured in the gastrointestinal chyme but using different and non-standardised activity assays, as well as different definitions of enzyme units), or in enzyme outputs (flow rates). Interpretation of such data, which is obtained from fluids generally collected from individuals either in the fasted state or after treatment to stimulate pancreatic secretion, is further complicated by the fact the caloric content, nutrient composition, and physical properties of the ingested meal are all known to be factors that can strongly

influences pancreatic enzyme secretory responses (Keller and Layer, 2005). Therefore, the comparison of data from different studies is challenging and makes it difficult to identify well-defined levels and activities of digestive enzymes to be included in *in vitro* test systems.

#### B.3.2.1. Gastric juice

Median values for the pepsin content of gastric aspirates in a study of 20 fasted healthy individuals ranged from 0.11 to 0.22 mg/mL depending on sampling time (Kalantzi et al., 2006). After eating, the levels increased from 0.26 to 0.58 mg/mL (Kalantzi et al., 2006). In another study, pepsin activity was determined using haemoglobin as a standard and found to be 349.1 (±14.5) μg/mL using porcine pepsin A (2,100 U/mg) as a standard (n =343) (Metheny et al., 1997). One drawback of the later study is that the patient population was heterogeneous (aged 15–95 years; approximately 54% were > 60 years of age). Around 50% were receiving H2-receptor antagonists and eight omeprazole, a PPI. Post-prandial measurement indicated that there was 87 U/mL pepsin in preterm infants (n = 28) (Armand et al., 1996) as compared to levels of 942–1,333 U/mL in adults (n = 6), levels depending on diet (Armand et al., 1995). In both the Armand and Metheny studies, pepsin activity was measured in gastric aspirates and using haemoglobin as a substrate. The much lower levels of pepsin, coupled with higher intragastric pH mean that gastric digestion in infants is very limited (Bourlieu et al., 2014).

The other major gastric enzyme is human gastric lipase. Levels of both gastric lipase and pepsin increase in infancy as the gastrointestinal tract matures, reaching those of adults by the age of 12 months (Bourlieu et al., 2014). In some instances, gastric fluid can contain bile and often contains salivary proteins and salivary mucins. The gastric mucus layer contains of surfactant, believed to enhance its barrier properties with the dominant lipid being phosphatidyl choline and phosphatidyl ethanolamine (Schmitz and Renooij, 1990). As a consequence, phospholipids find their way into the lumen, with contents of dipalmitoyl phosphatidyl choline ranging between 49 and 95 μg/mL in a group of 68 children (Chang et al., 2006). In another study, the concentration levels of total phospholipids found in the gastric juice of healthy adults ranged from 60.8 to 206.4 μg/mL (n = 15) (Wenner et al., 2000).

#### B.3.2.2. Pancreatic juice

Pancreatic secretions are highly complex mixtures of proteins, including a wide range of endoproteases and carboxy- and aminopeptidases, as recent proteomic profiling experiments have demonstrated. For example, proteomic profiling of gastroduodenal fluid collected during endoscopy allowed 71–89 proteins to be identified in six subjects demonstrated the presence of trypsin, enteropeptidase, chymotrypsin and elastase although many of the isoforms were not identified (Paulo et al., 2010). The gastrointestinal fluid also showed the presence of salivary proteins, such as the basic proline-rich salivary protein, together with a variety of mucins.

Analysis of pancreatic juice from healthy individuals sampled during endoscopic retrograde cholangiopancreatography has been undertaken in a limited number of individuals. One study in three female subjects identified 90 proteins with a high degree of confidence using MS–MS analysis (Doyle et al., 2012). Digestive enzymes identified that degrade proteins include carboxypeptidase A1 and A2, chymotrypsin C and chymotrypsinogen B1, elastase 2A, 3A, 3B, trypsin 1, peptide prolyl isomerase, together with other digestive enzymes such as α-amylase 1A, 2A, 2B, carboxyl ester lipase, lipase and co-lipase and phospholipase A2. The same proteins were identified in another proteomic profiling study of pancreatic juice from individuals with chronic pancreatitis (n = 9) and a similar number of control subjects, with the addition of aminopeptidase N and a Xaa-Pro dipeptidase. Semiquantitative analysis of the protein profiles showed that digestive enzymes were downregulated two- to -threefold in patients with pancreatitis (Paulo et al., 2012).

It has classically been defined that trypsin and chymotrypsin are the most abundant human pancreatic digestive enzymes making up about 19% and 9%, respectively, of total pancreatic juice protein (Whitcomb and Lowe, 2007). One of the first studies showed the levels of trypsin and chymotrypsin drop along the length of the intestine (Borgstrom et al., 1957), presumably as a consequence of their being inactivated or digested. Determination of trypsin in intestinal fluid has shown there to be 577 (± 44) mg/L immunoreactive trypsin (n = 8) with an enzymatic activity of 11.13 (± 0.12) kU/L of activity (as measured using hydrolysis of *N*-benzoyl-L-arginine ethyl ester (Rinderknecht et al., 1978a). A mean post-prandial trypsin activity of 150 U/mL of duodenal juice was determined in 6 healthy adults (DiMagno et al., 1977). In another study, trypsin was determined in aspirates using *N*-benzoyl-L-arginine-*p*-nitroanilide as a substrate and was determined to be 143

(± 6.7) μg/mL (n = 399) using bovine pancreatic trypsin as a standard (11,000 U/mg) (Metheny et al., 1997).

Analysis of enzyme activities in pancreatic juice of normal subjects found levels of trypsin of 24.4 mg/g (with a range of 7.7–60.5, n = 20) while chymotrypsin was found at 4.8 mg/g protein (range of 1–19.4, n = 27) (Rinderknecht et al., 1978a). These authors titrated the enzymes against bovine standards using *N*-carbobenzoxy-diglycyl-L-arginyl-2-naphthylamide hydrochloride (GANA) for the trypsin activity and γ-glutamyl-*p*-nitroanilide (GGPNA) for chymotrypsin although it appeared the human enzyme was four times less active with this substrate compared to the bovine enzyme when another substrate was used (Rinderknecht et al., 1978b). Yet another study measured chymotrypsin activity in duodenal aspirates of healthy individuals and found on average 120.5 international units/ml of chymotrypsin but no clear indication is given as to the substrate used (Gaia et al., 1984). There is also data on the analysis of pancreatic tissue from human fetal, newborn and adult origin which suggest that adult pancreas contains 100 mU/mg protein trypsin while chymotrypsin was around 6 mU/mg protein; however, the methods and reporting units are not clearly described and enzymatic assays may lack specificity (Track et al., 1975). Analysis of intestinal contents in a fed state suggested that trypsin and chymotrypsin levels vary greatly between 50 and 850 μg/mL (Track et al., 1975). While enzyme activities were measured no indication is given as to how the mass of enzyme in the intestinal contents was calculated in this study.

Phosphatidylcholine is the main phospholipid found in the lumen of the proximal small intestine, with an estimated concentration of 3 mM (Carey and Hernell, 1992). In another study, the total phospholipid concentration determined in the proximal jejunum of healthy adults (n = 6) varied between 0.1 and 3.9 mM, which resulted from the mixture of endogenous secreted phospholipids and phospholipids contained in the dietary meal given to the volunteers (Persson et al., 2006). As previously described, phospholipids have also been shown to affect the rate of protein digestion in the gastric and small intestinal environments (Macierzanka et al., 2009).

### B.3.3. Concentration of bile salts in gastric and intestinal fluids

Conjugated bile salts affect the digestion of proteins by pancreatic proteases such as trypsin and chymotrypsin, having an effect at concentrations as low as 2 mM, a level below the critical micelle concentration, although it was more pronounced at 3.5 mM. Thus, bile salts should be included in model intestinal fluid. While low concentrations of bile salts (0.2 ± 0.5 mM) can be found in the human fasting gastric fluid (16 out of 36 individuals; Lindahl et al., 1997), concentrations of bile salts in the intestinal lumen are variable but usually high, estimated in the medium millimolar range (Martínez-Augustin and Sánchez de Medina, 2008). In agreement with this, the typical concentration of bile salts found in healthy adults in the fed state is 10 mM in total fluid (n = 20) (Kalantzi et al., 2006); in another study, the overall bile salt concentration found in the proximal jejunum of healthy adults (n = 6) in response to a dietary meal varied between 0.5 and 8.6 mM depending on the sampling time (Persson et al., 2006). The mean bile salt concentration for the fasted and fed state determined in pre- and full-term infants (n = 214) was calculated to be 4.7 ± 1.4 mM and 1.0 ± 0.3 mM, respectively (Kampstrum et al., 2017).

## Annex B-4. Principles of *in vitro* protein digestibility tests for risk assessment

### B.4.1. Test conditions

#### B.4.1.1. Material for testing

With regard to the test protein to be used for the revised/refined *in vitro* digestibility assays, it is considered important that the test protein comes from the parts of the plant commonly considered as edible, at present mostly the seeds/grain, because this material will be consumed.

Therefore, testing of the newly expressed proteins purified from the edible raw plant material is a more realistic scenario, if sufficient quality and quantity of the protein is available.

Should strong evidence be provided that the protein cannot be purified from GM plant material, the use of recombinant protein may be justified. Such recombinant protein should be produced in a manner which allows it to display the same structural, biochemical and functional properties, e.g. post-translational modifications (glycosylation, phosphorylation, hydroxylation, carboxylation, disulfide bond formation) as the protein expressed in a plant tissue. An expression system closely matching the

natural host should increase the structural, biochemical and functional equivalence of the recombinant protein to the newly expressed plant protein (Hoffman-Sommergruber et al., 2008).

#### B.4.1.2. Digestion conditions

Digestion conditions should be selected based on the range of conditions found *in vivo* and which encompass the needs of special groups and those receiving medication, such as antacids. The current document provides a set of conditions to reflect the different situations experienced *in vivo* (Annex B-3). Different concentrations for proteolytic enzymes and/or biosurfactants have also been described in *in vitro* digestion models (Dupont et al., 2010; Mandalari et al., 2009b; Minekus et al., 2014, Thomas et al., 2004) and summarised in Annexes B-3.2 and B-3.3.

An elaboration of the current classical pepsin resistance test would build on both *in vivo* data and current *in vitro* test methodology thus maintaining a link to previous test results and building on best practice. Different test conditions combining low and high pepsin concentrations with low and high pH values could potentially be used to represent the range of conditions found *in vivo* (Figure B.1). For simulated gastric digestion, one set of conditions that should be considered for inclusion is the current pepsin test, which represents the low pH conditions observed in the fasted state or at the end of gastric emptying with an excess of pepsin. Alternative gastric digestion conditions could be included representing the fed state, and could be followed by an intestinal *in vitro* digestion step should digestion resistant fragments be observed.

### B.4.2. Data interpretation

Criteria such as length, persistence and abundance of peptides derived from *in vitro* gastrointestinal digestion of newly expressed proteins will play a key role in identifying potential hazards within the weight-of-evidence approach for the allergenicity assessment. The kinetic parameter of 'half-life' could provide means of defining 'transient' and 'persistent' peptides generated by a given set of digestion conditions. This has been used in other assessments of the allergenic risk of novel proteins, such as the ice structuring protein (Baderschneider et al., 2002). However, further evidence to support the application of such a parameter needs to be obtained. Figure B.2 displays different example scenarios and possible subsequent data interpretation:

- if a protein digest is composed of peptides < 9 amino acid residues in length, the allergenic potential can be considered to be low;
- if a protein digest is composed of peptides ≥ 9 amino acids residues in length but transient, the allergenic potential can be assumed to be low;
- if digestion fragments of ≥ 9 amino acids or longer are identified and are persistent, further consideration is required. In such a case, the abundance of the stable peptides ≥ 9 amino acids in length within the whole mixture of digestion products should be considered in the risk assessment process. If the abundance of the stable peptides ≥ 9 amino acids in length is considered to be significant, further assessment may be requested on a case-by-case basis. This may include information on the potential of such digestion resistant fragments to interact with the immune system. Although additional tests including *in vitro* cell based assays or *in vivo* tests on animal models have not been validated so far for regulatory purposes, they may be considered useful to provide additional information (EFSA GMO Panel, 2010, 2011). For non-IgE-mediated allergic reactions (celiac disease), specific HLA-DQ-peptide modelling and binding assays, as well as T-cell testing are proposed in this document (see section 2.1 describing *Non-IgE-mediated adverse immune reactions to foods*). For IgE-mediated allergic reactions, IgE reactivity and formation of immune complexes may depend on the proximity and number of IgE epitopes (Huby et al., 2000; Gieras et al., 2016). Therefore, in contrast to T-cell epitopes, the minimum size of peptides which might act as B-cell receptor epitopes and cause IgE cross-linking is less clear, and will require the presence of multiple epitopes (at least two) which can only be accommodated in peptides greater than 9 amino acids in length (Harwood et al., 2010; Handlogten et al., 2013).

Finally, an appropriate exposure assessment, on a case-by-case basis, is envisaged to allow a more reliable and accurate allergenicity assessment.

## Annex B-5. Interim phase for refining the conditions of *in vitro* digestion tests

*General considerations:*

During the interim phase, two aspects of the *in vitro* digestion test should be considered:

1) Test conditions, the interplay between pH, enzyme concentration and duration of the digestion are key aspects to be considered taking into account data on *in vivo* conditions.
2) End-points of digestion and objective measurements of the extent of digestion should be developed which can enable better comparison of test results. This includes (at least) semiquantitative assessment of the abundance of peptide fragments in a digest and a definition of transient and persistent digestion fragments, using concepts such as half-life. The identification of persistent peptide fragments which are $\geq$ 9 amino acids in length is critical, because these may indicate that further assessment is required.

*Specific and key considerations:*

A) Digestion model conditions: Because the physiological process of digestion is inherently dynamic, *in vitro* batch models of digestion inevitably restrict the range of conditions that exist *in vivo*. To better reflect the range of conditions found *in vivo*, without overly increasing the complexity of the test, it is proposed that a minimum of two gastrointestinal test conditions should be used to reflect the range found *in vivo*. The classical pepsin resistance test should form one part of this testing scenario, having the added value of enabling comparison with past data collected on resistance of newly expressed proteins to digestion. (a) pH conditions: The pH employed in the *in vitro* digestibility tests range from 1.2 to 4.2 with many using pH conditions of 1.2 or 2.0, the former based on the simulated gastric fluid conditions detailed in the US Pharmacopeia (1995) and used for drug dissolution tests. However, intragastric pH conditions found *in vivo* tend only to go down to around pH 2 towards the end of gastric emptying, and are generally much higher because of the buffering capacity of many foods (Kalantzi et al., 2006) (Annex B-3.1). (b) Pepsin:protein ratio: Many reports only describe it on a (w:w) basis and do not take differences in enzyme activity into account. However, the two available ring-trialled studies of digestibility tests for purified proteins used a ratio of 10 U (Thomas et al., 2004) and 0.165 U (Mandalari et al., 2009b) of pepsin per μg of tested protein, respectively. (c) Durations of digestion: Different durations of digestion based on digestion of whole foods have been proposed (Minekus et al., 2014). Test material is typically exposed to gastric digestion for 60 min followed by intestinal digestion for 30–60 min with corresponding intermediate sampling time points (Mandalari et al., 2009b; Dupont et al., 2010). Adaptation and integration of these approaches may provide the conditions for a refined *in vitro* digestion test. These adaptations should take *in vivo* physiological conditions, described in Annex B-3, into account.

Although a flexible framework for performing the digestion tests is considered in line with the current considerations, the following recommendations should ensure consistency between the different laboratories undertaking the tests:

i) The source, purity and specific activities of digestive enzymes (e.g. pepsin and intestinal endoproteases) should be determined using standardised protocols. Individual intestinal enzymes (i.e. trypsin and chymotrypsin) should be prioritised over the use of pancreatin. The use of individual duodenal enzymes instead of pancreatin should provide a more consistent, accurate and detailed analysis of protein hydrolysis. In the case of using pancreatin, proteolytic, lipolytic and amylolytic activity of the extract should be determined and the amount of pancreatin added should be based on the trypsin activity. Finally and under specific circumstances, the use of other intestinal lumen proteases such as elastase and carboxypeptidases and additional brush border enzymes as aminopeptidases could also be considered. Further considerations in such respect should take place with the laboratory(ies) involved in the interim phase.
ii) Enzyme concentrations and activity used in digestion tests should be specified and could be based on approaches previously used in interlaboratory studies performed with purified proteins (Thomas et al., 2004; Mandalari et al., 2009b) taking into consideration data on physiological levels reported in human studies (Annex B-3).
iii) Different gastric pH values should be specified which reflect those found *in vivo* and encompass those found in infants, adults, elderly or people with impaired digestive functions.

iv) The addition of biosurfactants (bile salts and phospholipids) at physiologically relevant levels should be considered. The use of individual and well-defined lipids and bile salt standards instead of extracts could provide more reproducible and standardised conditions unless the high quality of the extracts can be guaranteed and assured the non-variability of their composition among different lots.

It is proposed that, at a minimum, two gastrointestinal digestion test conditions should be considered. These conditions should encompass the most extensive (such as the current classical pepsin resistance test conditions) and the less extensive digestion conditions, reflective of those found in fed state in adults and children.

1) Low pH/high [pepsin], these conditions should include those used in the classical pepsin resistance test (Thomas et al., 2004). In the case that persistent peptides ≥ 9 amino acids residues in length are detected at the end of the digestion phase, a subsequent intestinal digestion should be performed (Figure B.1). Based on different physiological conditions described in Annex B-3, these set of conditions could simulate an adult in fasted state. As previously mentioned, the interplay between pH, enzyme concentration and duration of the digestion is a key aspect to be considered, together with the inclusion of gastric biosurfactants.
2) High pH/low [pepsin] followed by intestinal conditions, if persistent peptides ≥ 9 amino acids residues in length are detected at the end of the digestion phase. In the gastric phase physiological conditions reported in *in vivo* studies and to be considered are the following: a pH of 5.5 and a pepsin concentration of 1,000 U/mL of gastric juice (when using haemoglobin as a substrate – see Annex B-3) (Figure B.1). These conditions would reflect those found in the fed state. Again, an important aspect to consider further is the interplay between pH, enzyme concentration and duration of the digestion, together with the inclusion of biosurfactants.

In relation to the intestinal phase, a pH of 6.5 could be considered as a physiologically relevant value which has been reported in the proximal intestinal tract. The evidence available regarding the concentration and activity of intestinal enzymes observed *in vivo* is much weaker than for pepsin concentrations in gastric fluid. For trypsin, it is proposed that a reference value of 145 μg/mL of duodenal juice, equating to 1,595 U/mL, identified *in vivo* (when using *N*-benzoyl-L-arginine-*p*-nitroanilide as a substrate and bovine pancreatic trypsin with 11,000 U/mg as a standard – see Annex B-3). In relation to chymotrypsin, the data are even more sparse. One approach to defining a level would be to have a ratio of trypsin:chymotrypsin by mass which relates to the ratio observed *in vivo* of 2:1 (w/w). In line with previous deliberations, the inclusion of intestinal biosurfactants should be considered. Finally and given the complexity of pancreatic secretions, the use of other endo- and exopeptidases should also be considered.

Replication of the *in vitro* digestion experiments should also be implemented to obtain more reliable and statistically significant results.

B) Control proteins: Digestion studies should be performed including control proteins to demonstrate the effectiveness of the digestion system employed and allow benchmarking of different digestion models. The control proteins are not to provide an indication of allergenicity, but rather reflect the different susceptibilities of proteins to gastrointestinal digestion.

For proteins to act as appropriate controls in the *in vitro* digestion tests, they must be:

1) either commercially available, and/or purified in reasonable quantities using published methods and made available for use by the community (e.g. either as quality control (QC) or reference materials);
2) well characterised with regard to their primary sequence, post-translational modifications (if any) and physicochemical state (e.g. size, oligomerisation, pI, hydrophobicity, ligand binding/ prosthetic group);
3) previously subjected to *in vitro* digestion tests allowing their susceptibility to digestion to be classified as either highly resistant, moderately resistant or labile to the action of digestive enzymes.

One protein which meets these criteria, has been extensively tested in digestion tests and was found to be extremely resistant to gastric digestion is bovine β-lactoglobulin (Reddy et al., 1988;

Schmidt et al., 1995; Astwood et al., 1996; Yagami et al., 2000; Fu et al., 2002; Takagi et al., 2003; Thomas et al., 2004; Sanz et al., 2007; Herman et al., 2007; Lucas et al., 2008; Ofori-Anti et al., 2008; Macierzanka et al., 2009; Mandalari et al., 2009a,b; Misra et al., 2009; Dupont et al., 2010; Zheng et al., 2010; Bogh et al., 2013). β-Lactoglobulin has also been reported to be relatively stable under gastrointestinal digestion conditions (Mandalari et al., 2009a,b; Dupont et al., 2010; Borgh et al., 2013). On the other hand, phosphofructokinase and/or sucrose synthetase could also be recommended as control proteins because both enzymes have been described to be rapidly digested under simulated gastric (Astwood et al., 1996; Fu et al., 2002) and intestinal conditions (Fu et al., 2002). A representative gluten protein could also be used as control protein.

C) Digestion end-points and read-out considerations: The terms 'persistent' and 'transient' are used for classification of proteins and peptides with different kinetic behaviour and for fragments in relation to their rate of formation as well as their rate of further degradation. Kinetic studies are founded on following the time course of a reaction or process, and hence, there is a necessity for the digestion tests to take the form of time-course experiments. Consequently, sampling should be undertaken at various time points during the gastric and intestinal digestion steps, to allow the evolution of peptide fragments to be monitored. Sampling time points should be selected which are appropriate and will allow transient and persistent peptides to be distinguished based on kinetic parameters.

Standardised methodology for monitoring protein digestion needs to be used which is suitable for profiling of both large resistant fragments and lower molecular weight peptides of ~ 1,000 Da (the average mass of a 9 residue peptide fragment). Techniques should also allow at least semiquantitative profiling of residual intact protein and digestion products. Traditionally, protein digestibility has been measured using SDS-PAGE. However, while providing valuable data especially for intact proteins and large resistant fragments, this technique is essentially qualitative in nature, can provide inconsistent results between laboratories and is not an appropriate technique to carry out reliable quantification of peptide fragments. Tandem mass spectrometry, even with caveats with regard to peptide ionisation efficiency, is a more effective tool to carry out a comprehensive peptide mapping of digesta and identify stable digestion fragments ≥ 9 amino acids in length. Since at present no single methodology can readily characterise the digestion of both proteins and peptides effectively, a combination of the best available methodology, such as SDS-PAGE and mass spectrometry, should be used. These techniques can provide at least a semi-quantitative output of digestion following disappearance of the intact protein and appearance of digestion resistant fragments.

Replicate analysis should be undertaken using two biological replicates (i.e. digestion tests) being analysed in duplicate, which is considered the minimum. Best laboratory practice should be addressed for any methodology such as use of standards, use of appropriate protein stains able to provide a broad dynamic range and stain many proteins, molecular weight markers (SDS-PAGE) or internal peptide standards spanning a range of masses (mass spectrometry).

Special attention should be paid to the pretreatment of the digesta before analysis (e.g. analyses need to take account of maintaining the integrity of disulfide bonds during sample handling and analysis to remain as close as possible to physiological conditions). Hence, the use of reductants or derivatisation of the sulfhydryl groups of cysteine residues should be avoided if possible. It is necessary to not only take the presence of covalent linkages (such as disulfide bridges) into account, but also non-covalent interactions, which may lead to smaller peptides assembling to complexes of higher molecular weight and size.

D) Classification of digestion resistant proteins/peptides: A consensus definition of transient and persistent proteins and peptides is required that can apply to the different *in vitro* digestion test conditions. Control proteins accepted to be highly resistant, moderately resistant and highly digestible could be used to support development of such a definition. To do that, an approach based on the objective measurement of the extent of digestion could be explored. To this end, the half-life of the intact protein and resulting peptide fragments could be determined and used to establish definitions of transient and resistant peptides (Shan et al., 2002; Baderschneider et al., 2002, Herman et al., 2007; Ofori-Anti et al., 2008; Macierzanka et al., 2009; Defernez et al., 2010; Yao et al., 2013; Smith et al., 2015).

## References

Akimov M and Bezuglov V, 2012. Methods of protein digestive stability assay - state of the art. New Advances in the Basic and Clinical Gastroenterology, Prof. Tomasz Brzozowski (ed.), InTech.

Armand M, Hamosh M, DiPalma JS, Gallagher J, Benjamin SB, Philpott JR, Lairon D and Hamosh P, 1995. Dietary fat modulates gastric lipase activity in healthy humans. The American Journal of Clinical Nutrition, 62, 74–80.

Armand M, Hamosh M, Mehta NR, Angelus PA, Philpott JR, Henderson TR, Dwyer NK, Lairon D and Hamosh P, 1996. Effect of human milk or formula on gastric function and fat digestion in the premature infant. Pediatr Research, 40, 429–437.

Astwood JD, Leach JN and Fuchs RL, 1996. Stability of food allergens to digestion *in vitro*. Nature Biotechnology, 14, 1269–1273.

Baderschneider B, Crevel RW, Earl LK, Lalljie A, Sanders DJ and Sanders IJ, 2002. Sequence analysis and resistance to pepsin hydrolysis as part of an assessment of the potential allergenicity of ice structuring protein type II HPLC 12. Food and Chemical Toxicology, 40, 965–978.

Ballmer-Weber BK, Hoffmann A, Wüthrich B, Lüttkopf D, Pompei C, Wangorsch A, Kästner M and Vieths S, 2002. Influence of food processing on the allergenicity of celery: DBPCFC with celery spice and cooked celery in patients with celery allergy. Allergy, 57, 228–235.

Banerjee R, Reddy DN, Guda NM, Kalpala R, Mahurkar S, Darisetty S and Rao GV, 2010. Oral buffered esomeprazole is superior to i.v. pantoprazole for rapid rise of intragastric pH: a wireless pH metry analysis. Journal of Gastroenterology Hepatology, 25, 43–47.

Bartnikas LM, Sheehan WJ, Hoffman EB, Permaul P, Dioun AF, Friedlander J, Baxi SN, Schneider LC and Phipatanakul W, 2012. Predicting food challenge outcomes for baked milk: role of specific IgE and skin prick testing. The Annals of Allergy, Asthma, and Immunology, 109, 309–313.

Blanquet S, Marol-Bonnin S, Beyssac E, Pompon D, Renaud M and Alric M, 2001. The 'biodrug' concept: an innovative approach to therapy. Trends Biotechnology, 19, 393–400.

Bogh KL, Barkholt V, Rigby NM, Mills EN and Madsen CB, 2012. Digested Ara h 1 loses sensitizing capacity when separated into fractions. Journal of Agricultural and Food Chemistry, 60, 2934–2942.

Bogh KL, Barkholt V and Madsen CB, 2013. The sensitising capacity of intact β-lactoglobulin is reduced by co-administration with digested β-lactoglobulin. International Archives of Allergy and Immunology, 161, 21–36.

Borgstrom B, Dahlqvist A, Lundh G and Sjovall J, 1957. Studies of intestinal digestion and absorption in the human. Journal of Clinical Investigation, 36, 1521–1536.

Bourlieu C, Menard O, Bouzerzour K, Mandalari G, Macierzanka A, Mackie AR and Dupont D, 2014. Specificity of infant digestive conditions: some clues for developing relevant *in vitro* models. Critical Reviews in Food Science and Nutrition, 54, 1427–1457.

Bratten J and Jones MP, 2009. Prolonged recording of duodenal acid exposure in patients with functional dyspepsia and controls using a radiotelemetry pH monitoring system. Journal of Clinical Gastroenterology, 43, 527–533.

Brenna O, Pompei C, Ortolani C, Pravettoni V, Farioli L and Pastorello EA, 2000. Technological processes to decrease the allergenicity of peach juice and nectar. J Agric Food Chem 48, 493–497.

Brockow K, Kneissl D, Valentini L, Zelger O, Grosber M, Kugler C, Werich M, Darsow U, Matsuo H, Morita E and Ring J, 2015. Using a gluten oral food challenge protocol to improve diagnosis of wheat-dependent exercise-induced anaphylaxis. J Allergy Clin Immunol 135, 977–984.

Carey MC and Hernell O, 1992. Digestion and absorption of fat. Semin Gastrointest Dis 3, 189–208.

Chambers SJ, Wickham MS, Regoli M, Bertelli E, Gunning PA and Nicoletti C, 2004. Rapid *in vivo* transport of proteins from digested allergen across pre-sensitized gut. Biochem Biophys Res Commun 325, 1258–1263.

Chang AB, Hills YC, Cox NC, Cleghorn GJ, Valery PC, Lewindon PJ, Ee LC, Withers GD, Beem C and Hills BA, 2006. 'Free' surfactant in gastric aspirates and bronchoalveolar lavage in children with and without reflux oesophagitis. Intern Med J 36, 226–230.

Chicz RM, Urban RG, Gorga JC, Vignali DAA, Lane WS and Strominger JL, 1993. Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. J Exp Med 178, 27–47.

Chinthrajah RS, Hernandez JD, Boyd SD, Galli SJ and Nadeau KC, 2016. Molecular and cellular mechanisms of food allergy and food tolerance. J Allergy Clin Immunol 137, 984–997.

Codex Alimentarius, 2003. Foods Derived from Modern Biotechnology. Codex Alimentarius Commission, Joint FAO/WHO Food Standards Programme, Rome.

Codex Alimentarius, 2009. Foods Derived from Modern Biotechnology. Codex Alimentarius Commission, Joint FAO/WHO Food Standards Programme, Rome.

Defernez M, Mandalari G and Mills ENC, 2010. Quantitative assessment of multi-laboratory reproducibility of SDS-PAGE assays: Digestion pattern of beta-casein and beta-lactoglobulin under simulated conditions. Electrophoresis 31, 2838–2848.

Díaz-Perales A, Blanco C, Sánchez-Monge R, Varela J, Carrillo T and Salcedo G, 2003. Analysis of avocado allergen (Prs a 1) IgE-binding peptides generated by simulated gastric fluid digestion. J Allergy Clin Immunol 112, 1002–1007.

Diesner SC, Knittelfelder R, Krishnamurthy D, Pali-Scholl I, Gajdzik L, Jensen-Jarolim E and Untersmayr E, 2008. Dose-dependent food allergy induction against ovalbumin under acid-suppression: a murine food allergy model. Immunol Lett 121, 45–51.

DiMagno EP, Malagelada JR, Go VL and Moertel CG, 1977. Fate of orally ingested enzymes in pancreatic insufficiency. Comparison of two dosage schedules. N Engl J Med 296, 1318–1322.

Doyle CJ, Yancey K, Pitt HA, Wang M, Bemis K, Yip-Schneider MT, Sherman ST, Lillemoe KD, Goggins MD and Schmidt CM, 2012. The proteome of normal pancreatic juice. Pancreas 41, 186–194.

Dupont D, Mandalari G, Mollé D, Jardin J, Rolet-Répécaud O, Duboz G, Léonil J, Mills ENC and Mackie AR, 2010. Food processing increases casein resistance to simulated infant digestion. Mol Nutr Food Res 54, 1677–1689.

Egger L, Ménard O, Delgado-Andrade C, Alvito P, Assunção R, Balance S, Barberá R, Brodkorb A, Cattenoz T, Clemente A, Comi I, Dupont D, Garcia-Llatas G, Lagarda MJ, Le Feunteun S, JanssenDuijghuijsen L, Karakaya S, Lesmes U, Mackie AR, Martins C, Meynier A, Miralles B, Murray BS, Pihlanto A, Picariello G, Santos CN, Simsek S, Recio I, Rigby N, Rioux LE, Stoffers H, Tavares A, Tavares L, Turgeon S, Ulleberg EK, Vegarud GE, Vergères G and Portmann R, 2016. The harmonized INFOGEST *in vitro* digestion method: From knowledge to action. Food Res Int 88, 217–225.

Evans DF, Pye G, Bramley R, Clark AG, Dyson TJ and Hardcastle JD, 1988. Measurement of gastrointestinal pH profiles in normal ambulant human subjects. Gut 29, 1035–1041.

Fallingborg J, Christensen LA, Ingeman-Nielsen M, Jacobsen BA, Abildgaard K and Rasmussen HH, 1989. pH-profile and regional transit times of the normal gut measured by a radiotelemetry device. Aliment Pharmacol Ther 3, 605–613.

Faure C, Michaud L, Shaghaghi EK, Popon M, Turck D, Navarro J and Jacqz-Aigrain E, 2001. Intravenous omeprazole in children: pharmacokinetics and effect on 24-hour intragastric pH. J Pediatr Gastroenterol Nutr 33, 144–148.

Foster ES, Kimber I and Dearman RJ, 2013. Relationship between protein digestibility and allergenicity: comparisons of pepsin and cathepsin. Toxicology 308, 30–38.

Fu T-J, Abbott UR and Hatzos C, 2002. Digestibility of food allergens and nonallergenic proteins in simulated gastric fluid and simulated intestinal fluid. A comparative study. J Agric Food Chem 50, 7154–7160.

Gaia E, Andriulli A, Tappero G, Piantino P, Jayme A, Rocca G and Musorrofiti A, 1984. Influence of sex and body size on pancreatic bicarbonate and enzyme output. J Clin Gastroenterol 6, 429–433.

Gieras A, Linhart B, Roux KH, Dutta M, Khodoun M, Zafred D, Cabauatan CR, Lupinek C, Weber M, Focke-Tejkl M, Keller W, Finkelman FD and Valenta R, 2016. IgE epitope proximity determines immune complex shape and effector cell activation capacity. J Allergy Clin Immunol 137, 1557–1565.

Grimshaw KE, King RM, Nordlee JA, Hefle SL, Warner JO and Hourihane JO, 2003. Presentation of allergen in different food preparations affects the nature of the allergic reaction–a case series. Clin Exp Allergy 33, 1581–1585.

Handlogten MW, Kiziltepe T, Serezani AP, Kaplan MH and Bilgicer B, 2013. Inhibition of weak-affinity epitope-IgE interactions prevents mast cell degranulation. Nat Chem Biol 9, 789–795.

Herman RA, Woolhiser MM, Ladics GS, Korjagin VA, Schafer BW, Storer NP, Green SB and Kan L, 2007. Stability of a set of allergens and non-allergens in simulated gastric fluid. Int J Food Sci Nutr 58, 125–141.

Hoffmann-Sommergruber K, Mills ENC and Vieths S, 2008. Coordinated and standardized production, purification and characterization of natural and recombinant food allergens to establish a food allergen library. Mol Nutr Food Res 52, S159–S165.

Holst B and Williamson G, 2008. Nutrients and phytochemicals: from bioavailability to bioefficacy beyond antioxidants. Curr Opin Biotechnol 19, 73–82.

Hosking DJ, Moody F, Stewart IM and Atkinson M, 1975. Vagal impairment of gastric secretion in diabetic autonomic neuropathy. Br Med J 2, 588–590.

Huby RD, Dearman RJ and Kimber I, 2000. Why are some proteins allergens? Toxicol Sci 55, 235–246.

Kalantzi L, Goumas K, Kalioras V, Abrahamsson B, Dressman JB and Reppas C, 2006. Characterization of the human upper gastrointestinal contents under conditions simulating bioavailability/bioequivalence studies. Pharm Res 23, 165–176.

Kamstrup D, Berthelsen R, Sassene PJ, Selen A and Mullertz A, 2017. *In vitro* model simulating gastro-intestinal digestion in the pediatric population (neonates and young infants). AAPS PharmSciTech. 18, 317–329.

Kay L and Jorgensen T. Epidemiology of upper dyspepsia in a random population (1994) Prevalence, incidence, natural history, and risk factors. Scand J Gastroenterol 29, 2–6.

Keller J and Layer P, 2005. Human pancreatic exocrine response to nutrients in health and disease. Gut 54, 1–28.

Koziolek M, Grimm M, Becker D, Iordanov V, Zou H, Shimizu J, Wanke C, Garbacz G and Weitschies W, 2015. Investigation of pH and temperature profiles in the gi tract of fasted human subjects using the Intellicap(®) system. J Pharm Sci 104, 2855–2863.

Levi CS and Lesmes U, 2014. Bi-compartmental elderly or adult dynamic digestion models applied to interrogate protein digestibility. Food Funct 5, 2402–2409.

Lindahl A, Ungell AL, Knutson L and Lennernäs H, 1997. Characterization of fluids from the stomach and proximal jejunum in men and women. Pharm Res 14, 497–502.

Lucas JS, Cochrane SA, Warner JO and Hourihane JO, 2008. The effect of digestion and pH on the allergenicity of kiwifruit proteins. Pediatr Allergy Immunol 19, 392–398.

Macierzanka A, Sancho AI, Mills ENC, Rigby NM and Mackie AR, 2009. Emulsification alters simulated gastrointestinal proteolysis of β-casein and β-lactoglobulin. Soft Matter 5, 538–550.

Mackie A, Knulst A, Le TM, Bures P, Salt L, Mills EN, Malcolm P, Andreou A and Ballmer-Weber BK, 2012. High fat food increases gastric residence and thus thresholds for objective symptoms in allergic patients. Mol Nutr Food Res 56, 1708–1714.

Mandalari G, Mackie AR, Rigby NM, Wickham MSJ and Mills ENC, 2009a. Physiological phosphatidylcholine protects bovine b-lactoglobulin from simulated gastrointestinal proteolysis. Mol Nutr Food Res 53, S131–S139.

Mandalari G, Adel-Patient K, Barkholt V, Baro C, Bennett L, Bublin M, Gaier S, Graser G, Ladics GS, Mierzejewska D, Vassilopoulou E, Vissers YM, Zuidmeer L, Rigby NM, Salt LJ, Defernez M, Mulholland F, Mackie AR, Wickham MS and Mills ENC, 2009b. In vitro digestibility of β-casein and β -lactoglobulin under simulated human gastric and duodenal conditions: a multi-laboratory evaluation. Regul Toxicol Pharmacol 55, 372–381.

Martínez-Augustin O and Sánchez de Medina F, 2008. Intestinal bile acid physiology and pathophysiology. World J Gastroenterol 14, 5630–5640.

Mason S, 1962. Some aspects of gastric function in the newborn. Arch Dis Child 37, 387–391.

Matsuo H, Morimoto K, Akaki T, Kaneko S, Kusatake K, Kuroda T, Niihara H, Hide M and Morita, 2005. Exercise and aspirin increase levels of circulating gliadin peptides in patients with wheat-dependent exercise-induced anaphylaxis. Clin Exp Allergy 35, 461–466.

Ménard O, Cattenoz T, Guillemin H, Souchon I, Deglaire A, Dupont D and Picque D, 2014. Validation of a new in vitro dynamic system to simulate infant digestion. Food Chem 145, 1039–1045.

Metheny NA, Stewart BJ, Smith L, Yan H, Diebold M and Clouse RE, 1997. pH and concentrations of pepsin and trypsin in feeding tube aspirates as predictors of tube placement. J Parenter Enteral Nutr 21, 279–285.

Mercuri A, Passalacqua A, Wickham MS, Faulks RM, Craig DQ and Barker SA, 2011. The effect of composition and gastric conditions on the self-emulsification process of ibuprofen-loaded self-emulsifying drug delivery systems: a microscopic and dynamic gastric model study. Pharm Res 28, 1540–1551.

Metcalfe DD, Astwood JD, Townsend R, Sampson HS, Taylor SL and Fuchs RL, 1996. Assessment of the allergenic potential of foods derived from genetically engineered crop plants. Crit Rev Food Sci Nutr 36, S165–S186.

Michalek W, Semler JR and Kuo B, 2011. Impact of acid suppression on upper gastrointestinal pH and motility. Dig Dis Sci 56, 1735–1742.

Minekus M, Alminger M, Alvito P, Ballance S, Bohn T, Bourlieu C, Carrière F, Boutrou R, Corredig M, Dupont D, Dufour C, Egger L, Golding M, Karakaya S, Kirkhus B, Le Feunteun S, Lesmes U, Macierzanka A, Mackie A, Marze S, McClements DJ, Ménard O, Recio I, Santos CN, Singh RP, Vegarud GE, Wickham MS, Weitschies W and Brodkorb A, 2014. A standardised static in vitro digestion method suitable for food - an international consensus. Food Funct 5, 1113–1124.

Misra A, Prasad R, Das M and Dwivedi PD, 2009. Probing novel allergenic proteins of commonly consumed legumes. Immunopharmacol Immunotoxicol 31, 186–194.

Mohan JF and Unanue ER, 2012. Unconventional recognition of peptides by T cells and the implications for autoimmunity. Nat Rev Immunol 12, 721–728.

Moreno FJ, Mackie AR and Mills ENC, 2005a. Phospholipid interactions protect the milk allergen alpha-lactalbumin from proteolysis during in vitro digestion. J Agric Food Chem 53, 9810–9816.

Moreno FJ, Mellon FA, Wickham MS, Bottrill AR and Mills ENC, 2005b. Stability of the major allergen Brazil nut 2S albumin (Ber e 1) to physiologically relevant in vitro gastrointestinal digestion. FEBS J 272, 341–352.

Moreno FJ, 2007. Gastrointestinal digestion of food allergens: effect on their allergenicity. Biomed Pharmacother 61, 50–60.

Netting M, Makrides M, Gold M, Quinn P and Penttila I, 2013. Heated allergens and induction of tolerance in food allergic children. Nutrients 5, 2028–2046.

Ofori-Anti AO, Ariyarathna H, Chen L, Lee HL, Pramod SN and Goodman RE, 2008. Establishing objective detection limits for the pepsin digestion assay used in the assessment of genetically modified foods. Regul Toxicol Pharmacol 52, 94–103.

Omari TI and Davidson GP, 2003. Multipoint measurement of intragastric pH in healthy preterm infants. Arch Dis Child Fetal Neonatal 88, F517–F520.

Omari T, Davidson G, Bondarov P, Naucler E, Nilsson C and Lundborg P, 2007. Pharmacokinetics and acid-suppressive effects of esomeprazole in infants 1–24 months old with symptoms of gastroesophageal reflux disease. J Pediatr Gastroenterol Nutr 45, 530–537.

Ono S, Kato M, Ono Y, Imai A, Yoshida T, Shimizu Y and Asaka M, 2009. Immediate acid-suppressing effects of ranitidine hydrochloride and rabeprazole sodium following initial administration and reintroduction: A randomized, cross-over study using wireless pH monitoring capsules. J Gastroenterol Hepatol 24, 639–645.

Pabst O and Mowat AM, 2012. Oral tolerance to food protein. Mucosal Immunol 5, 232–239.

Paulo JA, Lee LS, Wu B, Repas K, Banks PA, Conwell DL and Steen H, 2010. Proteomic analysis of endoscopically (endoscopic pancreatic function test) collected gastroduodenal fluid using in-gel tryptic digestion followed by LC-MS/MS. Proteomics Clin Appl 4, 715–725.

Paulo JA, Kadiyala V, Lee LS, Banks PA, Conwell DL and Steen H, 2012. Proteomic analysis (GeLC-MS/MS) of ePFT-collected pancreatic fluid in chronic pancreatitis. J Proteome Res 11, 1897–1912.

Persson EM, Nilsson RG, Hansson GI, Löfgren LJ, Libäck F, Knutson L, Abrahamsson B and Lennernäs H, 2006. A clinical single-pass perfusion investigation of the dynamic in vivo secretory response to a dietary meal in human proximal small intestine. Pharm Res 23, 742–751.

Reddy IM, Kella NKD and Kinsella JE, 1988. Structural and conformational basis of the resistance of beta-lactoglobulin to peptic and chymotrypsin digestion. J Agric Food Chem 36, 737–741.
Rémond D, Shahar DR, Gille D, Pinto P, Kachal J, Peyron MA, Dos Santos CN, Walther B, Bordoni A, Dupont D, Tomás-Cobos L and Vergères G, 2015. Understanding the gastrointestinal tract of the elderly to develop dietary solutions that prevent malnutrition. Oncotarget 6, 13858–13898.
Rinderknecht H, Nagaraja MR and Adham NF, 1978a. Effect of bile acids and pH on the release of enteropeptidase in man. Am J Dig Dis 23, 332–336.
Rinderknecht H, Nagaraja MR and Adham NF, 1978b. Enteropeptidase levels in duodenal juice of normal subjects and patients with gastrointestinal disease. Am J Dig Dis 23, 327–331.
Robic S, Linscott KB, Aseem M, Humphreys EA and McCartha SR, 2011. Bile acids as modulators of enzyme activity and stability. Protein J 30, 539–545.
Rudensky AY, Preston-Hurlburt P, Hong S-C, Barlow A and Janeway CA Jr, 1991. Sequence analysis of peptides bound to MHC class II molecules. Nature 353, 622–627.
Sanz ML, Corzo-Martinez M, Rastall RA, Olano A and Moreno FJ, 2007. Characterization and in vitro digestibility of bovine β-lactoglobulin glycated with galactooligosaccharides. J Agric Food Chem 55, 7916–7925.
Schmidt DG, Meijer RJ, Slangen CJ and van Beresteijn EC, 1995. Raising the pH of the pepsin-catalysed hydrolysis of bovine whey proteins increases the antigenicity of the hydrolysates. Clin Exp Allergy 25, 1007–1017.
Scholl L, Untersmayr E, Bakos N, Roth-Walter F, Gleiss A, Boltz-Nitulescu G, Scheiner O and Jensen-Jarolim E, 2005. Antiulcer drugs promote oral sensitization and hypersensitivity to hazelnut allergens in BALB/c mice and humans. Am J Clin Nutr 81, 154–160.
Shakeri-Leidenmuhler S, Lukschal A, Schultz C, Bohdjalian A, Langer F, Birsan T, Diesner SC, Greisenegger EK, Scheiner O, Kopp T, Jensen-Jarolim E, Prager G and Untersmayr E, 2015. Surgical Elimination of the Gastric Digestion by Roux-en-Y Gastric Bypass Impacts on Food Sensitisation-a Pilot Study. Obes Surg 25, 2268–2275.
Shan L, Molberg O, Parrot I, Hausch F, Filiz F, Gray GM, Sollid LM and Khosla C, 2002. Structural basis for gluten intolerance in celiac sprue. Science 297, 2275–2279.
Smith F, Pan X, Bellido V, Toole GA, Gates FK, Wickham MS, Shewry PR, Bakalis S, Padfield P and Mills ENC, 2015. Digestibility of gluten proteins is reduced by baking and enhanced by starch digestion. Mol Nutr Food Res 59, 2034–2043.
Schmitz MG and Renooij W, 1990. Phospholipids from rat, human, and canine gastric mucosa. Composition and metabolism of molecular classes of phosphatidylcholine. Gastroenterol 99, 1292–1296.
Strait RT, Mahler A, Hogan S, Khodoun M, Shibuya A and Finkelman FD, 2011. Ingested allergens must be absorbed systemically to induce systemic anaphylaxis. J Allergy Clin Immunol 127, 982–989.
Takagi K, Teshima R, Okunuki H and Sawada J, 2003. Comparative study of in vitro digestibility of food proteins and effect of preheating on the digestion. Biol Pharm Bull 26, 969–973.
Thomas K, Aalbers M, Bannon GA, Bartels M, Dearman RJ, Esdaile DJ, Fu TJ, Glatt CM, Hadfield N, Hatzos C, Hefle SL, Heylings JR, Goodman RE, Henry B, Herouet C, Holsapple M, Ladics GS, Landry TD, MacIntosh SC, Rice EA, Privalle LS, Steiner HY, Teshima R, Van Ree R, Woolhiser M and Zawodny J, 2004. A multi-laboratory evaluation of a common in vitro pepsin digestion assay protocol used in assessing the safety of novel proteins. Regulatory Toxicology and Pharmacology, 39, 87–98.
Track NS, Creutzfeldt C and Bokermann M, 1975. Enzymatic, functional and ultrastructural development of the exocrine pancreas–II. The human pancreas. Comparative Biochemistry and Physiology A Comp Physiol, 51, 95–100.
Untersmayr E, Scholl I, Swoboda I, Beil WJ, Forster-Waldl E, Walter F, Riemer A, Kraml G, Kinaciyan T, Spitzauer S, Boltz-Nitulescu G, Scheiner O and Jensen-Jarolim E, 2003. Antacid medication inhibits digestion of dietary proteins and causes food allergy: A fish allergy model in Balb/c mice. The Journal of Allergy and Clinical Immunology, 112, 616–623.
Untersmayr E, Bakos N, Scholl I, Kundi M, Roth-Walter F, Szalai K, Riemer AB, Ankersmit HJ, Scheiner O, Boltz-Nitulescu G and Jensen-Jarolim E, 2005. Anti-ulcer drugs promote IgE formation toward dietary antigens in adult patients. FASEB Journal, 19, 656–658.
Untersmayr E, Vestergaard H, Malling HJ, Jensen LB, Platzer MH, Boltz-Nitulescu G, Scheiner O, Skov PS, Jensen-Jarolim E and Poulsen LK, 2007. Incomplete digestion of codfish represents a risk factor for anaphylaxis in patients with allergy. The Journal of Allergy and Clinical Immunology, 119, 711–717.
Untersmayr E, Diesner SC, Bramswig KH, Knittelfelder R, Bakos N, Gundacker C, Lukschal A, Wallmann J, Szalai K, Pali-Scholl I, Boltz-Nitulescu G, Scheiner O, Duschl A and Jensen-Jarolim E, 2008. Characterization of intrinsic and extrinsic risk factors for celery allergy in immunosenescence. Mechanisms of Ageing and Development, 129, 120–128.
USP, 1995. Simulated gastric fluid and simulated intestinal fluid. In: The United States Pharmacopeia 23, The National Formulary 18. Rockville, MD: United States Pharmacopeial Convention, Inc.; 1995. 2053 pp.
Wenner J, Gunnarsson T, Graffner H and Lindell G, 2000. Influence of smoking and Helicobacter pylori on gastric phospholipids. Digestive Diseases and Sciences, 45, 1648–1652.
Whitcomb DC and Lowe ME, 2007. Human pancreatic digestive enzymes. Digestive Diseases and Sciences, 52, 1–17.

Worm M, Hompes S, Fiedler EM, Illner AK, Zuberbier T and Vieths S, 2009. Impact of native, heat-processed and encapsulated hazelnuts on the allergic response in hazelnut-allergic patients. Clinical and Experimental Allergy, 39, 159–166.

Yagami T, Haishima Y, Nakamura A, Osuna H and Ikezawa Z, 2000. Digestibility of allergens extracted from natural rubber latex and vegetable foods. The Journal of Allergy and clinical Immunology, 106, 752–762.

Yao X, Bunt C, Cornish J, Quek SY and Wen J, 2013. Improved RP-HPLC method for determination of bovine lactoferrin and its proteolytic degradation in simulated gastrointestinal fluids. Biomedical Chromatography, 27, 197–202.

Zheng C, Liu Y, Zhou Q and Di X, 2010. Capillary electrophoresis with noncovalently bilayer-coated capillaries for stability study of allergenic proteins in simulated gastrointestinal fluids. Journal of Chromatography B, Analytical Technologies in the Biomedical and Life Sciences, 878, 2933–2936.

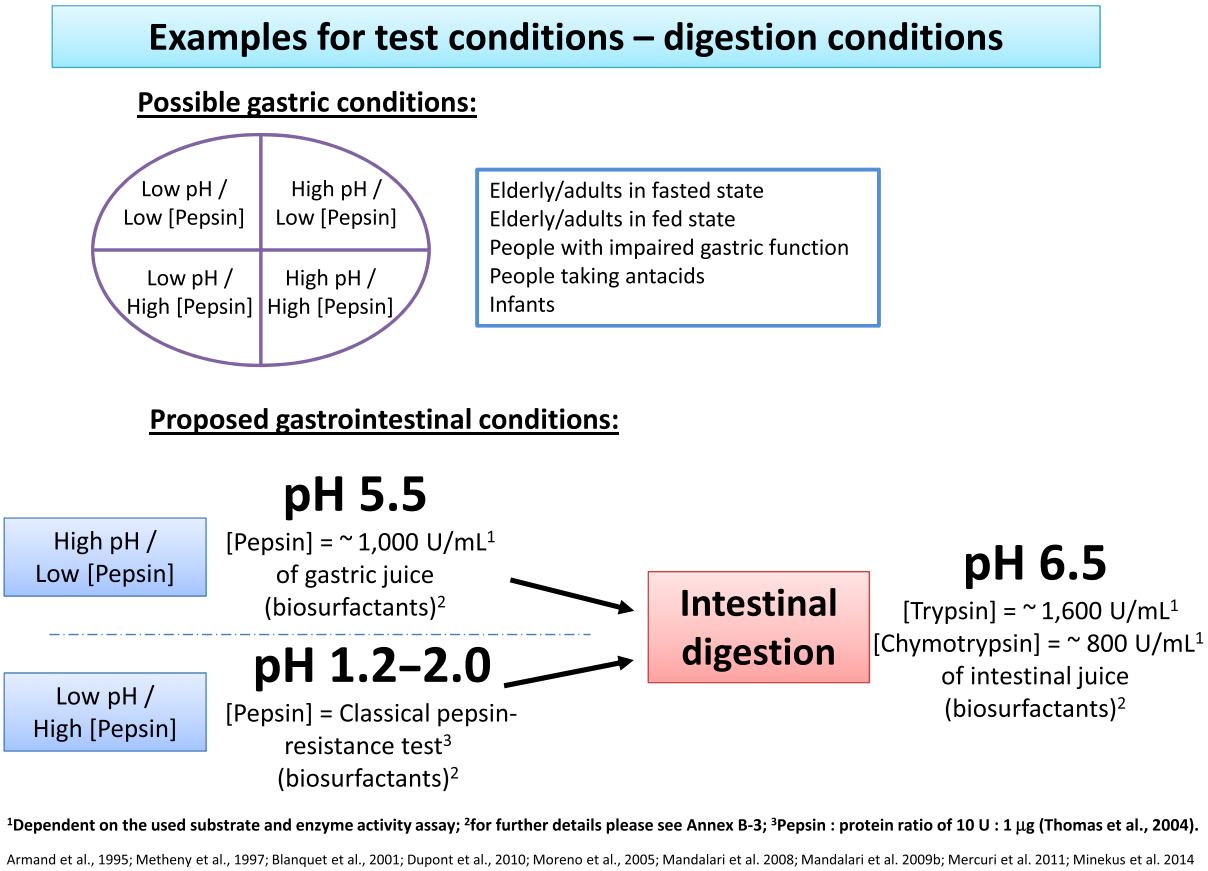



**Figure B.1:** Illustrative examples of *in vitro* gastrointestinal test conditions and proposed gastrointestinal conditions for the interim phase

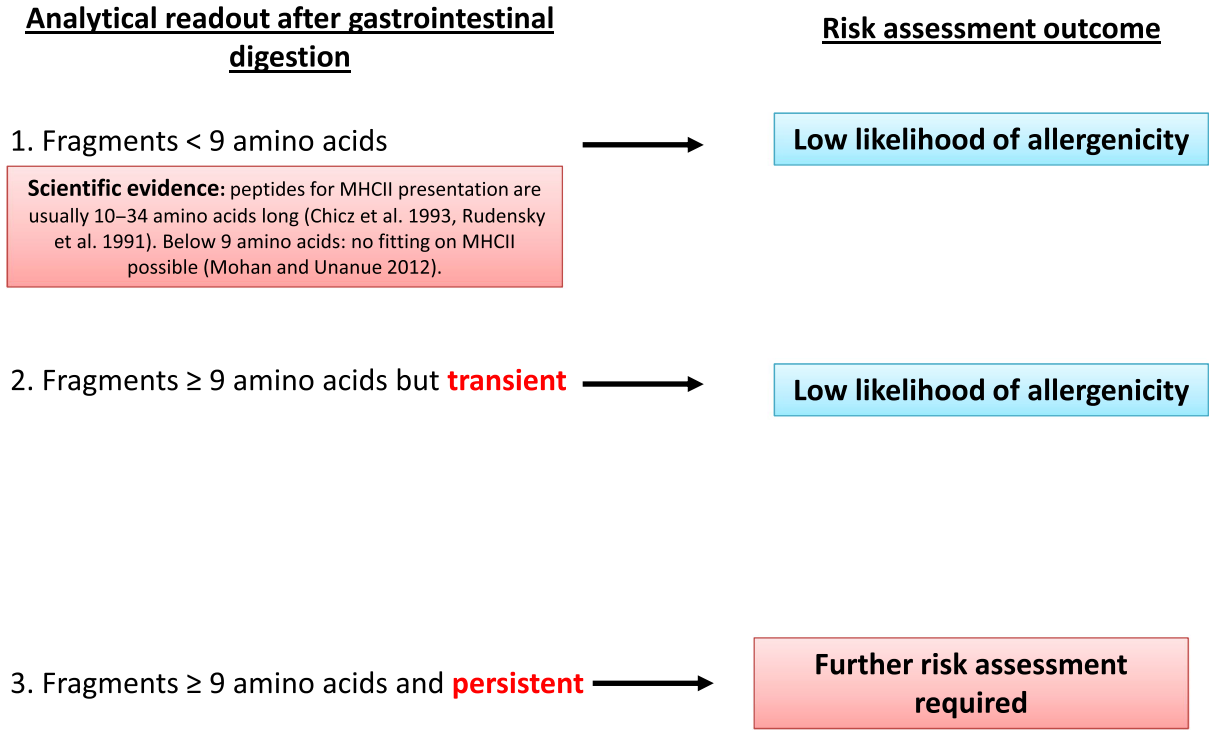



**Figure B.2:** Example of possible scenarios resulting from the *in vitro* digestibility tests and subsequent data interpretation

# Annex C – Endogenous allergenicity

*Annex C-1. History of endogenous allergenicity assessments*

Scientific Opinions on soybean including an assessment of endogenous allergenicity were published by the EFSA GMO panel (e.g. EFSA, 2007, 2011, 2015, 2016, 2017). Historically, the assessment was carried out using sera from allergic individuals, e.g. using 2D gel electrophoresis (immunoblot) in combination with spot quantification. More recently, however, novel absolute quantification methods were developed and applicants employed those in the latest applications submitted to EFSA. To date and considering the EFSA GMO Panel Scientific Opinions published, no significant changes in the allergen repertoire raising concern was observed in the GM plants assessed. The EFSA GMO Panel concluded that there was no evidence that the genetic modification changed the overall allergenicity of the GM plant when compared with its conventional counterpart. EFSA is committed to incorporate latest scientific developments in its risk assessment process and approaches for endogenous allergenicity evaluation should be revisited accordingly in the light of the latest novelties in this area.

*Annex C-2. Possible approach for selection of soybean allergens to be analysed*

The OECD list of 'potential allergens' (Table 20 of OECD, 2012) which is based on allergens listed in the WHO/IUIS database (2011 www.allergen.org) and a review (L'Hocine and Boyne 2007), was taken as a basis, and further complemented with other allergen databases including the updated WHO/IUIS, the FARRP allergen database of the University of Nebraska-Lincoln (www.allergenonline.org) and Allergome (http://allergome.org/). Sequence information was additionally reviewed using the NCBI and Uniprot databases. Other allergen databases, listed in the EFSA 2010 scientific opinion (EFSA GMO Panel, 2010), were also considered, but could not be screened in detail because they did not provide additional information or were no longer available.

Using these sources, data on individual 'potential allergens' and on the allergic individuals tested were retrieved. Furthermore, information was collected connecting clinical reactivity to total soy preparations with data on specific allergens, as well as possible clinical reactions to single purified allergens (either isolated from the plant or produced as recombinant allergens). To date, clinical data for soy allergy were obtained by double-blind placebo-controlled food challenge (DBPCFC) using whole soybean protein extract or soy flour (Ballmer-Weber et al., 2007), while only limited clinical research studies, mostly restricted to component resolved diagnostics (Tuano and Davis, 2015), are available for single proteins. The available information retrieved on the soybean 'potential allergens' are listed in Table C.1 (available in Supporting Information; please see also Selb et al., 2017).

It is challenging to unequivocally connect clinical data obtained on soy extracts by DBPCFC with reactivity to single proteins. In this particular example, taking the information collected for Table C.1 as starting point and following the strategy proposed above, defined criteria were applied for a possible selection of relevant allergens. WHO/IUIS defines criteria to include a protein in the allergen list. These are: (i) a minimum of five sera from allergic individuals (allergic to the allergen source in question) have to contain IgE which binds to the protein in question, or (ii) at least 5% of the allergic individuals' sera tested react with the allergen in question by IgE binding. It is noteworthy that WHO/IUIS database was the main contributor for listing allergens in the OECD document on soybean. WHO/IUIS is also mentioned as a key reference for listing allergens in the EFSA NDA Panel opinion on food allergens (See section 8 of EFSA NDA Panel opinion 2014). WHO/IUIS committee meetings deciding the inclusion of new allergen in the list are held in the frame of EAACI and AAAAI international allergy meetings. Notably, because an application must be filed to include a protein into the WHO/IUIS database, not all proteins clearly fulfilling the requirements are included in the current database. In our example approach following the current WHO/IUIS criteria, the following allergens listed in Table C.1 are also part of the WHO/IUIS database and should be measured accordingly: Gly m 1, Gly m 2, Gly m 3, Gly m 4, Gly m 5, Gly m 6, Gly m 7 and Gly m 8. Out of these, Gly m 2 can currently not be measured since the sequence is unknown. The following proteins potentially fulfil the primary WHO/IUIS criteria of required tested allergic individuals, but did not undergo an expert peer-review by the WHO/IUIS committee: Gly m Bd 28 K, Gly m Bd 30 K, soybean lectin, lipoxygenase, Kunitz trypsin inhibitor and Gly m 50 KD. Out of these proteins, for Gly m Bd 28 K, Gly m Bd 30 K and Kunitz trypsin inhibitor considerable peer-reviewed literature is available and endogenous allergenicity measurement was previously suggested by other scientists (Ladics et al., 2014). In this context, international recognition and/or clinical relevance should be the main aspects to consider for the selection of proteins that will require an endogenous allergenicity assessment.

In contrast, the evidence for the involvement of soybean lectin and lipoxygenase in soy allergy is more scarce (evidence from peer-reviewed scientific literature is limited/non-existent), and additional evidence from clinical studies would be useful. Gly m 50KD is currently not measurable since the sequence is unknown. The other listed proteins, Gly m 39KD, P22-25, Gly m CPI, Gly m EAP and the 'unknown possible allergen' currently do not fulfil the primary WHO/IUIS criteria. Measurement of these would therefore not be performed at the current time. However, more evidence might become available in the future. It should be noted that the above stated considerations (i) might be incomplete and (ii) do not prevent a revision of the scientific progresses in science at any time in the future, which might considerably change the list of potential allergens suitable for assessment as well as the necessity to rank their relevance for the safety assessment.

Even though food allergy to soybean is known in animals (Suto et al., 2015; Kang et al., 2014), importantly, limited data are available on relevant individual soybean allergens in animals. However, certain proteins known to cause soybean allergy in humans (e.g. Gly m 5, Gly m 6) also caused reactions in calves, piglets, dogs and other animals (Taliercio et la., 2014; Lalles and Peltre, 1996; Dreau et al., 1994).

*Annex C-3. Methodology*

For ELISAs, individual allergens should be quantified using purified monoclonal or polyclonal antibodies raised against each purified allergen molecule. The measurement should be performed together with calibrated standards to ensure adequate information on allergen quantities in a sample. Protocols for allergen quantification by quantitative ELISA were developed previously (Geng et al., 2015, 2017; Chen et al., 2014; Hei et al., 2012; Liu et al., 2012, 2013).

Mass spectrometry (MS) approaches can also be used to specifically detect and quantify single allergens. MS protocols for the assessment of endogenous allergenicity of soybean were developed for some potential allergen molecules (Kuppannan et al., 2014; Stevenson et al., 2012; Houston et al., 2011).

The standardisation and harmonisation of these analytical methods among applicants would be beneficial to enhance measurement comparability. This would support the future establishment of an allergen database, including data on the natural variability, which would provide useful additional information to improve the robustness of the safety assessment.

*Annex C-4. Data interpretation and risk assessment considerations*

According to IR503/2013, after comparison to a conventional counterpart within the comparative analysis, natural variability is currently considered the main tool to identify significant and potential relevant changes in allergen content. The European Commission clarified this issue further.[9]

Currently, the experimental field design for comparative compositional analysis requires the inclusion of at least six non-GM reference commercial varieties – the selection of which should be justified according to defined criteria (see section 1.3.2 of IR503/2013). These varieties are used to estimate the overall natural variability to which consumers are routinely exposed. The application of such an approach allows an objective comparative evaluation independently of the absolute content of endogenous allergens ensuring a high degree of protection for consumers.

It is recognised that the natural variability of any given endogenous allergen content estimated from the reference commercial varieties, even if appropriately selected, may not capture the full range of its variability, which is currently unknown. Further efforts in such respect would be beneficial, as described in Annex C-3.

Finally, specific considerations associated with a particular allergenic risk have to be taken into account. Among others (see also Selb et al., 2017), the following considerations are of principle relevance:

- exposure to a certain allergen should be taken into account as a last step in the risk assessment process with particular interest in understanding levels of allergens in foods derived from soybean varieties consumed in Europe by humans and animals at a given time;
- efforts to reduce the uncertainty could be defined according to the single allergen in question for which enhanced allergen content was encountered and for which there is a potential

[9] Letter from European Commission to EuropaBio (Ref. SANCO/El/SP/mb sanco.ddg2.e.l(2014)l 140685) where it was clarified that 'allergens included in the compositional analysis should be treated as any other compound meaning that allergens in the reference varieties included in the assessment should also be analysed (so that an equivalence test as well as a different test can be performed)'.

increase in allergenicity to be communicated to risk managers. The relevance of any given increase might be evaluated with the help of clinical scientist, taking into account: (a) the magnitude of the increase, (b) the potency of the increased potential allergen in question, (c) the overall percentage of the allergen in the non-GM crop and the consequent possible impact of an increase, (d) the route how the allergen is encountered by the allergic individual, (e) the frequency of the potential allergen in question in various products;

- depending on the level of uncertainties identified and, in the case of need, the allergenicity of a GM plant can be compared with that of its appropriate comparator by DBPCFC. This would imply performing a clinical study investigating the reactivity of selected allergic individuals to the varieties under assessment. Therefore, well-characterised allergic individuals reacting with the allergen in question would have to be challenged;
- as a future consideration/perspective, the probability of elicitation of an allergic reaction could be further investigated with the help of dose–distribution curves (Ballmer-Weber et al., 2015) obtained by DBPCFC to single allergens (Kinaciyan et al., 2016) and with reference values (e.g. threshold of elicitation) more precisely evaluated.

## References


Ballmer-Weber BK, Holzhauser T, Scibilia J, Mittag D, Zisa G, Ortolani C, Oesterballe M, Poulsen LK, Vieths S and Bindslev-Jensen C, 2007. Clinical characteristics of soybean allergy in Europe: a double-blind, placebo-controlled food challenge study. The Journal of Allergy and Clinical Immunology, 119, 1489–1496.

Ballmer-Weber BK, Fernandez-Rivas M, Beyer K, Defernez M, Sperrin M, Mackie AR, Salt LJ, Hourihane JO, Asero R, Belohlavkova S, Kowalski M, de Blay F, Papadopoulos NG, Clausen M, Knulst AC, Roberts G, Popov T, Sprikkelman AB, Dubakiene R, Vieths S, van Ree R, Crevel R and Mills EN, 2015. How much is too much? Threshold dose distributions for 5 food allergens. The Journal of Allergy and Clinical Immunology, 135, 964–971.

Chen J, Wang J, Song P and Ma X, 2014. Determination of glycinin in soybean and soybean products using a sandwich enzyme-linked immunosorbent assay. Food Chemistry, 162, 27–33.

Dreau D, Lalles JP, Philouze-Rome V, Toullec R and Salmon H, 1994. Local and systemic immune responses to soybean protein ingestion in early-weaned pigs. Journal of Animal Sciences, 72, 2090–2098.

EC, 2013. Commission Implementing Regulation (EU) No. 503/2013 of 3 April 2013 on applications for authorisation of genetically modified food and feed in accordance with Regulation (EC) No. 1829/2003 of the European Parliament and of the Council and amending Commission Regulations (EC) No. 641/2004 and (EC) No. 1981/2006. Official Journal of European Union L157, 1–48.

EFSA GMO Panel (EFSA Panel on Genetically Modified Organisms, 2010. Scientific Opinion on the assessment of allergenicity of GM plants and microorganisms and derived food and feed. EFSA Journal 2010; 8(7):1700, 168 pp. https://doi.org/10.2903/j.efsa.2010.1700

EFSA GMO Panel (EFSA Panel on Genetically Modified Organisms), 2007. Opinion of the Scientific Panel on genetically modified organisms (GMO) on an application (Reference EFSA-GMO-NL-2005-18) for the placing on the market of the glufosinate tolerant soybean A2704-12, for food and feed uses, import and processing under Regulation (EC) No 1829/2003 from Bayer CropScience. https://doi.org/10.2903/j.efsa.2007.524.

EFSA GMO Panel (EFSA Panel on Genetically Modified Organisms), 2011. Scientific Opinion on application (EFSA-GMO-UK-2007-43) for the placing on the market of herbicide tolerant genetically modified soybean 356043 for food and feed uses, import and processing under Regulation (EC) No 1829/2003 from Pioneer. EFSA Journal 2011;9(7):2310, 40 pp. https://doi.org/10.2903/j.efsa.2011.2310

EFSA GMO Panel (EFSA Panel on Genetically Modified Organisms), 2015. Scientific Opinion on application (EFSA-GMO-NL-2012-108) for the placing on the market of the herbicide-tolerant genetically modified soybean MON 87708 × MON 89788 for food and feed uses, import and processing under Regulation (EC) No 1829/2003 from. EFSA Journal 2015;13(6):4136, 26 pp. https://doi.org/10.2903/j.efsa.2015.4136

EFSA GMO Panel (EFSA Panel on Genetically Modified Organisms), 2017. Naegeli H,Birch AN, Casacuberta J, D e Schrijver A, Gralak MA, Jones H, Manachini B, Messpean A, Nielsen EE,Nogupe F, Robaglia C, Rostoks N, Sweet J, Tebbe C, Visioli F, Wal J-M,pAlvarez F, Ardizzone M, Liu Y,Neri FM and Ramon M, 2017. Scientific opinion on an application by Dow AgroSciences LLC (EFSA-GMO-NL-2012-106) for the placing on the market of genetically modified herbicide-tolerant soybean DAS-44406-6 for food and feed uses, im port and processing under Regulation (EC) No 1829/2003. EFSA Journal 2017;15(3):4738, 33 pp. https://doi.org/10.2903/j.efsa.2017.4738

EFSA GMO Panel (EFSA Panel on Genetically Modified Organisms), 2016. Naegeli H,Birch AN, Casacuberta J, De Schr ijver A, Gralak MA, Guerche P, Jones H, Manachini B, Messpean A,Nielsen EE, Nogupe F, Robaglia C, Rostoks N, Sweet J, Tebbe C, Visioli F, Wal J-M, Divpeki Z, Fernpandez-Dumont A, Gennaro A, Lanzoni A, Neri FM and Paraskevopoulos K, 2016. Scientific Opinion on an application by Dow AgroSciences (EFSA-GMO-NL-2013-116) for placing on the market of genetically modified insect-resistant soybean DAS-81419-2 for food and feed uses, import and processing under Regulation (EC) No 1829/2003. EFSA Journal 2016;14(12):4642, 23 pp. https://doi.org/10.2903/j.efsa.2016.4642

EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2014. Scientific Opinion on the evaluation of allergenic foods and food ingredients for labelling purposes. EFSA Journal 2014;12(11):3894, 286 pp. https://doi.org/10.2903/j.efsa.2014.3894

Geng T, Liu K, Frazier R, Shi L, Bell E, Glenn K and Ward JM, 2015. Development of a Sandwich ELISA for Quantification of Gly m 4, a Soybean Allergen. Journal of Agricultural and Food Chemistry, 63, 4947–4953.

Geng T, Stojšin D, Liu K, Schaalje B, Postin C, Ward J, Wang Y, Liu ZL, Li B and Glenn K, 2017. Natural Variability of Allergen Levels in Conventional Soybeans: Assessing Variation across North and South America from Five Production Years. Journal of Agricultural and Food Chemistry, 65, 463–472.

Hei W, Li Z, Ma X and He P, 2012. Determination of beta-conglycinin in soybean and soybean products using a sandwich enzyme-linked immunosorbent assay. Analytica Chimica Acta, 734, 62e68.

Houston NL, Lee D, Stevenson SE, Ladics GS, Bannon GA, McClain S, Privalle L, Stagg N, Herouet-Guicheney C, MacIntosh SD and Thelen JJ, 2011. Quantitation of soybean allergens using tandem mass spectrometry. Journal of Proteome Research, 10, 763–773.

Kang MH, Kim HJ, Jang HJ and Park HM, 2014. Sensitization rates of causative allergens for dogs with atopic dermatitis: detection of canine allergen-specific IgE. Journal of Veterinary Science, 15, 545–550.

Kinaciyan T, Nagl B, Faustmann S, Kopp S, Wolkersdorfer M and Bohle B, 2016. Recombinant Mal d 1 facilitates sublingual challenge tests of birch pollen-allergic patients with apple allergy. Allergy, 71, 272–274.

Kuppannan K, Julka S, Karnoup A, Dielman D and Schafer B, 2014. 2DLC-UV/MS assay for the simultaneous quantification of intact soybean allergens Gly m 4 and hydrophobic protein from soybean (HPS). The Journal of Agricultural and Food Chemistry, 62, 4884–4892.

Ladics GS, Budziszewski GJ, Herman RA, Herouet-Guicheney C, Joshi S, Lipscomb EA, McClain S and Ward JM, 2014. Measurement of endogenous allergens in genetically modified soybeans–short communication. Regulatory Toxicology and Pharmacology, 70, 75–79.

Lalles JP and Peltre G, 1996. Biochemical features of grain legume allergens in humans and animals. Nutrition Reviews, 54, 101–107.

L'Hocine L and Boye JI, 2007. Allergenicity of soybean: new developments in identification of allergenic proteins, cross-reactivities and hypoallergenization technologies. Critical Reviews in Food Science and Nutrition, 47, 127–143.

Liu B, Teng D, Wang X, Yang Y and Wang J, 2012. Expression of the soybean allergenic protein P34 in *Escherichia coli* and its indirect ELISA detection method. Applied Microbiology and Biotechnology, 94, 1337–1345.

Liu B, Teng D, Wang X and Wang J, 2013. Detection of the soybean allergenic protein Gly m Bd 28K by an indirect enzyme-linked immunosorbent assay. Journal of Agricultural and Food Chemistry, 61, 822–828.

OECD, 2012. Revised consensus document on compositional considerations for new varieties of soybean [Glycine max (L.) Merr.]: key food and feed nutrients, antinutrients, toxicants and allergens. Series on the Safety of Novel Foods and Feeds No. 25.

Selb R, Wal JM, Moreno FJ, Lovik M, Mills C, Hoffmann-Sommergruber K and Fernandez A, 2017. Assessment of endogenous allergenicity of genetically modified plants exemplified by soybean - Where do we stand? Food and Chemical Toxicology, 101, 139–148.

Suto A, Suto Y, Onahara N, Tomizawa Y, Yamamoto-Sugawara Y, Okayama T and Masuda K, 2015. Food allergens inducing a lymphocyte-mediated immunological reaction in canine atopic-like dermatitis. Journal of Veterinary Medical Science, 77, 251–254.

Stevenson SE, Woods CA, Hong B, Kong X, Thelen JJ and Ladics GS, 2012. Environmental effects on allergen levels in commercially grown non-genetically modified soybeans: assessing variation across North America. Frontiers in Plant Science, 3, 1–13.

Taliercio E, Loveless TM, Turano MJ and Kim SW, 2014. Identification of epitopes of the beta subunit of soybean beta-conglycinin that are antigenic in pigs, dogs, rabbits and fish. Journal of the Science of Food and Agriculture, 94, 2289–2294.

Tuano KS and Davis CM, 2015. Utility of component-resolved diagnostics in food allergy. Current Allergy and Asthma Reports, 15, 32.