https://doi.org/10.15388/vu.thesis.962
https://orcid.org/0000-0003-1910-1509

VILNIUS UNIVERSITY

Juta Dvareckienė

Production and Characterization of Diagnostic-Grade Recombinant Allergens: from Expression Platform Selection to Clinical Application

DOCTORAL DISSERTATION

Technological Sciences,
Chemical Engineering (T 005)

VILNIUS 2026

The dissertation was prepared between 2018 and 2022 during the studies at Vilnius University Life Sciences Centre Institute of Biotechnology.
The research was supported by European Regional Development Fund (Project No. 01.2.2-LMT-K-718-01-0008) under grant agreement with Research Council of Lithuania (LMTLT).

The dissertation is defended on an external basis.

Academic consultant – Dr. Rasa Petraitytė-Burneikienė (Vilnius University, Technological Sciences, Chemical Engineering – T 005).

This Doctoral Dissertation will be Defended in a Public Meeting of the Dissertation Defence Panel:

Chair – Dr. Inga Matijošytė (Vilnius University, Technological Sciences, Chemical Engineering – T 005).
Members:
Assoc. Prof. Dr. Violeta Jonušienė (Vilnius University, Natural Sciences, Biochemistry – N 004),
Dr. Andris Kazaks (Latvian Biomedical Research and Study Centre, Technological Sciences, Chemical Engineering – T 005),
Prof. Dr. Laura Malinauskienė (Vilnius University, Medical and Health Sciences, Medicine – M 001),
Dr. Jolanta Sereikaitė (Vilnius University, Technological Sciences, Chemical Engineering – T 005).

The dissertation shall be defended at a public meeting of the Dissertation Defense Panel at 13:00 on 18 September 2026 in Room R-401, Life Sciences Center, Vilnius University, and remotely. Address: Saulėtekio av. 7, LT-10257, Vilnius, Lithuania.

https://doi.org/10.15388/vu.thesis.962
https://orcid.org/0000-0003-1910-1509

VILNIAUS UNIVERSITETAS

Juta Dvareckienė

Diagnostinės kokybės rekombinantinių alergenų sintezė ir apibūdinimas: nuo raiškos sistemos pasirinkimo iki klinikinio pritaikymo

DAKTARO DISERTACIJA

Technologijos mokslai,
chemijos inžinerija (T 005)

VILNIUS 2026

Disertacija rengta 2018-2022 metais Vilniaus universitete, Gyvybės mokslų centro Biotechnologijos institute.
Mokslinius tyrimus rėmė Europos regioninės plėtros fondo lėšos (projekto Nr. 01.2.2-LMT-K-718-01-0008) pagal dotacijos sutartį su Lietuvos mokslo taryba (LMTLT).

Disertacija ginama eksternu.

Mokslinė konsultantė – dr. Rasa Petraitytė-Burneikienė (Vilniaus universitetas, technologijos mokslai, chemijos inžinerija – T 005).

Gynimo taryba:
Pirmininkė – dr. Inga Matijošytė (Vilniaus universitetas, technologijos mokslai, chemijos inžinerija – T 005).
Nariai:
doc. dr. Violeta Jonušienė (Vilniaus universitetas, gamtos mokslai, biochemija – N 004),
dr. Andris Kazaks (Latvijos biomedicinos tyrimų ir mokymosi centras, technologijos mokslai, chemijos inžinerija – T 005),
prof. dr. Laura Malinauskienė (Vilniaus universitetas, medicinos ir sveikatos mokslai, medicina – M 001),
dr. Jolanta Sereikaitė (Vilniaus universitetas, technologijos mokslai, chemijos inžinerija – T 005).

Disertacija ginama viešame Gynimo tarybos posėdyje 2026 m. Rugsėjo mėn. 18 d. 13 val. Vilniaus universiteto Gyvybės mokslų centro R-401 auditorijoje ir nuotoliniu būdu. Adresas: Saulėtekio al. 7, LT-10257, Vilnius, Lietuva.

Contents

ABBREVIATIONS 9
INTRODUCTION 11
1. LITERATURE REVIEW 15
1.1 Immune and non-immune mediated adverse reactions to food 15
1.2 IgE-mediated allergy: mechanism and pathophysiology 18
1.2.1 Sensitization 18
1.2.2 Early- and late-phase reactions 24
1.2.3 Chronic allergic inflammation 28
1.3 Epidemiology of allergies 29
1.4 Risk factors for allergic disease 33
1.4.1 Genetic factors 33
1.4.2 Environmental factors 35
1.4.3 Allergen molecular characteristics 39
1.5 Allergy diagnostics 43
1.5.1 Recombinant allergens vs. Allergen extracts 44
1.5.2 *In vivo* vs. *in vitro* allergy diagnostics 47
1.5.3 Allergen immunotherapy 49
1.6 Molecular allergology: potential adaptations 51
1.6.1 Food allergy: *Penaeus monodon* allergen Pen m 4 . . . 51
1.6.2 Pet allergy: *Canis familiaris* allergen Can f 6 53
1.6.3 Insect venom allergy: *Vespula vulgaris* allergen Ves v 5 56
2. MATERIALS AND METHODS 59
2.1 Materials and Resources 59
2.1.1 Oligonucleotides and plasmid vectors 59
2.1.2 Bacteria and yeast strains 63
2.1.3 Protein coding sequences 64
2.1.4 Human sera 64
2.1.5 Bioinformatic and programming tools 65
2.2 Methods with associated reagents 67
2.2.1 DNA restriction and dephosphorylation 67
2.2.2 Oligonucleotide hybridization 68
2.2.3 DNA electrophoresis 68
2.2.4 DNA fragment purification from an electrophoresis gel 68

2.2.5 DNA ligation 69
2.2.6 Polymerase chain reaction (PCR) 69
2.2.7 Plasmid DNA purification from *E. coli* 70
2.2.8 Preparation of chemically competent *E. coli* cells and transformation via heat shock 72
2.2.9 Preparation of chemically competent *S. cerevisiae* cells and transformation via heat shock 73
2.2.10 Preparation of electrocompetent *P. pastoris* cells and transformation via electroporation 74
2.2.11 Induction of recombinant protein expression in *E. coli* 75
2.2.12 Induction of recombinant protein expression in *S. cerevisiae* 75
2.2.13 Induction of recombinant protein expression in *P. pastoris* 76
2.2.14 Protein SDS-PAGE and polyacrylamide gel staining . . 77
2.2.15 *E. coli* cell lysis for protein purification 78
2.2.16 Mechanical tearing of yeast cells 79
2.2.17 Rapid total protein extraction from yeast by alkaline lysis 79
2.2.18 Manual Ni-NTA batch purifications 80
2.2.19 Purification of proteins using ZYMO Reasearch columns 81
2.2.20 Protein dialysis 82
2.2.21 Protein concentration from yeast media with methanol 83
2.2.22 Ion chromatography 84
2.2.23 Ni-NTA chromatography 85
2.2.24 Protein concentration determination 86
2.2.25 Protein long term storage 87
2.2.26 Western blot (WB) 88
2.2.27 Glycosylation assay 89
2.2.28 Deglycosylation assay 90
2.2.29 Electrospray ionization mass spectometry (ESI-MS) . . 90
2.2.30 Indirect enzyme-linked immunosorbent assay (ELISA) 90
2.2.31 Reverse-phase protein microarray 92
2.2.32 Statistical analysis for ELISA datasets 93
2.2.33 Concentration selection after reverse-phase microarray experiment 94
2.2.34 Statistical analysis for microarray datasets 94

3. RESULTS AND DISCUSSION. 96
3.1 Selection of expression platforms for the production of recombinant allergens 97
3.2 Recombinant allergen Pen m 4 production and immunological characterization 100
3.2.1 Generation of codon-optimized Pen m 4 allergen sequence 100
3.2.2 Cloning, expression, and purification of rPen m 4 from *E. coli* 100
3.2.3 Characterization of rPen m 4 variants purified from *E. coli* 103
3.2.4 Molecular cloning and expression of rPen m 4 in *P. pastoris* 104
3.2.5 Purification and characterization of *P. pastoris*-produced rPen m 4 106
3.2.6 Purification and characterization of *P. pastoris*-produced rMBP-Pen m 4 110
3.2.7 Immunoreactivity of rPen m 4 and rMBP-Pen m 4 . . . 113
3.2.8 CCD interference in IgE binding to rPen m 4 120
3.2.9 IgE reactivity to rPen m 4 in Ca^{2+}-depleted environment 122
3.3 Recombinant allergen Can f 6 production and immunological characterization 125
3.3.1 Generation of codon-optimized Can f 6 allergen sequence 125
3.3.2 Cloning, expression, and purification of rCan f 6 from *E. coli* 125
3.3.3 Characterization of rCan f 6 variants purified from *E. coli* 127
3.3.4 Molecular cloning and expression of rCan f 6 in *P. pastoris* 129
3.3.5 Purification and characterization of *P. pastoris*-produced rCan f 6 131
3.3.6 Purification and characterization of *P. pastoris*-produced rMBP-Can f 6 132
3.3.7 Immunoreactivity of rCan f 6 and rMBP-Can f 6 135
3.4 Streamlining PichiaPink™ recombinant allergen production . 141
3.4.1 Construction of optimized vectors for allergen expression and purification 141
3.4.2 Validation of the expression platform using yellow jacket venom allergen Ves v 5 142

SUMMARY. 151
CONCLUSIONS 152
BIBLIOGRAPHY 153
SANTRAUKA. 167
PUBLICATION LIST 203
CONFERENCES 204
CURRICULUM VITAE 205
PADĖKA 206
NOTES. 207

ABBREVIATIONS

Note: In this list, "#" denotes a variable number (e.g., T_H1, T_H2).

AD	Atopic dermatitis
AEX	Anion exchange chromatography
AFR	Adverse food reactions
AGE	Advanced glycation end product
AIT	Allergen immunotherapy
APC	Antigen-presenting cells
AVMA	American Veterinary Medical Association
BSA	Bovine serum albumin
CCD	Carbohydrate determinants
CCL2	CC-chemokine ligand 2
CD#	Cluster of differentiation #
CD#L	CD # ligand
CEX	Cation exchange chromatography
CRD	Component-resolved diagnosis
DBPCFC	Double-blind placebo-controlled food challenge
DC	Dendritic cells
DEP	Diesel exhaust particle
EDTA	Ethylenediaminetetraacetic acid
EGID	Eosinophilic gastrointestinal disorder
ELISA	Enzyme-linked immunosorbent assay
ER	Endoplasmic reticulum
FAO	Food and Agriculture Organization of the United Nations
FEDIAF	European Pet Food Industry Federation (fr: *Fédération Européenne de l'Industrie des Aliments pour Animaux Familiers*)
FcεRI	High-affinity IgE receptor I
FcεRII	Low-affinity IgE receptor II, also known as CD23
FLG	Filaggrin gene
FPIES	Food protein-induced enterocolitis syndrome
GlcNAc	N-acetylglucosamine
GRAS	Generally regarded as safe
GWAS	Genome-wide association studies
HLA	Human leukocyte antigen
ICAM-1	Intercellular adhesion molecule 1
IgE	Immunoglobulin E
IgG	Immunoglobulin G
IL#	Interleukin #

LFA-1	Lymphocyte function-associated antigen 1
LTB_4	Leukotriene B4
LTC_4	Leukotriene C4
MBP	Maltose-binding protein
MeOH	Methanol
MFI	Median fluorescence intensity
MHC	Major histocompatibility complex
Neu5Gc	N-glycolylneuraminic acid
Ni-NTA	Nickel-nitrilotriacetic acid
PCR	Polymerase chain reaction
PG_2	Prostaglandin 2
PGD_2	Prostaglandin D2
PM	Particulate matter
PTM	Post-translational modification
RT	Room temperature
SCP	Sarcoplasmic calcium-binding protein
sIgE	Specific IgE
SPT	Skin prick test
TCR	T cell receptor complex
T_H#	T helper # cells
TNF-α	Tumor necrosis factor α
T_{reg}	T regulatory cells
VEGFA	Vascular endothelial growth factor A
VIT	Venom immunotherapy
VU LSC IBT	Vilnius University Life Sciences Center Institute of Biotechnology
WB	Western blot
WHO	World Health Organization
w/v	Weight-on-volume
w/w	Weight-on-weight

INTRODUCTION

Allergic diseases have emerged as a major public health concern, with their prevalence steadily rising over recent decades (Shin et al. 2023). Rapid urbanization, industrialization, and lifestyle changes have been implicated in allergy development, acting through increased exposure to pollutants, altered dietary habits, and reduced microbial diversity (Shin et al. 2023; Q Zhang et al. 2025; Kirjavainen et al. 2019). These environmental influences shape immune development and may help explain the growing burden of allergic disorders worldwide. Most clinically relevant allergic diseases involve inappropriate immune recognition of otherwise harmless environmental or food-derived antigens.

IgE-mediated allergy is a type I hypersensitivity reaction with a type 2 immune response, characterized by the involvement of allergen-specific IgE and T helper 2 cells that recognize allergen-derived antigens. Allergic inflammation can manifest in diverse ways, from rhinitis and urticaria to anaphylaxis (Renz et al. 2018; Anvari et al. 2019).

At the molecular level, only a small amount of the proteome act as allergens. Their allergenic potential arises not from a single defining property but from a complex interplay of abundance, structural stability, biochemical modifications, and the context of exposure (Costa et al. 2022). Yet, considerable variability exists even within closely related protein families, underscoring that additional factors, such as three-dimensional structure, epitope accessibility, and host-specific responses, may contribute to sensitization (Lukacs and Hogan 2025). Posttranslational modifications, particularly glycosylation, further complicate allergenicity by modulating IgE binding and enabling cross-reactivity. Additionally, common food-processing methods, including heating or pasteurization, can drastically alter antigenicity (Zeng et al. 2025). Together, these findings highlight that understanding the determinants of allergenicity remains one of the central challenges of molecular allergology.

This complexity is directly reflected in the field of allergy diagnostics, where distinguishing clinically relevant sensitization from cross-reactivity remains an ongoing challenge. Historically, allergen extracts derived from natural sources were used to diagnose and treat allergic diseases. However, their variability, instability, lack of standardization, and contamination with

non-allergenic components limited their accuracy and safety (Zimmer et al. 2021). Recombinant allergens have revolutionized the field of allergology by enabling a more consistent and precise approach to allergy diagnostics. They facilitate the identification of specific allergens responsible for allergic reactions, a process termed component-resolved diagnosis, which allows clinicians to distinguish between genuine sensitization and cross-reactivity, refine risk assessment, and provide more personalized therapeutic strategies (Sánchez-Borges et al. 2018).

Despite these advances, major challenges remain. Allergies can be managed either by avoidance, self-medication or by allergen immunotherapy. Allergen immunotherapy is a form of therapeutic vaccination that aims to develop functional immune system tolerance towards allergens (Durham and Shamji 2023). However, the process is lengthy, not universally effective, and still relies largely on crude allergen extracts. Next-generation strategies, including recombinant allergens, engineered hypoallergens, and emerging peptide- or nucleic acid-based vaccines hold promise for safer and more effective interventions (Zemelka-Wiacek et al. 2024). However, their success ultimately depends on precise molecular characterization and robust biotechnological production platforms capable of generating clinically relevant proteins.

Selecting an expression system for recombinant allergens requires careful consideration of factors such as posttranslational modifications, structural complexity needed to preserve conformational epitopes, and protein solubility. Practical constraints, including production time, efficiency, and regulatory compliance for therapeutic applications, must also be addressed (M Liu et al. 2025). Advancing our understanding of allergen expression in suitable recombinant systems is therefore crucial for both diagnostics and the development of next-generation therapeutic approaches.

The aim of this research

To produce and characterize recombinant black tiger shrimp *Penaeus monodon* Pen m 4, dog *Canis familiaris* Can f 6, and yellow jacket *Vespula vulgaris* Ves v 5 allergens in *Pichia pastoris* while assessing this system's suitability for generating diagnostic-grade allergens.

Work objectives

1. To compare *Saccharomyces cerevisiae* and *P. pastoris* expression systems for recombinant allergen protein production.

2. To synthesize Pen m 4 and Can f 6 allergens as native and fused to the maltose-binding protein (MBP) in *E. coli* and *P. pastoris* expression systems.

3. To characterize IgE-binding reactivity of all recombinant allergen variants using sera from allergic patients.

4. To compare IgE reactivity between *E. coli*- and *P. pastoris*-produced allergens to assess the importance of posttranslational modifications.

5. To validate the optimized *P. pastoris* production workflow using Ves v 5 as a model insect venom allergen and evaluate diagnostic performance across all produced allergens.

Relevance and scientific novelty

This study represents the first systematic evaluation of Pen m 4 and Can f 6 allergen production in *P. pastoris*, providing comparative data against the traditional *E. coli* system. Protein expression was enhanced using maltose-binding protein (MBP) fusion, an established strategy in *E. coli* but rarely investigated in *P. pastoris*. Here, MBP improved recombinant allergen production without compromising downstream diagnostic use, as MBP does not bind IgE antibodies. System performance was further validated with Ves v 5, which demonstrated high-yields and system-dependent effects on antigenicity. This confirmed the adaptability of the created workflow for the production of recombinant allergens. Mechanistically, the work revealed proteolytic cleavage occurring downstream of the MBP domain, consistent with earlier reports, and, to our knowledge, novel N-terminal cleavage of MBP fusion protein in *P. pastoris*. These findings highlight important considerations for MBP fusion protein design in *P. pastoris*. In addition, *P. pastoris*-produced Pen m 4 was found to form unusually stable dimers resistant to chemical and thermal denaturation while maintaining IgE reactivity – a glycosylation-modulated behavior seen only in two other cases (Z Li et al. 2010; AK Walker et al. 2013).

Clinically, this research addresses gaps in allergy diagnostics. Commercial shrimp tests rely predominantly on tropomyosin (Pen m 1), a panallergen with extensive cross-reactivity that causes false-positive results, while dog allergy diagnostics suffer from extract variability and incomplete allergen representation. By producing and characterizing Pen m 4 and Can f 6, allergens known to cause monosensitization and define distinct patient subpopulations,

this work provides new molecular tools for component-resolved diagnostics. The created workflow is also beneficial to known clinically important major allergens like the yellow jacket venom allergen Ves v 5, by creating a high-yield, high-quality production tool. The study also identified key variables that affect diagnostic reliability: EDTA, a common anticoagulant in clinical blood collection, reduced IgE reactivity to Pen m 4 by 61-73% through calcium chelation; and lyophilization was found to degrade both Pen m 4 and Can f 6, explaining their instability in commercial extracts. Collectively, these findings establish evidence-based protocols for recombinant allergen production and highlight critical preanalytical and stability factors that must be controlled to ensure diagnostic accuracy.

Defending statements

1. *P. pastoris* PichiaPink™ expression system can be used to produce recombinant allergens that retain proper folding structure required for IgE recognition, and are suitable for use in *in vitro* diagnostic systems.

2. MBP fusion enhances allergen yield in *P. pastoris* without compromising diagnostic functionality.

3. N-terminal proteolytic processing in *P. pastoris* MBP-fusion constructs represent a previously undocumented susceptibility to proteolysis, providing new insights for fusion protein design in *P. pastoris* systems.

4. Standard preanalytical practices, from EDTA anticoagulant use in blood collection to lyophilization-based allergen preservation, can introduce unrecognized variables significantly affecting diagnostic sensitivity and specificity in component-resolved allergy testing.

5. *P. pastoris*-derived glycosylation modifications can induce significant structural changes in recombinant allergens, including enhanced protein stability, while preserving immunological functionality for diagnostic applications.

1. LITERATURE REVIEW

1.1. Immune and non-immune mediated adverse reactions to food

An allergy is an inappropriate immune response to otherwise harmless environmental, food-derived, or venom-derived antigens. These antigens, termed allergens, can induce sensitization and trigger recurrent inflammatory reactions upon subsequent exposure (Galli, Tsai, and Piliponsky 2008). Depending on the underlying immune mechanism, allergic reactions may involve IgE antibodies, cellular immune responses, or a combination of both, resulting in clinical manifestations ranging from mild local symptoms to systemic anaphylaxis. The literature review first introduces allergic disease in the broader context of adverse reactions and then narrows to IgE-mediated mechanisms.

"Adverse food reactions" (AFR) affect up to 20-30% of the population in developed countries, with nearly half of these remaining undiagnosed (Onyimba et al. 2021; Mazur et al. 2022). Symptoms range from mild to severe, substantially reducing quality of life and imposing a considerable socioeconomic burden (Zuberbier et al. 2014). These reactions comprise abnormal clinical responses following ingestion of specific food or food components, classified according to their pathophysiological mechanisms into two primary categories: non-immune mediated food intolerances and immune-mediated hypersensitivities (Fig. 1.1). Despite their clinical similarities, accurate distinction is essential for appropriate management, as immune-mediated reactions can trigger severe systemic responses from trace exposures, while non-immune intolerances typically demonstrate dose-dependency and variable severity (Cianferoni and Spergel 2009; Onyimba et al. 2021).

Non-immune mediated food intolerances arise from pharmacologic, toxicological, or host-specific factors that impair normal food processing. The most prevalent example, lactose intolerance, results from reduced intestinal lactase activity, leading to malabsorption of milk disaccharides. This produces dose-dependent gastrointestinal symptoms (diarrhea, bloating, abdominal pain) that may clinically resemble milk allergy but permit consumption of lactose-free dairy products (Walsh et al. 2016). Similarly, non-celiac wheat sensitivity elicits short-term bloating and abdominal pain following gluten ingestion, which can be alleviated by a low-gluten diet rather than complete

elimination (Onyimba et al. 2021). These conditions illustrate the characteristic features of food intolerances: threshold doses for symptom onset, day-to-day variability, and management through reduction rather than avoidance (Nelson and Ogden 2008).

In contrast to the dose-dependent, variable nature of food intolerances, immune-mediated AFR involves specific immunological mechanisms that can be triggered by trace amounts of otherwise harmless antigens, producing reproducible and potentially life-threatening responses. These immune-mediated reactions are classified into four subcategories: IgE-mediated, non-IgE-mediated, mixed IgE/cellular, and cell-mediated hypersensitivities. Each category demonstrates distinct pathophysiology, clinical manifestations, and diagnostic approaches, necessitating different management strategies (Anvari et al. 2019; Renz et al. 2018).

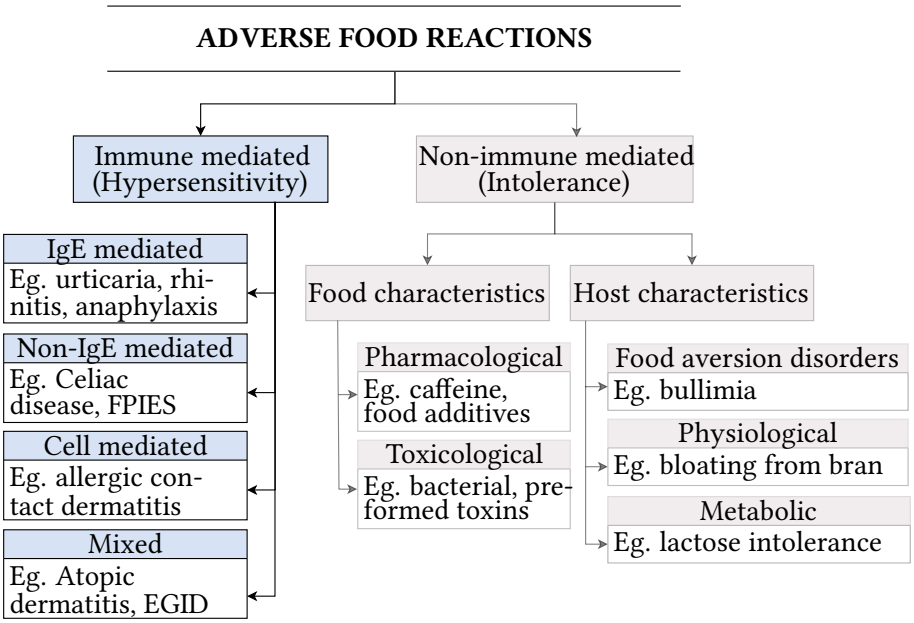


Figure 1.1: Classification of adverse food reactions. Adapted and modified from Anvari et al. 2019 and Onyimba et al. 2021. Made with Diagrams.net (see 2.4).

IgE-mediated food allergies, representing type I immediate hypersensitivity reactions, constitute the most clinically relevant subset of AFR. They involve an exaggerated immune response characterized by allergen-specific IgE (sIgE) antibodies bound to mast cells and basophils. Upon re-exposure, cross-linking of these antibodies triggers rapid degranulation, releasing in-

flammatory mediators that produce symptoms ranging from urticaria and rhinitis to life-threatening anaphylaxis within minutes to hours of ingestion (Anvari et al. 2019). The immediate onset and potential severity necessitate strict avoidance of all forms of the offending food, including products with precautionary ("may contain") labeling.

Non-IgE-mediated food allergies comprise a heterogeneous group of disorders primarily affecting the gastrointestinal tract through cellular immune mechanisms. These include allergic proctocolitis, the most common form presenting as mild rectal bleeding in infants; food protein-induced enterocolitis syndrome (FPIES), which can cause severe vomiting and hypovolemic shock; and celiac disease, an autoimmune enteropathy triggered by gluten, that activates gluten-specific CD4+ T cells, leading to cytokine-driven cytotoxic T-cell destruction of intestinal epithelial cells (Connors et al. 2018; Berin 2015). Unlike IgE-mediated reactions, these conditions typically manifest hours to weeks after ingestion and lack reliable biomarkers, requiring diagnosis through elimination diets, oral food challenges, medical history, endoscopy, and biopsy findings (Nowak-Węgrzyn et al. 2015).

Mixed mechanism disorders demonstrate IgE and cellular immunity features, as observed in conditions such as eosinophilic gastrointestinal disorder (EGID) and atopic dermatitis (AD). EGID patients often test positive for food-specific IgE yet rarely experience anaphylaxis, instead developing dense eosinophilic infiltration at gastrointestinal inflammation sites (Rothenberg 2004). Similarly, in 90% of AD patients, tests show elevated IgE levels and T-cell infiltration in skin lesions (Spergel 2006). AD is an internal skin condition triggered by ingesting milk, eggs, soy, peanuts, or wheat. Allergic contact dermatitis (ACD) is often confused with AD. However, ACD is a pure cell-mediated condition, caused by low-molecular-weight haptens, which require binding to proteins capable of associating with major histocompatibility complex (MHC) class II molecules. These haptens can be found in certain foods (oleosins in fruits, vegetables, spices) and metals (nickel, cobalt, chromium). Cell-mediated reactions occur through T-cell sensitization without IgE involvement, manifesting as delayed skin reactions upon direct contact (Kostner et al. 2017).

While a comprehensive understanding of all AFR mechanisms remains clinically significant, the increasing prevalence, severity, and socioeconomic impact of IgE-mediated allergies have driven substantial research into improved diagnostics and therapeutics (Mazur et al. 2022). This literature review,

therefore, concentrates on IgE-mediated allergic mechanisms, examining allergy pathophysiology, epidemiology, risk factors, allergen characterization, current diagnostic capabilities, and therapeutics. Through detailed analysis of several molecular allergology adaptations, this work aims to contribute to the improved understanding and management of these allergies.

1.2. IgE-mediated allergy: mechanism and pathophysiology

Food allergies were documented as early as 2700 BC by Chinese emperors; however, a definitive link between food allergens and antibodies was not established until the 1920s (Lukacs and Hogan 2025). The Prausnitz-Küstner experiment was a pivotal study demonstrating that immediate-type hypersensitivity reactions could be transferred via serum. In this experiment, Heinz Küstner, who exhibited systemic allergic responses to cooked fish, provided his serum to Carl Prausnitz, who then injected this serum intracutaneously into his skin. Twenty-four hours later, a classic wheal-and-flare reaction was observed upon challenge with boiled fish extract at the same site. This experiment provided the first empirical evidence linking allergic reactions to a circulating, serum-derived factor, establishing a foundation for the modern understanding of IgE-mediated hypersensitivity (Lukacs and Hogan 2025).

IgE-mediated allergy (hereinafter referred to as "allergy") is a type I hypersensitivity reaction with a type 2 immune response, characterized by the involvement of allergen-specific IgE and T helper 2 (T_H2) cells that recognize allergen-derived antigens. Allergy development is impacted by many different factors: genetic predisposition, allergen type, frequency and/or route of exposure, environmental allergen concentration, whether first contact is made with other immunomodulating agents, and so on (Renz et al. 2018; Shi et al. 2025). However, the mechanisms by which immune cells, involved in type 2 immune reactions, recognize a wide array of unrelated allergens, is unclear. To define allergy and its triggers, the molecular drive inducing type 2 inflammation needs to be understood (Lambrecht 2025; Shi et al. 2025).

1.2.1 Sensitization

The first contact with an allergen is called sensitization, which reflects the allergen's ability to induce a T_H2 cell response, in which IL-4, IL-5, and IL-13 cytokines drive IgE production by promoting immunoglobulin class-switch recombination in B cells (Galli, Tsai, and Piliponsky 2008). This process begins when an allergen breaches the epithelial barrier, coming into contact

with professional antigen-presenting cells (APC). APCs, such as dendritic cells (DC), macrophages, and B cells, are located at environmental interfaces, including the epithelial barrier, mucous membrane, digestive system, and respiratory tract. DCs are the most effective APCs, capable of activating naive T cells and initiating the T cell response. Immature DCs capture allergens via three primary mechanisms: 1) receptor-mediated endocytosis, 2) constitutive macropinocytosis, and 3) phagocytosis (Lambrecht 2001; Hammad and Lambrecht 2008).

Receptor-mediated endocytosis occurs via C-type lectin receptors such as langerin, DC-SIGN (dendritic cell-specific intercellular adhesion molecule 3-grabbing non-integrin), dectin (dendritic cell-associated C-type lectin-1), macrophage mannose receptor, and carbohydrate receptor DEC-205 (dendritic endocytic receptor-205 kDa) (Lambrecht 2001; K Peters and M Peters 2021). Studies have shown that these C-type lectin receptors interact with various allergens – for example, allergenic pollen starch granules via C-type lectin and integrin receptors (Currie et al. 2000), and glycoproteins in allergenic foods via the DC-SIGN receptor (Kamalakannan et al. 2016), among others. Investigation of DC receptor-mediated endocytosis for specific allergens may provide a means to modulate the inflammatory response associated with allergies.

Constitutive macropinocytosis, also known as fluid-phase endocytosis, is a non-specific receptor-ligand-independent endocytosis characterized by plasma membrane ruffling, leading to the internalization of extracellular fluid (Lin et al. 2018). Allergens such as birch allergen Bet v 1 and grass allergen Phl p 1 have been shown to be internalized primarily via macropinocytosis (Noirey et al. 2000).

Immature DCs also use phagocytosis for the capture of antigenic particles such as latex beads, apoptotic and necrotic cell fragments, viruses, bacteria, and intracellular parasites (Thery and Amigorena 2001). Unlike the previously described uptake mechanisms, phagocytosis not only enables APCs to process and present allergen-derived peptides but also serves to remove cellular debris and pathogens. Consequently, it remains unclear whether an increase of phagocytic activity would reduce inflammation by removing allergenic proteins or increase it by promoting antigen presentation (Poole et al. 2009).

The allergens taken up by any of the three mechanisms accumulate in the endocytic compartment. Internalization of the allergen (antigen) triggers the DCs' maturation process, characterized by functional changes (increased

synthesis of T cell stimulating molecules) and cell migration from peripheral tissues to secondary lymphoid organs (Thery and Amigorena 2001). DCs process the internalized allergens during maturation into short peptides within endosomal and lysosomal vesicles. In parallel, biosynthesis of MHC II class molecules in the endoplasmic reticulum (ER) increases, and their transport to the endosomes can be seen. Processed peptides associate with MHC II molecules following fusion of the vesicular transport pathway, which carries newly synthesized MHC II molecules, and the lysosomal degradation pathway, which carries proteolyzed proteins. This peptide-MHC II complex is then transported to the cell surface, a process directly induced by DC maturation. Immature DCs retain most MHC II molecules within late endosomal and lysosomal compartments. In contrast, mature DCs display nearly all MHC II class molecules on the cell surface (Lambrecht 2001; Murray et al. 2013).

Antigen presentation to naive T cells occurs in secondary lymphoid organs, which include lymph nodes, spleen, and mucosa-associated lymphoid tissue. Activation of the antigen-specific T cell response requires coordinated cytokine and cell-to-cell receptor interactions initiated by the T cell receptor complex (TCR) and the MHC II-antigen complex (Fig. 1.2).

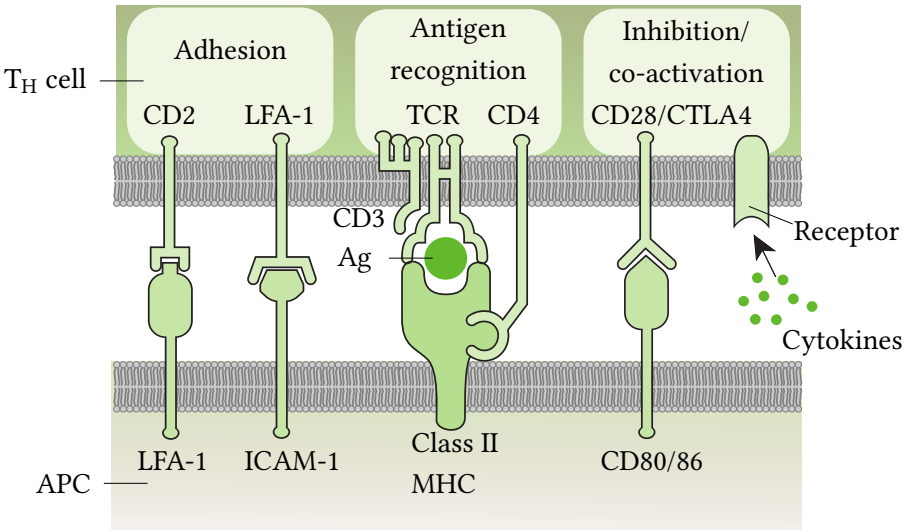


Figure 1.2: Activation of CD4+ T cells. CD4+ T cell TCR interaction with APC MHC II class molecule in complex with the antigen supported by the interaction of adhesion and inhibition/co-activation co-stimulatory molecules. Ag, Antigen; CTLA4, cytotoxic T lymphocyte A4; ICAM-1, intercellular adhesion molecule-1; LFA-1/3, lymphocyte function-associated antigen-1/3. Adapted from Murray et al. 2013.

Antigen recognition by TCR is stabilized by linking adhesion molecules on the APC and the T cell. This stabilization is needed because TCR and MHC II-antigen complex interaction often has low affinity. Adhesion molecules lengthen the interaction time, allowing for T cell activation. T cell CD2 and LFA-1 (lymphocyte function-associated antigen 1) adhesion receptors bond with the APC ligands LFA-3 and ICAM-1 (intercellular adhesion molecule-1), respectively. Activation also depends on the co-stimulatory signal mediated by the binding of CD80/86 ligand on the APCs to the CD28 receptor on the T cells (Murray et al. 2013). This activation is regulated by the inhibitory receptor CTLA-4 (cytotoxic T lymphocyte antigen 4), which directly antagonizes CD28 signaling by binding CD80/86 ligands with much greater affinity. However, when APCs express a sufficient amount of CD80/86 ligands, they can bind to both CTLA-4 and CD28 (Magee et al. 2012).

Naive CD4+ T cells can differentiate into at least four types of TH cells: T_H1, T_H2, T_H17, and regulatory T cells (T_{reg}). Differentiation into T_H2 cells requires an early source of IL-4, which can be provided by basophils (Sokol et al. 2008). Other potential sources include mast cells, eosinophils, natural killer T cells, and T cells. T cell differentiation is also regulated by T cell receptor Notch and APC ligand Jagged interaction, which promotes GATA3 (GATA-binding protein 3) expression. GATA3 is essential for T_H2 cell differentiation. Another regulating factor is the TCR signal strength. Weak TCR signaling favors T_H2 cell differentiation, whereas strong TCR signaling results in T_H1 cells (Galli, Grimbaldeston, et al. 2008; Paul and Zhu 2010).

Allergen-derived antigens induce a T_H2 cell phenotype. These CD4+ T cell effectors produce various cytokines, including IL-4, IL-5, IL-9, IL-13, and IL-25, which recruit and activate other immune cells at the inflammation site. In IgE-mediated allergy, the principal role of T_H2 effectors is the regulation of B cell class switch recombination through the production of IL-4 and IL-13 in conjunction with CD40-CD40L interactions (Paul and Zhu 2010). B cells subsequently produce both allergen-specific and nonspecific IgE. These IgE molecules diffuse locally into lymphatic vessels and enter the bloodstream (Galli, Tsai, and Piliponsky 2008).

Discovered in 1968 as the last of the five classes of human antibodies, IgE is now commonly associated with parasitic infestations and allergic diseases (Stanworth 1993). The two principal receptors for IgE are the high-affinity receptor FcεRI and the low-affinity receptor CD23 (also known as FcεRII). FcεRI is expressed as an $\alpha\beta\gamma2$ tetramer on mast cells and basophils, and as an

$\alpha\gamma 2$ trimer on APCs, monocytes, eosinophils, platelets, and smooth-muscle cells (Fig. 1.3a). CD23 is a C-type lectin receptor, which, in its membrane-bound form, has three lectin domains "heads" (Fig. 1.3b). Although considered a low-affinity IgE receptor, the avidity effect achieved by the three lectin receptors together can reach the high-affinity of the FcεRI receptor (Gould and Sutton 2008; Okayama et al. 2020). Interestingly, while Fig. 1.3b shows the schematic representation of membrane-bound CD23, this receptor can also be found in soluble monomeric or oligomeric forms. The coiled-coil stalk region of CD23 is susceptible not only to endogenous proteases, like ADAM10, but also exogenous, like the allergenic dust mite protease Der p 1 (Dhaliwal et al. 2017).

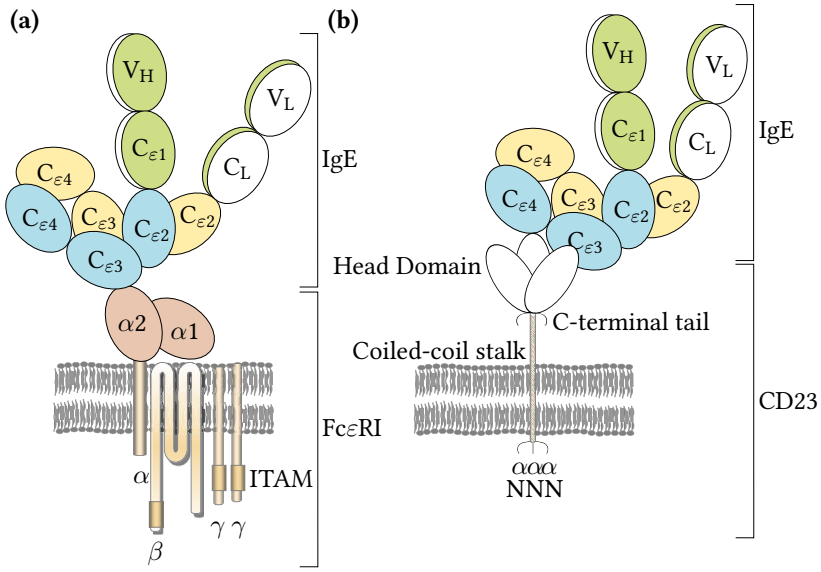


Figure 1.3: The structure of the **(a)** FcεRI (tetramer) α-chain and **(b)** CD23 receptors in complex with IgE. Monomeric IgE consists of two identical heavy chains (H- ε chains) and two identical light chains (L). IgE heavy ε-chains contain four Ig-like constant domains (Cε1-Cε4). IgE binds to the receptors through the $(C\varepsilon 3)_2$ domain of the Fc region. (a) Tetrameric FcεRI consists of the IgE binding α-chain and a β-chain and two γ-chains with immunoreceptor tyrosine-based activation motifs (ITAM) for signal transduction. (b) Membrane-bound CD23 is a trimeric type II membrane protein comprising α-helical coiled-coil stalk formed by three α-helices, topped by three C-type lectin "Head" domains. The C-terminal ends ("tails") extend extracellularly, while the N-terminal ends are cytoplasmic. Three N-glycosylation sites near the base of the coiled-coil stalk are also shown. Adapted from Okayama et al. 2020, figure modified with Inkscape (see 2.4).

B cells produce allergen-specific and nonspecific IgE that bind to FcεRI, CD23, and other receptors through the $(C\varepsilon 3)_2$ domain of its Fc region. Binding of allergen-specific IgE to FcεRI sensitizes mast cells and other effector cells, priming them to release mediators upon re-exposure to the target allergen (Gould and Sutton 2008; Galli and Tsai 2012). The sensitization process (Fig. 1.4) does not induce allergy symptoms; it only prepares the immune system to react upon subsequent encounters. The following step in the allergic response is the development of allergic inflammation, which is divided into three phases. The early-phase occurs within minutes of re-exposure to the allergen, and transitions after several hours into the late-phase reaction. Repeated allergen exposure can lead to chronic inflammation, resulting in structural changes at the sites of inflammation (Galli, Tsai, and Piliponsky 2008).

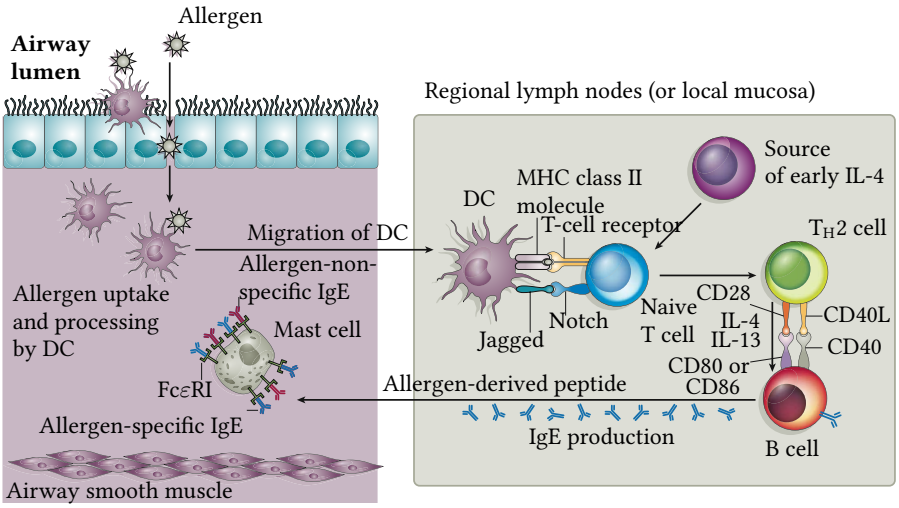


Figure 1.4: Sensitization to allergens. Allergens can enter the body via the respiratory tract, skin lesions, digestive system, or mucous membranes. This example illustrates respiratory tract sensitization. After crossing the epithelial barrier, allergens are captured by professional APCs, predominantly DCs, which migrate to peripheral lymphoid organs. There, DCs present allergen-derived peptides on MHC class II molecules to naive CD4+ T cells. In the presence of IL-4 and Notch receptor stimulation by DC-associated Jagged ligand, T cells differentiate into T_H2 cells. T_H2 cytokines IL-4, IL-5 and IL-13, and co-stimulatory signals (CD28 ligand with the CD80/CD86 receptor, CD40 ligand with the CD40 receptor) drive B cell class switching to IgE production. IgE enters the bloodstream via lymphatic vessels and binds to FcεRI on mast cells, sensitizing them to the specific allergen. Adapted from Galli, Tsai, and Piliponsky 2008.

1.2.2 Early- and late-phase reactions

Early-phase reactions begin with allergen-mediated cross-linking of allergen-specific IgE bound to the surface of mast cells via FcεRI. This type I immediate hypersensitivity reaction occurs within seconds to minutes of allergen re-exposure (Galli, Tsai, and Piliponsky 2008). Cross-linking of adjacent IgE molecules through multivalent or bivalent allergen epitopes induces FcεRI aggregation, initiating a phosphorylation-driven signaling cascade. Aggregation of tetrameric FcεRI on mast cells and basophils triggers anaphylactic degranulation – a rapid form of regulated secretion resulting in the coordinated release of three classes of biologically active molecules: 1) pre-formed mediators stored in cytoplasmic granules, 2) lipid-derived mediators, and 3) newly synthesized cytokines, chemokines, and growth factors (Fig. 1.5a) (Dvorak 2005). In contrast, aggregation of trimeric FcεRI on APCs and eosinophils leads to enhanced allergen uptake, antigen presentation to naive T cells, and up-regulated proinflammatory cytokine production (Fig. 1.5b) (Kraft and Kinet 2007).

Early-phase reactions are mainly mediated by the release of preformed mediators by mast cells and basophils, while the actions of APCs and eosinophils contribute to late-phase responses. Mast cells and basophils produce a similar, but distinct spectrum of proinflammatory mediators; however, allergic inflammation depends more heavily on mast cells, which is why it is the focus of this review (Gould and Sutton 2008).

Mast-cell cytoplasmic granules store diverse preformed mediators that drive early-phase allergic inflammation. These include vasoactive amines, proteoglycans, proteases, and cytokines/growth factors, each contributing to the initiation and amplification of the immune response.

Histamine is the most thoroughly characterized mediator modulating the immune response and inflammation among the biogenic vasoactive amines (Branco et al. 2018). Serglycin proteoglycans, bearing heparin- or chondroitin sulfate-type glycosaminoglycan (GAG) side chains, are essential for storing other preformed mediators in mast cell cytoplasmic granules and regulating the enzymatic response of mast cell-specific proteases. This function is thought to depend on the strong electrostatic interaction between the anionic GAG side chains and the basic charge of stored mediators (Rönnberg et al. 2012).

Proteases represent another major class of mast-cell mediators. Tryptases and chymases are serine protease family members, whereas carboxypeptidase A is a zinc-dependent metalloprotease (Pejler et al. 2007). Chymase recruits

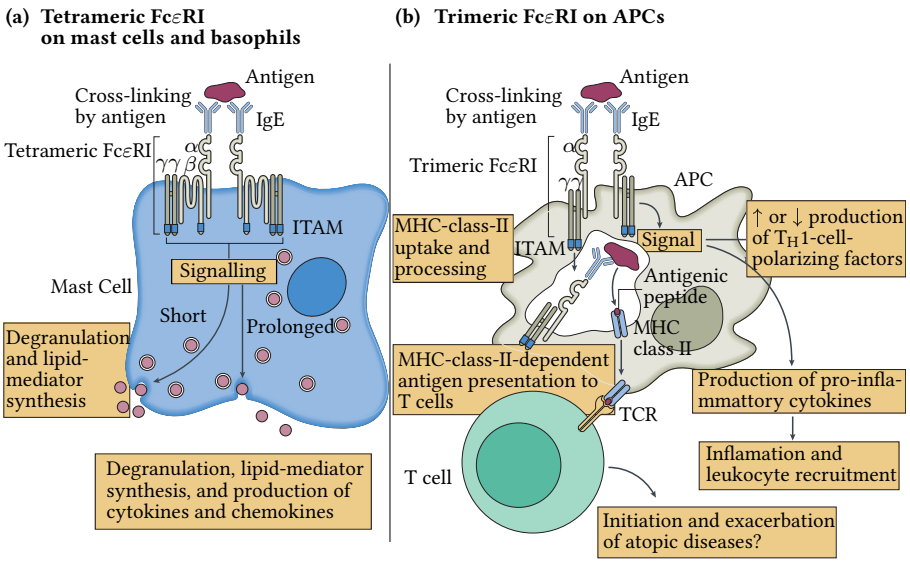


Figure 1.5: Signaling via tetrameric and trimeric FcεRI after antigen engagement. **(a)** Tetrameric FcεRI, expressed on mast cells and basophils, consists of an α-chain, β-chain and two γ-chains with ITAM for signaling. Cross-linking of adjacent IgE molecules bound to tetrameric FcεRI triggers, within seconds to minutes, a signaling cascade leading to the release of pre-formed mediators such as histamine, as well as lipid mediator synthesis. Prolonged allergen exposure promotes late-phase reactions through the production of cytokines and chemokines. **(b)** Trimeric FcεRI lacks the β-chain, and is found on APCs, eosinophils. Cross-linking induces uptake and processing of the FcεRI-IgE-bound antigen complex into the endocytic compartment. Antigen-derived peptides are loaded onto the MHC II class molecules and presented to T cells, activating their effector functions. FcεRI cross-linking also leads to the production of pro-inflammatory cytokines that activate other immune cells. Adapted from Kraft and Kinet 2007, figure modified with Inkscape (see 2.4).

inflammatory cells, including neutrophils and eosinophils, and promotes epithelial barrier permeability by proteolyzing occludin and fibronectin (Ebihara et al. 2005). Tryptase exhibits similar proinflammatory activities, although its precise mechanisms of action remain unclear (Pejler et al. 2007). Carboxypeptidase A participates in innate immunity, but its role in allergic inflammation is not well characterized (Atiakshin et al. 2022).

Mast cell cytoplasmic granules also contain preformed cytokines and growth factors, which have immunomodulatory (e.g., influencing the biological functions of immune cells) and effector (e.g., regulating the phenotypic output

of structural cells) functions. Notable examples include the tumor-necrosis factor α (TNF-α), the CC-chemokine ligand 2 (CCL2), the vascular endothelial growth factor A (VEGFA), and multiple interleukins (Galli, Kalesnikoff, et al. 2005). IL-4 and IL-13 are hallmark cytokines, regulating the allergic response through T_H2 cell differentiation, B-cell class switch recombination, mast cell proliferation, and enhanced cytokine production (McLeod et al. 2015). Another example would be the synergetic effect of TNF-α and IL-8 in the recruitment of neutrophils to the inflammation site (Salamon et al. 2005).

Mast cell degranulation also initiates the *de novo* biosynthesis of lipid mediators through arachidonic acid metabolism via the cyclooxygenase and lipoxygenase pathways. These newly released eicosanoids include leukotriene LTC_4 and LTB_4, prostaglandin PGD_2 and PG_2, as well as the precursor of the cysteinyl leukotrienes (cysLTs). Eicosanoids influence allergic inflammation in ways similar to preformed cytokines and growth factors: they can initiate, amplify, or attenuate the reaction and modulate its duration, magnitude, and nature of the response. These effects are mediated through specific G-protein-coupled receptors, expressed by bronchial smooth muscle cells, leukocytes, neutrophils, basophils, macrophages, mast cells, and other immune and structural cells (Boyce 2007).

The broad spectrum of released preformed and *de novo* synthesized mediators contributes to the acute signs and symptoms associated with the early-phase reactions of the allergic inflammation response (Fig. 1.6). The clinical manifestations vary according to the reaction site. However, they can be characterized by tissue-specific effects, including vasodilation, which is accompanied by local microvascular leakage, which is a key feature of allergic diseases, resulting in increased plasma protein and neutrophil migration (Owen-Woods et al. 2020). Vasodilation is readily observed during skin-prick testing in individuals predisposed to allergy, where allergen injection induces a wheal-and-flare response (Barnes 2011). Vasodilation is also followed by increased vascular permeability, leading to tissue swelling and tears in the eyes. Bronchoconstriction is another common manifestation, with airflow obstruction and wheezing ranging from mild discomfort to life-threatening suffocation (Galli, Tsai, and Piliponsky 2008). Additional symptoms include increased mucus production (runny nose, aggravation of airflow obstruction in the lower airways) and the stimulation of a subpopulation of nociceptors of the sensory nerves (itching, coughing) with the initiation of the peptidergic pathway (sneezing) (F Li et al. 2021).

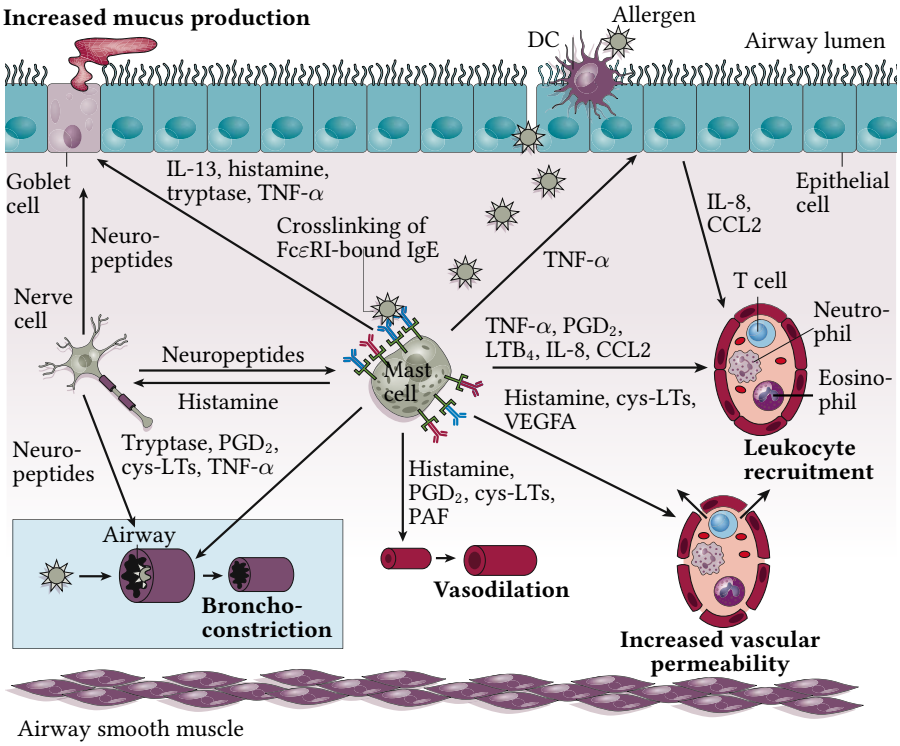


Figure 1.6: Mechanisms of the early-phase allergic reaction (e.g., airway inflammation). Tetrameric FcεRI on the mast cell surface are bound with allergen specific and non-specific IgE. Upon allergen recognition, FcεRI aggregates and induces degranulation. Pre-formed mediators (histamine, tryptase, chymase, TNF-α, VEGFA, CCL2, IL-8 and other) resident in the cytoplasmic granules of mast cells are released. Degranulation also induces the *de novo* biosynthesis of lipid-mediators (PGD_2, PG_2, LTC_4, LTB_4, and PAF). The combined action of these pro-inflammatory mediators causes bronchoconstriction, vasodilation, increased vascular permeability, and mucus hypersecretion. Recruitment and activation of other immune cells, particularly leukocytes, contributes to progression toward the late-phase reaction. Adapted from Galli, Tsai, and Piliponsky 2008.

Mast cells upon FcεRI crosslinking will not only release mediators through degranulation but also begin slowly synthesizing new cytokines, chemokines, and growth factors. These newly produced molecules contribute to the late-phase reaction alongside the actions of mast cell recruited leukocytes, neutrophils, eosinophils, and other immune cells (Galli, Tsai, and Piliponsky 2008). Allergic inflammation is further amplified by the recruitment of inflammatory cells, the crosslinking of tetrameric and trimeric FcεRI, and the induction of

T_H2-type effector responses (Paul and Zhu 2010). These processes collectively constitute the late-phase reaction, which typically develops within 2-6 hours of allergen exposure and peaks around 6-9 hours. There is no clear distinction between the two phases; differences lie in the variety and quantity of cells involved, the observed effects of prolonged stimulation by inflammatory mediators, and the severity of specific symptoms (e.g., increased mucus production, epithelial damage, spreading of hives). Late-phase reactions can resolve without treatment in the absence of the offending allergen. For example, this occurs at the end of the pollen season for pollen-sensitive individuals. However, apart from this, the mechanisms that underpin the resolution of allergic inflammation are not fully understood (Galli, Tsai, and Piliponsky 2008; Valenta, Hochwallner, et al. 2015). A more well-studied area is what happens when there is a persistent, repetitive allergen exposure.

1.2.3 Chronic allergic inflammation

Continuous exposure to an allergen leads to chronic allergic inflammation. Persistent inflammation in the affected area leads to structural changes and, in many cases, to altered functions of the affected organs. Immune cell homing remodels the local cell environment: IL-9 promotes recruitment and proliferation of FcεRI-expressing mast cells, leading to mast-cell hyperplasia. High densities of allergen-sensitized mast cells perpetuate inflammation by releasing preformed and *de novo* synthesized mediators. IL-5-mediated eosinophil recruitment results in tissue eosinophilia, while APC recruitment via cytokines enhances FcεRI expression and up-regulation of co-stimulatory molecules (CD40, CD134, CD80), promoting T_H2-type differentiation and sensitization to additional allergens or epitopes. These processes reinforce T_H2-type effector functions in a cycle of constant amplification (Galli, Tsai, and Piliponsky 2008; Paul and Zhu 2010; Barnes 2011; McLeod et al. 2015).

T_H2 cytokines also stimulate mucus hypersecretion, goblet-cell metaplasia, and airway hyperresponsiveness through IL-4, IL-9, and IL-13 acting on epithelial cells, and IL-4/IL-13 acting on smooth-muscle cells. Additional T_H2-effector functions include activation of T cells (IL-4, IL-9, IL-25), macrophages (IL-4, IL-13), and B cells (IL-25) (Paul and Zhu 2010).

Interactions between resident structural cells and infiltrating inflammatory cells drive site-specific tissue changes in chronic allergic disease. In the airways, for example, formation of the epithelial-mesenchymal trophic unit is a hallmark of chronic asthma and sustains T_H2-driven inflammation (Galli, Tsai, and

Piliponsky 2008). In other IgE-mediated disorders, chronic inflammation can lead to nasal polyps in allergic rhinitis, or epithelial damage in AD, which increases susceptibility to *Staphylococcus aureus* infection (Ogulur et al. 2025).

A major concern in severe or chronic allergy is the heightened risk of developing new allergic sensitizations, a progression known as the atopic march (or allergic march). Persistent allergic inflammation damages the epithelial barrier, facilitating access of pathogens, pollutants, and allergens to immune cells (Fig. 1.7). This increased exposure promotes presentation of novel epitopes to APCs, driving epitope or molecular spreading (more information in 1.4.3). Spreading can occur within the same allergen when homologous epitopes exist in related allergens, or via the low-affinity receptor CD23, which can bind and present structurally unrelated allergens (Gould and Sutton 2008). Population-based studies further support a positive correlation between epithelial barrier dysfunction and an increased risk of food allergy. For instance, infants with AD at 3 months of age are estimated to have a sixfold higher risk of developing food allergies compared to healthy infants (Tsakok et al. 2016).

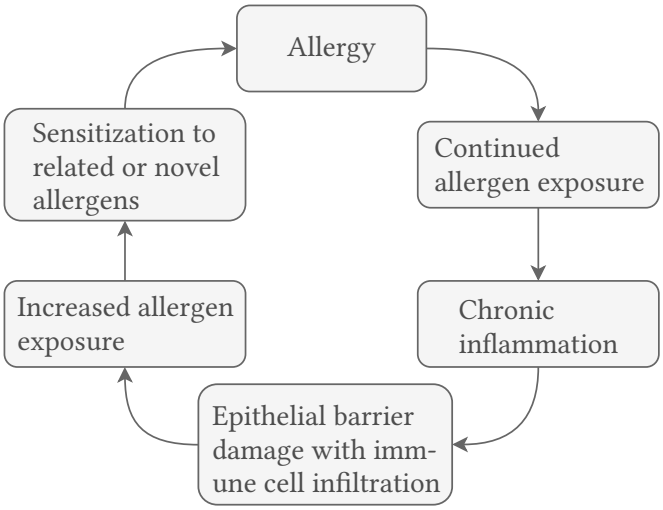


Figure 1.7: Atopic march – a cycle of epitope spreading. The graph shows a proposed atopic march model for atopic dermatitis. Drawn with Diagrams.net (see 2.4).

1.3. Epidemiology of allergies

Allergic diseases are a major cause of global morbidity and an increasing burden on health and medical systems of both developed and emerging countries. According to the World Allergy Organization (WAO), the sensitization

rates worldwide to one or more common allergens among school children are approaching 40-50% (Pawankar et al. 2013). While this comprehensive global estimate remains widely cited, more recent investigations have focused on national or allergy-specific prevalence (Spolidoro et al. 2023). Updated population-based studies indicate that although the individual risk of developing allergic conditions such as asthma and AD may not be increasing, the total number of cases worldwide has risen (Fig. 1.8). This increase is mainly attributed to global population growth rather than a rise in per-capita incidence (Shin et al. 2023).

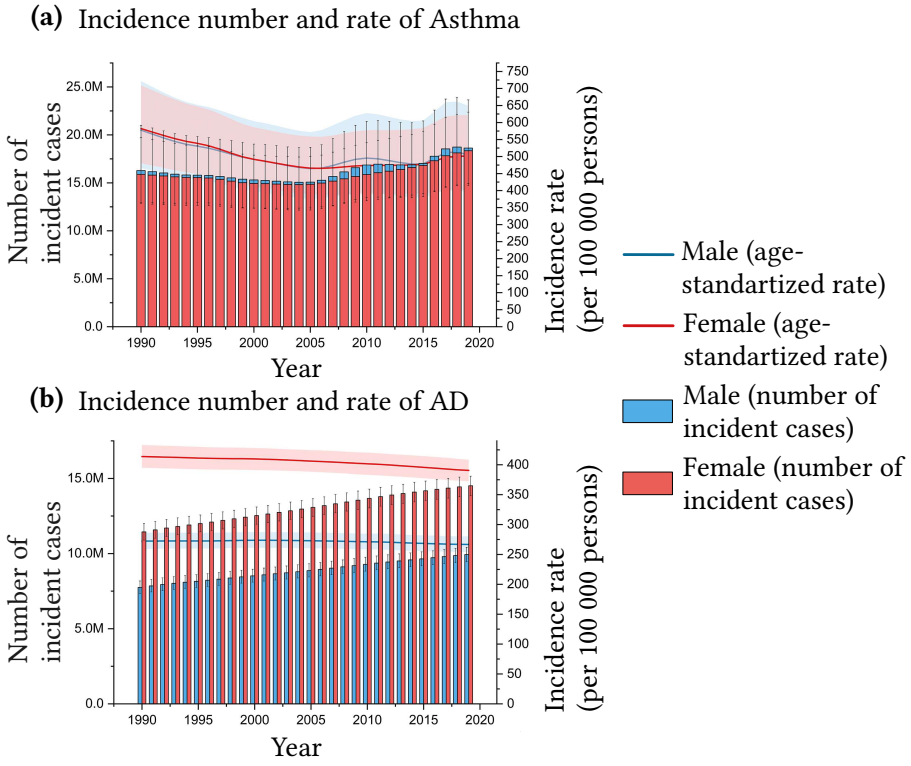


Figure 1.8: Total number of incident cases and global incidence rates of asthma **(a)** and atopic dermatitis **(b)** by sex from 1990 to 2019. Adapted from Shin et al. 2023, figure modified with Inkscape (see 2.4).

Nevertheless, this growing absolute burden places increasing pressure on healthcare systems and highlights the need for improved allergy prevention and management strategies. The burden of allergic diseases is not uniformly distributed – substantial variation exists across age groups, socio-economic status, and geographic regions. For instance, higher prevalence rates are gen-

erally seen in countries with a high socio-demographic index, likely reflecting better reporting and greater exposure to urbanized environments. In contrast, low- and middle-income countries, while reporting lower prevalence, experience disproportionately worse outcomes, including higher mortality and fewer treatment options (Shin et al. 2023).

In Europe, the European Academy of Allergology and Clinical Immunology (EAACI) reports that approximately 30% of the population experiences allergy symptoms, with some countries showing even higher rates – Germany, for example, reports a prevalence of 41% (Zuberbier et al. 2014; Mazur et al. 2022). Recognizing the global impact, the World Health Organization (WHO) has designated allergy as the fourth leading cause of immune disease in its 11th revision of the International Classification of Diseases (ICD), projecting that by 2050, half of the world's population will be affected (Tanno and Demoly 2022).

Despite this being a major health issue, allergic diseases remain underrecognized by the public and health institutions. This results in delayed diagnoses, insufficient resource allocation, and limited awareness of the consequences of untreated allergic inflammation, especially in vulnerable populations (Sánchez-Borges et al. 2018). For example, food allergy is a leading cause of life-threatening allergic reactions such as anaphylaxis. Anaphylaxis is a systemic hypersensitivity reaction characterized by the rapid onset of symptoms, which, if not promptly treated with intramuscular epinephrine and followed by emergency care, can be fatal (Matricardi et al. 2016). In the United Kingdom (UK), an analysis of hospital records from 1998 to 2018 identified 30 700 admissions for food-triggered anaphylaxis, revealing a concerning 5.7% annual increase in hospitalizations for this condition (Baseggio Conrado et al. 2021; Motosue et al. 2022).

One of the most compelling arguments for stronger public awareness comes from a high-profile case in the UK. Natasha Ednan-Laperouse, severely allergic to sesame, consumed a prepackaged sandwich that did not disclose the presence of sesame seeds. Without access to epinephrine, she suffered a fatal reaction. This tragedy led to the introduction of *Natasha's law* in 2021, requiring all food outlets to provide complete ingredient lists with clear allergen labeling on prepacked food intended for direct sales (Natasha allergy research foundation).

Despite such legislative advances, no Food and Drug Administration (FDA)-approved safe or effective methods for preventing or curing food allergies

are currently available. Avoidance of trigger foods remains the only viable option, yet this is especially challenging when the most common allergens in the USA are peanuts, cow's milk, and eggs, which are widely used. This creates an unsafe environment, greatly diminishing the quality of life (Soon and Abdul Wahab 2021). The risk is not limited to ingestion: severe, even fatal, allergic reactions have been reported from allergens transmitted through saliva (Steensma 2003; McKibbin et al. 2021). These realities highlight the importance of warning atopic patients about such risks, as epinephrine auto-injectors are prescription-only and unavailable in standard first-aid kits. Preventing future tragedies will require a multifaceted approach combining proactive legislation, widespread public education, and improved accessibility to diagnosis and treatment.

The trivialization of allergic disorders is evident across all parts of society, from education to healthcare services. An epidemiological survey on the treatment of rhinitis symptoms in Europe reported that 52.6% of patients had not visited a medical professional for their symptoms, and 30.2% opted for nonprescription medicine as it does not require visiting the healthcare facility (Maurer and Zuberbier 2007). Allergic rhinitis median prevalence world-wide is estimated to range from 18.1% to 28.5%, with studies indicating a rising prevalence in the general population (Licari et al. 2023). The undertreatment of allergic disorders has direct and indirect socio-economic consequences. It was inferred that the absence from work or reduced work productivity due to allergies can result in between €55 to €151 billion annually of avoidable costs for the European Union (Zuberbier et al. 2014). Beyond economic impact and reduced quality of life, allergic disorders have also been linked to impaired cognitive performance in children, further underscoring the importance of adequate diagnosis and management.

Several large-scale studies have demonstrated that allergic symptoms, particularly during pollen season, can measurably impair academic performance. In Norway, fluctuations in pollen count were associated with lower test scores among allergic students compared to their non-allergic peers, with a one-standard-deviation increase in pollen linked to a 2.5% reduction in scores relative to the mean (Bensnes 2016). In the United States, a similar rise in ambient pollen was associated with declines in proficiency rates of 0.2-0.3 standard deviations in English language exams and 0.3-0.4 standard deviations in mathematics assessments (Marcotte 2015). A UK study found that 40% of students experiencing allergic rhinitis symptoms on examination day were more

likely to drop a grade between winter and summer assessments, a likelihood that increased to 70% among those taking sedating antihistamines (S Walker et al. 2007). The caveat of such large studies is the inherent impossibility of ruling out all competing explanations for these results. However, while it is important to acknowledge the possible bias, the consistency of these findings across populations underscores the need for greater attention to the societal impacts of allergic disease. This includes further research into underlying mechanisms, pathophysiology, improved diagnostics, effective treatments, and preventative strategies across all allergy types.

1.4. Risk factors for allergic disease

Increasing attention is being devoted to research on risk factors and preventive measures to mitigate the rising prevalence of allergies. Existing research has recognized that heritable factors can be involved in allergy development: if one parent is allergic, a child has a 33% chance of developing allergies, which rises to 70% if both parents are allergic (Steinke and Borish 2006). However, such estimates are not fixed and can vary from 15% to 82% when factoring in allergy phenotypes, severity, onset start and the possibility of environmental factors (X Hong et al. 2015). For example, food allergy is most common in infants and young children, pattern likely driven more by an immature gut microbiota and immune system rather than a genetic predisposition (Lee et al. 2020). Therefore, while researchers have agreed that genetics play a role in allergy development, environmental factors are no less important.

1.4.1 Genetic factors

Twin studies support the genetic basis of allergy. For example, in a cohort of 58 twin pairs with a family history of peanut allergy, concordance was significantly higher in monozygotic twins (64.3%) than in dizygotic twins (6.8%), corresponding to an estimated heritability of 81.6% (Sicherer et al. 2000). Similar trends were observed in other twin cohorts: peanut allergy was diagnosed in 59% of monozygotic twins and 29% of dizygotic twins in a 60-pair study (Kivistö et al. 2019), and in 53% versus 29%, respectively, in a broader study of 826 twin pairs aged 12-18 years (X Liu et al. 2009). These findings consistently indicate a substantial heritable component in allergic disease.

Risk factors are identified through genome-wide association studies (GWAS) or a candidate-gene approach. GWAS rapidly scans large numbers of genetic markers across entire genomes to detect disease-associated variants. In con-

trast, the candidate-gene approach is a hypothesis-based study in which genes are identified based on their potential role in the mechanism of the disease. Both methods have reported multiple candidate genes; however, due to the small sample sizes and high error rates, considerable uncertainty is often associated with the results (X Hong et al. 2015). Other approaches include family studies, analyses of monogenic disorders in which food allergy is the primary clinical manifestation, and animal models. However, these methods provide additional insights but are similarly limited by small cohorts, rarity of phenotypes, and translational constraints (Arnau-Soler et al. 2025).

Among the most consistently replicated candidates is the filaggrin gene (FLG), which encodes a key structural protein of the epidermal barrier. Loss-of-function mutations in FLG have been associated with an increased risk of food allergy, likely through impaired skin barrier integrity that facilitates allergen penetration; higher allergen concentrations do not positively correlate with higher sensitization rates in individuals with wildtype FLG (Stefanovic and Irvine 2024; Arnau-Soler et al. 2025). However, other studies report that FLG mutations do not increase the risk of food allergy or asthma when occurring independently of AD, suggesting that the link may be mediated by skin inflammation rather than the genetic variant alone (Lukacs and Hogan 2025). This highlights the importance of gene-environment and gene-phenotype interactions in developing allergic disease.

Another GWAS study provided convincing evidence that gene regions encoding human leukocyte antigen (HLA)-DR and -DQ isotypes confer risk for peanut allergy in subjects of European ancestry. HLA-DR is an MHC class II cell surface receptor encoded by the HLA complex. HLA-DQ is a cell surface receptor protein found on APCs. Both molecules are involved in antigen presentation and the development of antigen-mediated immune response. Several peanut allergy-associated single-nucleotide polymorphisms (SNPs) have been elucidated in those regions (X Hong et al. 2015).

A GWAS meta-analysis proved that the frequent co-occurrence of asthma, allergic rhinitis, and eczema is partly attributable to a shared genetic origin. In this study, 136 independent risk variants were identified that dysregulate the expression of immune-related genes. Further analysis elucidated that these target genes are involved in T and B cell activation, proliferation, isotype switching, cytokine production, and other pathways central to allergic inflammation (Ferreira et al. 2017). Most risk variants identified by such studies will be of no clinical relevance as they cannot predict allergic sensitization

or are involved in a minimal capacity in the pathogenesis mechanism. However, some risk factors have been identified to cause at least 25% of allergic sensitizations in a population, suggesting that targeting them would have a significant impact on public health (Bønnelykke et al. 2015).

Most risk variants are identified from the pathways which are involved in T_H2 immunity, T-cell differentiation, TGF-β signaling, T_{reg} function, and skin or mucosal barrier integrity. That is, from genes involved in antigen recognition, presentation to inflammatory response regulators (Bønnelykke et al. 2015; Arnau-Soler et al. 2025). Interestingly, GWAS studies also discovered that quite a few identified candidate genes have a nearby CpG site. This may suggest that environmental factors can influence methylation levels on these CpG sites, affecting candidate gene expression and, by extension, influencing allergy risk. This would also shed light on the dramatic rise of the prevalence of allergies if the cause were epigenetic factors, likely influenced by environmental and lifestyle changes (Xu et al. 2021; Arnau-Soler et al. 2025). An example of this would be the significant association between smoking and the methylation site of PITPNM2 gene, encoding a PYK2-binding protein, which is involved in neutrophil function (Ferreira et al. 2017). This interaction may explain the link between maternal smoking in pregnancy and the increasing risk of early-onset persistent asthma (London et al. 2001).

Other epigenetic factors would include posttranslational histone modifications, such as acetylation, phosphorylation, methylation, ubiquitination, etc., which may directly impact the expression levels of candidate genes, affecting the development of allergic diseases. Investigating genetic and epigenetic factors may provide new ideas into allergy prognosis, regulation, and inhibition (Alaskhar Alhamwe et al. 2018).

1.4.2 Environmental factors

Rapid urbanization and industrialization have been linked with the growing prevalence of allergic diseases through two main mechanisms: (i) greater exposure to environmental pollutants along with altered dietary patterns, which can exacerbate disease development; and (ii) reduced exposure to diverse childhood infections, as described by the "Hygiene hypothesis" (Pawankar et al. 2013; Q Zhang et al. 2025).

The causal link between air pollution, climate change, and allergic diseases remains the subject of extensive scientific debate. Ambient air pollution predominantly comprises particulate matter and gaseous pollutants. Particulate

pollutants include diesel exhaust particles (DEPs), one of the most well-studied forms, as well as particulate matter (PM) with aerodynamic diameters smaller than 2.5 μm ($PM_{2.5}$), or 10 μm (PM_{10}). $PM_{2.5}$ includes organic compounds, metallic particles, and combustion particles, whereas PM_{10} includes mold, dust, and pollen. Gaseous pollutants mainly comprise CO, O_3, NO_2, SO_2, and oxidants (Lu et al. 2021).

Most airborne pollutants act as mucosal adjuvants, interacting with innate and adaptive immune cells. These pollutants drive the antigen-mediated immune response towards a T_H2-type inflammation in about 11.1-40% of atopic individuals, depending on the country (Lu et al. 2021). Exposure to DEPs increases inflammatory cell infiltration, cytokine production, and chemokine release (Takizawa 2004). Gaseous pollutants, including O_3, NO_2, and SO_2, damage the ocular surface by acidifying tears, oxidizing tissues, and facilitating allergen penetration; O_3 and NO_2 can also modify aeroallergens, triggering cytokine release and aggravating allergic sensitization (Lu et al. 2021). Damage to the airway mucosa, impaired mucociliary clearance, and airway inflammation increase hyper-responsiveness, allowing greater immune access to inhaled allergens and thereby promoting sensitization or exacerbating disease (D'amato et al. 2010). By contrast, low concentrations of pollutants do not appear to enhance the inflammatory response, as was demonstrated in a controlled study investigating the effect of ozone exposure on late inflammatory and early bronchoconstrictor responses to inhaled house dust mite allergen in sensitized individuals (Chen et al. 2004).

PM_{10} particles represent another important aspect of unintended air pollution. 40% of Europeans are affected by pollen allergies, and their prevalence continues to rise. Hay fever typically lasts from late March to September, and symptom severity correlates with ambient pollen concentrations (Stevanovic et al. 2025). The EU promotes green initiatives through the *EU Green Infrastructure Strategy* (CELEX 52013DC0249) and the *EU Biodiversity Strategy for 2030* (CELEX 52020DC0380) initiatives. Urban planning in major cities ensures residents live within walking distance of a park or natural area. In 2025, Vilnius was named the European Green Capital, recognizing its commitment to climate action, sustainability, and smart urban planning. Around 500 000 trees can be inventoried in the capital. According to the inventory, maples (86 000), lindens (25 000), birches (33 000), and pines (32 000) account for the largest number of trees in Vilnius (www.zaliasvilnius.lt). However, the green initiatives do not account for allergies when landscaping projects are being

implemented. A notable example is the birch tree, which produces pollen containing the major allergen Bet v 1. Pollen from birch and related trees is a major cause of allergic rhinitis and asthma symptoms (Stevanovic et al. 2025).

Air pollution, when combined with rising global temperatures, can exert an even greater impact on public health, particularly in individuals with respiratory diseases. Weather changes have long been linked to asthma outbreaks. For the past century, researchers have studied the strange connection between reports of asthma epidemics following thunderstorms. Such cases have been reported in the UK, Europe, Australia, Canada, and Mexico. It was hypothesized and later confirmed that pollen grains break up under humid conditions due to osmotic shock and release part of their contents, including allergen-carrying starch granules, into the atmosphere. These 0.5-2.5 μm starch granules can reach the lower airways and induce asthma reactions in individuals allergic to pollen (Taylor and Jonsson 2004; D'Amato et al. 2007; J Hong et al. 2016; Lu et al. 2021).

Other environmental factors, such as moving to another country, can expose an individual to a new set of pollutants and allergens. Socioeconomic factors also influence allergy prevalence, as individuals of lower status are more likely to be exposed to sub-optimal living conditions, such as mold, dust, or unhealthy dietary habits. Another spectrum of environmental factors is associated with occupational allergies, which arise from high-concentration exposures in the workplace (e.g., latex in health-care workers and laboratory technicians; flavoring, oil, or antioxidants in bakers; dyes, latex, rubber, and fragrances in hairdressers) (Pawankar et al. 2013; Stevanovic et al. 2025).

Epidemiological studies have shown that dietary habits, which include foods high in advanced glycation end-products (AGEs), positively correlate with increased risk of food allergy. AGEs are generated during thermal food processing, such as grilling meats, frying foods, or caramelizing products. Population-based studies have confirmed a causal link, and mouse experiments demonstrate that an AGE-rich diet can induce an adjuvant-like effect on allergens, disrupt intestinal barrier integrity, and promote gut dysbiosis, thereby facilitating immune interactions with bacterial products or allergens and perpetuating intestinal allergic inflammation. Long-term AGE consumption also promotes sensitization to unrelated allergens, providing a clear example of an atopic march (see 1.2.3) (Q Zhang et al. 2025).

Nevertheless, environmental factors can also influence allergy development in a contrary manner. This phenomenon is known as the "Hygiene

hypothesis". It proposes that the rising prevalence of allergic diseases is linked to reduced exposure to infections, resulting from urbanization and improved healthcare in Western countries (Okada et al. 2010). The concept dates back to the 1870s; however, the term "Hygiene hypothesis" was introduced in 1989 by David P. Strachan, who after observing 17 000 British children born in 1958, noted an inverse correlation between hay fever and the number of older siblings in the family (Strachan 1989). Another British survey of 29 238 children corroborated the findings, concluding that the presence of older siblings or being an offspring of a multiple birth (twins, triplets, and so on) will have a protective effect against developing eczema, hay fever, and asthma diagnosed after the age of two (Mckeever et al. 2001). A similar protective effect was observed with early exposure to older children in daycare settings (Homas et al. 2000).

The hypothesis explained that some infectious agents can have a protective effect against allergic sensitization. The apparent differences in the prevalence of allergies in rural and urban territories also correlated with these findings. In Ethiopia, for example, asthma was significantly less common in rural villages than in urban areas. Studies in Germany found that farmers' children living in Bavarian counties had a lower prevalence of hay fever than peers living in non-agricultural areas of Bavaria (Yemaneberhan et al. 1997; Von Ehrenstein et al. 2000). A Canadian longitudinal study similarly reported a reduced risk of developing asthma when living in an agricultural environment (Midodzi et al. 2007). More recent studies elucidated that the protective effect may be associated with rich home dust microbiota. Although the mechanism is not fully understood, it may involve a higher proportion of environmental bacteria, including those of animal origin, compared with human-associated bacteria (Kirjavainen et al. 2019). On the other hand, pet ownership early in childhood is associated with a higher prevalence of allergies. However, the evidence remains inconclusive as more longitudinal studies are required (Pyrhönen et al. 2015).

Increasing evidence indicates that changes in human microbiota composition and function are associated with allergy development (Lee et al. 2020). In newborns, reduced microbiome diversity has been linked to increased allergy risk through indirect evidence, including mode of delivery, infant feeding practices, maternal and infant dietary fiber intake, childhood exposure to agricultural environments, and findings from animal models (Lambrecht and Hammad 2017). However, such associations have been challenged by recent

longitudinal studies from Japan; one study of 2 114 children found no link between cesarean delivery and allergic diseases up to nine years of age. These findings may not be directly applicable to Western countries due to cultural and dietary differences, but they underscore the need for further research in allergology (Tamai et al. 2025).

1.4.3 Allergen molecular characteristics

Allergens are a heterogeneous group of molecules, typically proteins, glycoproteins, lipoproteins, or protein-conjugate chemicals and drugs, that elicit an IgE-mediated immune response (Costa et al. 2022). Considering the diversity of allergic diseases, only a very small amount of the proteome acts as an allergen. According to the AllFam database (www.meduniwien.ac.at/allfam/), allergens are currently classified into 151 protein families (accessed on 2025-08-18). The major allergens can be further grouped into eight allergen superfamilies: prolamin, EF-hand, profilin-like, tropomyosin-like, cupin, Bet v 1-like, calycin, and double-psi beta barrel. Allergens in these superfamilies share high sequence identity or structure similarities, which could induce cross-reactive responses in an atopic individual exposed to homologous allergens. However, this is not universal, as allergens differ not only in their molecular characteristics but also in their ability to induce sensitization (sensitizing vs. non-sensitizing allergens), their capacity to initiate IgE responses (initiator vs. secondary allergens), and in their recognition by IgE antibodies (major vs. minor allergens) (Matricardi et al. 2016).

Longitudinal studies have elucidated a few patterns of allergy development. In dust mite allergy, for example, individuals often sensitize to one of the major allergen groups (Der p 1/Der f 1 or Der p 2/Der f 2) and remain sensitized to that group throughout childhood, or they become co-sensitized to both groups without further progession. Sensitization patterns can evolve with continued exposure to the allergen source in other allergies, such as timothy grass. Diversification of the IgE reactivity profile over time, termed "molecular spreading", has been observed in timothy grass allergy: individuals typically first sensitize to the initiator allergens Phl p 1 and Phl p 5, only then to Phl p 2, 4, and 6, and at later times even to Phl p 7, 11, and 12 (Custovic et al. 2015). It is clear that not all antigens that pass through the epithelial barrier can induce an IgE response, and the induced reaction is not universal. Some features are a consequence of an extrinsic factor, such as antigen presentation in the presence of bacterial cell wall components, whereas the role of intrinsic

allergen features remains uncertain (Matricardi et al. 2016).

Numerous studies have sought to determine what defines an allergen. However, the results remain inconclusive. Even the most basic questions, such as whether allergenic potential can be predicted from protein abundance remain unclear. For example, caseins, the most abundant proteins in milk, are considered major allergens in milk allergy and the whole casein family ranks fourth in allergenic potential upon ingestion within the food allergy category. Conversely, allergens present in limited quantities, such as tropomyosins, parvalbumins, or serum albumin, also can cause severe allergic reactions. This may indicate that the limited quantity can also be a factor in the severity of an allergic reaction. Similarly, no single parameter, whether it is structural stability, ligand binding, glycation, aggregation, heat stability, pressure lability, or lipid interactions, can reliably predict allergenicity (Costa et al. 2022).

Despite these ambiguities, some studies have identified direct links between protein function and allergenic potential. Pore-forming proteins from various species can trigger an allergic reaction through two mechanisms: one causes an influx of Ca^{2+}, which activates the type 2 immune response, the other triggers the alarmin IL-33 release, which results in an enhanced allergic response (Fig. 1.9) (Shi et al. 2025).

It is speculated that a common feature of type 2 response inducers appears to be their ability to cause different types of tissue perforations, which would make them immunogenic. However, while all allergens are antigenic, not all allergens are immunogenic. Therefore, allergens can be categorized into two different classes. Class A allergens are inherently capable of triggering immune responses on their own. They are often enzymatic proteins, such as proteases or lipases, simultaneously acting as antigens and immunogenic stimuli. Class B allergens lack intrinsic immunogenicity and require the presence or co-localization of adjuvants in space or time to become immunologically active. Adjuvants can be Class A allergens or any other substance capable of inducing a type 2 response. Hence, Class B allergens are also aptly named "guilty-by-association" (Kopp et al. 2023).

Class A allergens are thought to initiate the T_H2 response through three different modes: (i) retrospective detection of tissue damages, in which DCs respond to site of injury by recognizing stress- or damage-associated molecules (e.g., alarmins, ATP, uric acid); (ii) prospective detection via sensory cells, in which specialized epithelial or neuronal cells sense immunogenic activity and, in turn, release cytokines or neuropeptides to activate DCs; and (iii)

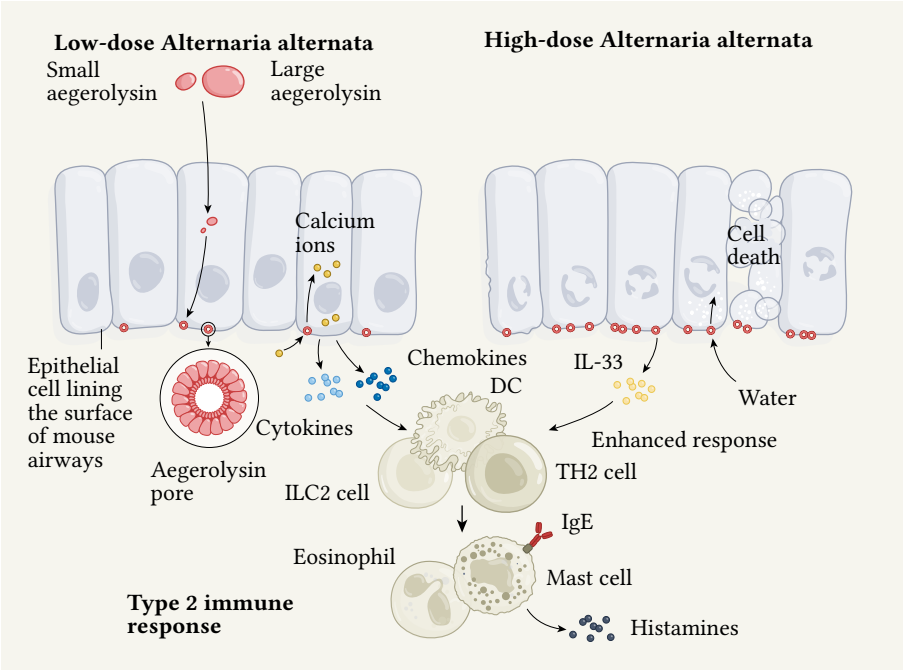


Figure 1.9: Allergy-triggering mechanisms of the common mold *Alternaria alternata* in mouse airways. The principal immunostimulatory components are large and small aegerolysin proteins, which assemble into transmembrane pores. Although the precise epithelial site of pore formation is unknown, aegerolysins can elicit allergic responses through two distinct mechanisms. Under low-dose exposure to *A. alternata* extract, Ca^{2+} influx through the pores activates the MAPK pathway and downstream inflammatory gene expression. The resulting release of cytokines and chemokines attracts and activates dendritic cells, type 2 innate lymphoid cells (ILC2s), and T_H2 cells, driving a type 2 response mediated by eosinophils and mast cells. In contrast, high-dose exposure damages epithelial cells, stimulating the release of the alarmin IL-33, which amplifies the allergic response, while excessive pore formation can also cause osmotic lysis of epithelial cells. Adapted from Lambrecht 2025, which is based on Shi et al. 2025.

direct detection, in which DCs directly recognize adjuvant activity from low-doses of bacterial lipopolysaccharides, peanut extract, or house dust mite allergens, leading to their maturation and the initiation of a T_H2 response. Aegerolysins, shown in Fig. 1.9, can be categorized as Class A allergens, and the study proposed two allergy mechanisms. When aegerolysin concentrations are low, DC activation relies on prospective cues from epithelial or sensory cells. However, the resulting tissue damage at higher concentrations releases

alarmins that drive retrospective sensing (Shi et al. 2025; Kopp et al. 2023).

While functional properties of Class A allergens illustrate how allergens can actively initiate a T_H2 response, structural and biochemical modifications also play a decisive role in shaping allergenic potential. PTMs, glycosylation in particular, can influence IgE binding and cross-reactivity (Altmann 2016). Protein glycosylation can be ambiguous in allergy diagnostics. On the one hand, glycosylation can be critical for the folding and activity of the recombinant allergen (Bonnet et al. 2018). On the other hand, glycosylation can elicit false positive reactions by reacting with non-specific IgE (Kozlov et al. 2025).

IgE-binding N-glycans, known as cross-reactive carbohydrate determinants (CCDs), are classified into several groups, some with proven clinical relevance and others still under investigation (Fig. 1.10). Group A carbohydrates, or 'classical CCDs,' are found on horseradish peroxidase, pineapple bromelain, and insect venoms. They are N-glycans with a core of two N-acetyl glucosamine (GlcNAc) sugars and two or three mannose residues. These structures are further characterized by the addition of a β1,2-linked xylose to the proximal mannose and/or an α1,3-linked fucose to the innermost GlcNAc of the core. Classical CCDs feature foreign monosaccharides, non-human linkage, or aberrant non-human glycosylation patterns. Group B includes non-human mammalian oligosaccharides, such as the galactose disaccharide with an α1,3 linkage (α-Gal epitope). This IgE-binding α-Gal epitope has been implicated in anaphylaxis associated with red meat allergy and with the therapeutic mAb cetuximab allergy. In group B, the non-human sialic acid N-glycolylneuraminic acid (Neu5Gc) has also been shown to interact with antibodies. Group A and B are recognized for their clinical significance and their importance is evident from *in vitro* molecular diagnostics, where the presence of these specific modifications is analyzed. Other carbohydrate groups are still under investigation for their role in allergology as no study has directly linked these carbohydrates with specific IgE (Homann et al. 2017; Platts-Mills et al. 2021).

Taken together, these findings highlight that allergenicity is not determined by a single property, but rather by a complex interplay of abundance, function, molecular modifications, and the context of exposure. Yet, the variability observed even within closely related protein families suggests that additional, as of yet unidentified factors – potentially linked to three-dimensional structure, epitope accessibility, or host-specific responses – also contribute. Even conventional processing methods (e.g., boiling, baking, frying, pasteur-

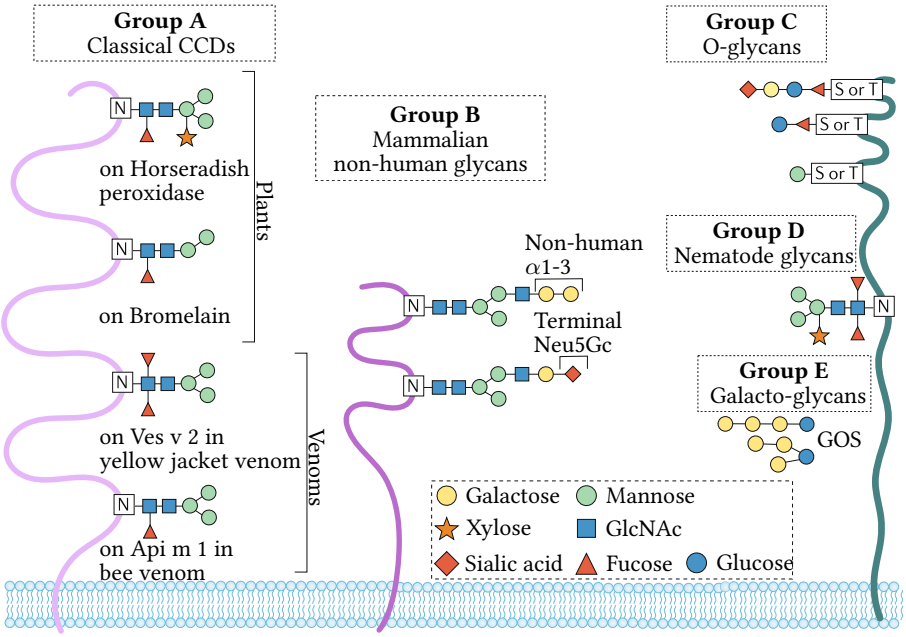


Figure 1.10: Groups of oligosaccharides recognized by IgE. Group A: classical CCD examples from plants and venoms. Group B: mammalian non-human oligosaccharides, such as the disaccharide Gal-α1,3-Gal and the monosaccharide N-glycolylneuraminic acid (Neu5Gc), another type of sialic acid. Groups A and B are clinically significant, whereas the clinical relevance of other groups is still under investigation. Group C: O-glycans. Group D: glycans derived from nematode parasites. Group E: short-chain galacto-oligosaccharides (GOS), known as prebiotic carbohydrates. These categories summarize the main oligosaccharide structures currently implicated in IgE recognition. Information was taken from Platts-Mills et al. 2021, design based on Reily et al. 2019, made with Inkscape (see 2.4).

ization, sterilization, and others) can drastically influence allergen antigenicity, underscoring that understanding these determinants remains a critical challenge for allergology and highlighting the need for precise molecular characterization. (Zeng et al. 2025). This complexity is directly reflected in the field of allergy diagnostics, where distinguishing clinically relevant sensitization from cross-reactivity remains an ongoing challenge.

1.5. Allergy diagnostics

Allergic diseases have a profound impact on quality of life, ranging from chronic skin inflammation and respiratory symptoms to life-threatening anaphylaxis. Given the broad spectrum of clinical manifestations and the lack

of a universal treatment, the first essential step in patient management is the accurate identification of the causative allergen. Advances in molecular allergology now enable diagnostics not only with traditional allergen extracts but also with recombinant allergen components, facilitating improved risk assessment and guiding both *in vivo* and *in vitro* diagnostic approaches, as well as subsequent immunotherapy.

1.5.1 Recombinant allergens vs. Allergen extracts

Historically, allergen extracts derived from natural sources were used to diagnose and treat allergic diseases. However, these extracts exhibit considerable variability, lack standardization, and contain a complex mixture of allergenic and non-allergenic components (Zimmer et al. 2021). Several studies have shown that using such extracts in *in vitro* or *in vivo* assays can lead to misdiagnosis when relevant allergens are absent. For instance, lipophilic allergens such as oleosins or defensins cannot be purified using conventional extract production methods, hence these allergens would not be found in commercial extracts (Jappe and Schwager 2017). Similarly, allergens with naturally low abundance or poor stability may escape detection (Nugraha et al. 2021). Misdiagnosis can also occur in the presence of homologous proteins (Y Zhang, Matsuo, et al. 2006) and due to undefined non-allergic materials, contaminants (Valenta, Karaulov, et al. 2018). These limitations are inherent to natural allergen sources and extract manufacturing methods and cannot be entirely overcome.

The understanding that sufficient and consistent presence of major allergens is decisive for the quality and efficacy of both immunotherapy and diagnostic products led to the revolutionized field of allergy diagnostics (Zimmer et al. 2021). One of the key objectives of large-scale European research initiatives, e.g., *CREATE* (Consortium for the Standardization of Allergen Extracts), *BSP090* (Biological Standardization Programme), and *MeDALL* (Mechanisms of the Developmental Allergy), was the production of high-quality, standardized allergen extracts/molecules with robust reproducibility. While this endeavor was achieved with varying degrees of success, it advanced the understanding of molecular mechanisms underlying allergic sensitization and clarified the relationships between sensitization patterns and clinical symptoms. Importantly, they also explored using native allergen molecules versus recombinant allergens (Anto et al. 2017).

Recombinant allergens, produced through genetic engineering, offer a

more consistent and precise approach to allergy diagnosis. They facilitate the identification of specific allergens responsible for allergic reactions, a process termed component-resolved diagnosis (CRD). CRD allows clinicians to distinguish between genuine sensitization and cross-reactivity, refine risk assessment, and provide more personalized therapeutic strategies. Furthermore, recombinant allergens enable the production of hypoallergenic variants, which promise a safer and more effective allergen-specific immunotherapy (Matricardi et al. 2016). With the global rise in allergic diseases, the demand for reliable and accurate diagnostic tools is increasingly urgent. Using recombinant allergens in allergy diagnostics represents a critical step toward achieving more targeted and effective management of allergic disorders (Sánchez-Borges et al. 2018).

Insect venom allergens provide a clear example of the progression from natural allergen extracts to recombinant allergens. Stings by members of the Hymenoptera order account for 20.2% and 48.2% of anaphylaxis cases in individuals below and over the age of 18, respectively (Worm et al. 2014). Severity varies between individuals; however, those who have experienced an allergic reaction have a 60% chance of a similar or worse reaction if stung again. The only long-term solution is venom immunotherapy (VIT). Although VIT is highly effective, systemic allergic side effects to injections occur in 20-40% of patients (Perez-Riverol et al. 2015). The underlying cause is the use of extracts, which contain not only allergens with immunogenic glycosylation patterns but also a variety of pharmacologically active amines and enzymes that can exacerbate the allergic responses (Bazon et al. 2018). This example illustrates the inherent drawbacks of natural extracts and underscores the potential of recombinant allergens to provide safer, more standardized alternatives for both diagnostics and therapy.

Given the advantages of recombinant allergens, it may seem surprising that allergen extracts still dominate routine practice. This is particularly notable given that several allergens have very limited natural availability. For recombinant allergen production, the key challenge lies in the selection of an appropriate expression system, that balances cost, labor, and inherent protein characteristics (Jayakrishnan et al. 2024). Selection requires careful consideration of the need for PTMs, structural complexity for preserving conformational epitopes, required solubility, as well as practical constrains such as production time, efficiency, consumer acceptance, and regulatory compliance for therapeutic applications (M Liu et al. 2025).

In practice, recombinant allergen production is typically performed in *E. coli* hosts due to its cost-effectiveness, rapid growth, and simple cultivation conditions. However, *E. coli* cannot perform PTMs or secrete proteins, complicating downstream purification. In addition, heterologous proteins often misfold into inclusion bodies, losing biological activity and requiring extensive refolding protocols, which can result in losses of 75-85% of the recovered protein (Jayakrishnan et al. 2024). For example, the peanut allergen Ara h 2 expression requires specialized *E. coli* strains such as Origami, that enable cytoplasmic disulfide bond formation necessary to maintain IgE reactivity (Lehmann et al. 2003). However, specialized strains are associated with low yields, contrary to the intended outcome.

Consequently, scientists and industry specialists have employed eukaryotic systems to produce recombinant allergens. Eukaryotic expression systems offer the advantage of performing PTMs essential for allergen recognition by sIgE antibodies: proteolytic processing of signal peptides, subunit assembly, phosphorylation, acetylation, acylation, disulfide bond formation, and glycosylation (Rapp and Reichl 2021).

S. cerevisiae is generally regarded as safe (GRAS) and is utilized as a model organism to synthesize a variety of food-safe products: natural colorants (e.g., betanin, anthocyanins, catechins, lutein), food flavorings (e.g., valence), functional dietary supplements (e.g, glutathione, vitamins), and others (M Liu et al. 2025). Although relatively uncommon, *S. cerevisiae* has been used to produce recombinant allergens with successful and unsuccessful outcomes reported (e.g., immunoreactive procalin allergen, insoluble dust mite (*Dermatophagoides pterosynus*) Der p 1 allergen, immunobinding Der p 2 allergen) (Schmidt and Hoffman 2002). Despite its GRAS status and capacity for PTMs, *S. cerevisiae* is known to hyperglycosylate recombinant proteins, which reduces its therapeutic use (Jayakrishnan et al. 2024).

The methylotrophic yeast *P. pastoris* is an established platform for industrial production of pharmaceutical- and food-grade recombinant proteins (M Liu et al. 2025). Compared to *S. cerevisiae*, *P. pastoris* offers several advantages, including higher heterologous protein expression levels, reduced background secretion of host-derived proteins, and more efficient secretion into the culture medium. Its ability to perform PTMs, such as glycosylation, makes it suitable for producing complex mammalian proteins. Unlike *S. cerevisiae*, which typically modifies glycoproteins with hypermannosylated N-glycan structures, containing α-1,2/3/6-mannoses and phosphomannoses that can

exceed 100 residues, *P. pastoris* produces shorter N-linked glycans with a lower polymerization degree, incorporating α-1,2/3/6-, β-1,2-mannose residues and occasional phosphomannoses (Rapp and Reichl 2021). Mannosylation as a form of glycosylation is preferred as mannosyl-type glycans appear to be a dominant feature among allergens (Al-Ghouleh et al. 2012).

Insect cell and plant-based systems offer more authentic PTMs for certain allergens, but their higher costs and longer production times limit their application to specialized cases. Insect cells employing baculovirus expression vectors can yield properly folded insect venom allergens such as Api m 1. However, glycosylation differences, particularly α1,3-fucose residues, can cause diagnostic interference (see 1.4.3 Fig. 1.10 Group A CCDs). Mammalian cell systems, while providing human-compatible glycosylation, remain economically unfeasible for routine allergen production (M Liu et al. 2025; Blank, Michel, et al. 2011).

The continued dominance of *E. coli* and *P. pastoris* in commercial recombinant allergen production reflects a practical reality: most clinically relevant allergens can be produced in functional form without complex PTMs. For diagnostic applications, where CCD interference must be minimized, non-glycosylated variants are often preferred regardless of the native protein's glycosylation status. However, glycosylation influences a protein's biological activity, half-life, and interactions with other molecules. There is evidence that certain glycans are clinically relevant and that glycan modifications can act as adjuvants, enhancing the immunogenicity of various allergens (Hale et al. 2024). A systematic understanding of how glycosylation contributes to allergy is still lacking. Thus, optimal expression system selection should increasingly depend on the intended application: simplified proteins for diagnostics versus more native-like structures for immunotherapy. The requirement for PTMs should be assessed on a case-by-case basis to advance our understanding of molecular allergology.

1.5.2 *In vivo* vs. *in vitro* allergy diagnostics

The double-blind placebo-controlled food challenge (DBPCFC) remains the gold standard for IgE-mediated food hypersensitivity diagnosis, even if its false-positive rate exceeding 50% highlights fundamental limitations in current diagnostic approaches (Renz et al. 2018; Osguthorpe 2014). More concerning is the inherent risk in *in vivo* diagnostics: intentional exposure to suspected allergens carries a substantial risk of severe adverse reactions in

sensitized patients. To maintain optimal safety measures, DBPCFC requires highly qualified personnel to oversee the procedure, multi-hour monitoring periods, and immediate access to specialized rescue services and equipment, making it resource-intensive and inaccessible for routine screening (Kowalski et al. 2016).

These practical constraints shifted the diagnostic approach towards quantitative *in vitro* assays over the past decade (Fig. 1.11b). Modern immunoassay platforms, notably ImmunoCAP® (Phadia, Thermo Fisher Scientific), Immulite (Siemens Healthineers), and HYTEC-88 (Hycor Biomedical), offer compelling alternatives by quantifying sIgE antibodies against specific allergen components rather than crude extracts. These assays minimize patient risk, can be standardized across laboratories, and offer improved sensitivity and specificity compared with traditional extract-based tests (Renz et al. 2018).

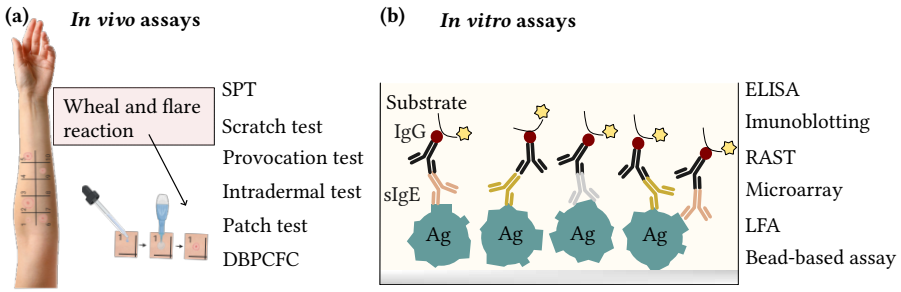


Figure 1.11: Overview of allergy diagnostic platforms. **(a)** *In vivo* assays detect local tissue reactions (wheal and flare) resulting from type 2 immune-mediated inflammation following allergen exposure through various routes: epidermal (SPT, scratch test), intradermal, mucosal (provocation test), transdermal (patch test), or oral (DBPCFC). **(b)** *In vitro* assays detect patient sIgE by their binding to immobilized allergens (Ag) using labeled anti-IgE antibodies, with signal readouts adapted to different platforms including ELISA, immunoblotting, radioallergosorbent test (RAST), microarrays, lateral flow assays (LFA), and bead-based multiplex assays. Figure created with Inkscape (see 2.4).

Despite the shift, *in vivo* practices remain standard in most hospitals. However, while DBPCFC is the reference, the skin prick test (SPT) dominates the clinical practice due to practical necessity rather than diagnostic superiority. Within 15-20 minutes and at minimal cost, SPT reveals mast cell-bound sIgE through characteristic wheal-and-flare responses to epidermal allergen exposure (Fig. 1.11a). Despite its clinical utility, the accuracy of SPT is limited by extract variability, interference from antihistamines, and the presence of

extensive eczema or other skin conditions. *In vivo* approaches also cannot distinguish genuine sensitization from cross-reactivity when using heterogeneous allergen mixtures (L Heinzerling et al. 2013; Renz et al. 2018).

In contrast, *in vitro* assays such as ImmunoCAP®, ELISA, or microarray-based systems allow for standardized measurement of allergen-specific IgE, independent of patient skin reactivity or medication status. In the cases of well-characterized allergies, serum sIgE levels correlate with clinical reaction probability, providing risk assessment without imposing any risk to the patient. Beyond their safety and convenience, the significant advantage of *in vitro* assays is their suitability for CRD. CRD platforms can simultaneously assess IgE reactivity to hundreds of allergen components. This process is impractical for *in vivo* testing because each component must be purified from natural sources or undergo extensive safety testing before recombinant allergens can be approved for human use.

In vitro CRD platforms employ a standard immunoassay format: allergen components immobilized on inert substrates (e.g., paper discs, cellulose matrices, polycarbonate plates) are incubated with patient sera. IgE that recognize allergen epitopes via their Fab regions are detected using labeled anti-human IgE directed against the Fc region. The signal is visualized using colorimetric, fluorimetric, or chemiluminescent readouts (Fig. 1.11b) (Osguthorpe 2014).

Commercial *in vitro* assays are either singleplex, such as ImmunoCAP® or Immulite, which quantify sIgE against one allergen extract or component at a time; or multiplex, like ISAC™ microarrays or ALEX® assays, that enable simultaneous measurement of IgE reactivity to over 100 allergen components from a small serum sample. Precise molecular sensitization profiles are particularly valuable when planning allergen immunotherapy (AIT), as adverse reactions often correlate with the complexity of the patient's sensitization pattern. For instance, in grass pollen allergy, comprehensive sIgE profiling helps identify high-risk patients and optimize AIT protocols, potentially improving safety and efficacy (Barber et al. 2021).

1.5.3 Allergen immunotherapy

Allergies can be managed in four ways: (i) allergen avoidance; (ii) self-treatment with antihistamines, corticosteroids, and adrenaline; (iii) monoclonal antibody treatment; and (iv) AIT. AIT is a form of therapeutic vaccination that aims to develop functional immune system tolerance towards allergens (Durham and Shamji 2023). The classical protocol involves two steps.

First, desensitization, during which effector cells of the type 2 response become less reactive or nonreactive; this state requires regular allergen administration to be maintained. Second, tolerance, also referred to as sustained unresponsiveness, in which desensitization remains a "permanent" state. This state resembles the immune tolerance of healthy individuals; however, it is unclear whether the underlying mechanisms are the same (Renz et al. 2018; Yu et al. 2016).

Mechanistically, AIT promotes the increase of T_{reg} cells and allergen-specific IgG4 levels. T_{reg} cells secrete inhibitory factors such as IL-10, TGF-β, and IFN-γ, which are essential in modulating the type 2 immune response. For example, IL-10 suppresses histamine release during mast cell degranulation, while IFN-γ inhibits the production of hallmark T_H2 cytokines such as IL-4, IL-13, and IL-5. In addition, IL-4 and IL-13 drive plasma cells to switch isotype from sIgE to allergen-specific IgG4 (Pavón-Romero et al. 2022). IgG4, unlike IgE, is a non-inflammatory IgG subclass that does not activate effector mechanisms of the immune system (Qin et al. 2022). AIT also promotes regulatory B (B_{reg}) cell responses and differentiation into the T_H1 cell response (Durham and Shamji 2023).

Tolerance remains the goal of AIT; however, it is not achieved in all patients and is not applicable to all types of allergies. AIT treatment can be administered through various routes, including sublingual (SLIT), subcutaneous (SCIT), oral (OIT), epicutaneous (EPIT), and by direct injection into lymph nodes (ILIT) (Zemelka-Wiacek et al. 2024). Immunotherapy is still not fully understood, in part because the use of allergen extracts complicates precise determination of the molecular doses required for tolerance induction. Moreover, different clinical outcomes can be observed with the same allergen when administered via different routes (Durham and Shamji 2023; Pechsrichuang and Jacquet 2020). For example, in VIT, the administration of native allergens can induce not only adverse side-effects but also severe anaphylactic reactions (Brunetto et al. 2025). These limitations highlight the need for next-generation AIT strategies that employ recombinant allergens, hypoallergenic variants, T-cell based synthetic peptide immuno-regulatory epitopes, or allergoids (Pechsrichuang and Jacquet 2020).

Another disadvantage of AIT is its treatment length, typically requiring allergen administration over two to five years, which results in poor patient commitment to treatment. New strategies aim to overcome the disadvantages of classical AIT by selectively targeting immune pathways. Approaches di-

rected at T cell responses seek to reduce or eliminate the risk of anaphylaxis. In contrast, those targeting B cell responses aim to induce the production of IgG and IgA with IgE-blocking abilities, which would hasten tolerance development (Durham and Shamji 2023).

Although these next-generation strategies differ in their immunological targets and mechanisms, they share a fundamental requirement: the therapeutic proteins must be produced through gene engineering approaches. Whether in the form of recombinant allergens, hypoallergenic variants, or novel fusion proteins, their development relies on molecular biotechnology platforms capable of providing consistent, safe, and standardized products. This convergence highlights not only the potential of engineered therapies to transform allergen immunotherapy, but also the central role of recombinant protein technology in the future of allergology.

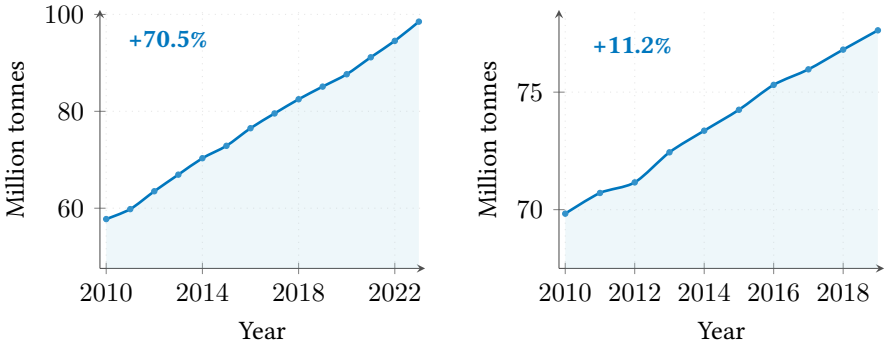
1.6. Molecular allergology: potential adaptations

Having established the molecular mechanisms underlying IgE-mediated allergies, this section examines specific allergens selected for investigation in this thesis based on their clinical significance, structural characteristics, and potential for therapeutic development. These allergens: black tiger shrimp *Penaeus monodon* allergen Pen m 4, dog *Canis familiaris* allergen Can f 6, and yellow jacket *Vespula vulgaris* allergen Ves v 5 represent diverse protein families and offer unique opportunities to explore recombinant allergen production in the eukaryotic expression system.

1.6.1 Food allergy: *Penaeus monodon* allergen Pen m 4

Shellfish allergies represent one of the most prevalent and persistent food allergies worldwide, affecting up to 0.2-5% of the population, though regional rates may reach up to 16% with shrimp being the predominant sensitizing source (Z Zhang et al. 2023; Johnston et al. 2019). The clinical significance of shrimp allergies has intensified with expanding global consumption, which increased 11.2% between 2010-2019, driven by aquaculture production that reached 98.5 Mt in 2023 – a 70.5% increase since 2010 (Fig. 1.12). Global consumption of aquatic animals also expanded, rising by 11.2% between 2010 and 2019 (Fig. 1.12b). Shrimp allergy is common in Western countries such as Europe, the United States and Australia, but appears to be particularly prevalent in Asian countries where shellfish consumption is higher. Black tiger shrimp (*Penaeus monodon*) and Pacific white shrimp (*Litopenaeus vannamei*)

are arguably the most important commercially farmed crustaceans based on FAO data.



(a) Aquaculture production

(b) Utilization and apparent consumption

Figure 1.12: World fishery and aquaculture production and consumption statistics provided by FAO via FishStat (see 2.4). **(a)** Global aquaculture production statistics from 2010 to 2023 of aquatic animals (fish, crustaceans, mollusks and others) in inland waters and marine areas by live-weight volume. **(b)** Global utilization and apparent consumption of aquatic products (2010-2019).

Current diagnostic approaches for shrimp allergy combine patient's medical history, SPT, and sIgE measurements. However, even the most widely used *in vitro* test ImmunoCAP® Shrimp (f24) misdiagnoses patients as it offers high sensitivity but low specificity (Gámez et al. 2011; Johnston et al. 2019). ImmunoCAP® Shrimp (f24) includes raw and cooked extracts from four shrimp species: *Pandalus borealis*, *Penaeus monodon*, *Metapenaeopsis barbata* and *Metapenaus joyneri*. The heterogeneity of shrimp extracts complicates diagnostics, as allergen content varies with species origin, processing methods, and storage conditions. Additionally, heat treatment during food preparation can alter allergen structure and IgE-binding capacity. However, the main limitation primarily stems from tropomyosin, which despite being the major allergen, accounting for 70-100% of shrimp allergies, cross-reacts extensively with other crustaceans, mollusks, and arthropods including house dust mites and cockroaches. Such cross-reactivity leads to false-positive results in dust mite- or cockroach-sensitized patients and fails to identify patients with exclusive sensitization to minor allergens (Costa et al. 2022; Papia et al. 2021; Asero, Pravettoni, et al. 2020).

Among alternative diagnostic targets, sarcoplasmic calcium-binding protein (SCP) presents compelling advantages. First identified as Pen m 4 in

P. monodon with its IgE-binding activity demonstrated in *L. vannamei*, this 20 kDa allergen ellicits positive sIgE responses in 29-50% of shrimp allergic patients and up to 85% of allergic children (Ruethers et al. 2018; Shiomi et al. 2008; Ayuso et al. 2009). Importantly, SCP can cause monosensitization without concurrent tropomyosin reactivity, as demonstrated in occupational asthma cases (Sogo et al. 2018). This suggests the need for standardization of allergen content in commercial assays and the need to include less studied allergens such as arginine kinase, SCP and hemocyanin for improved diagnosis (Johnston et al. 2019; Lopata et al. 2017).

SCP is a water-soluble protein that mimics the function of the vertebrate equivalent, parvalbumin. SCPs are members of the EF-hand superfamily characterized by their common calcium-binding structural motifs. These helix-loop-helix motifs may undergo a conformational change upon binding to Ca^{2+}, allowing the proteins either to act as Ca^{2+} sensors/modulators involved in downstream processes in a Ca^{2+}-dependent manner; or Ca^{2+} buffers that help maintain free cytosolic Ca^{2+} concentration (Y Gao et al. 2006; White et al. 2011). Black tiger shrimp SCP, also known as the allergen Pen m 4, is predicted to have three Ca^{2+} binding sites at positions 18-29, 70-81, and 114-125 as analyzed by the PROSITE program (Hulo et al. 2006; Morii et al. 2013). Previously published studies on the importance of Ca^{2+} binding to shrimp SCP do not clarify whether all binding sites are relevant for IgE binding, as is the case for fish parvalbumin (Swoboda et al. 2007). However, it was shown that IgE reactivity to Pen m 4 was significantly reduced (38-83%) in the presence of a Ca^{2+} chelating agent (Morii et al. 2013). While recombinant Pen m 4 has been successfully produced in *E. coli* with retained IgE reactivity (Mita et al. 2013), no studies have explored eukaryotic expression systems that might better preserve native conformational epitopes.

Given these diagnostic limitations and the clinical importance of Pen m 4, this research aimed to identify optimal production systems for recombinant Pen m 4 that balance three critical parameters: (i) preservation of native IgE epitope recognition, (ii) minimal cross-reactivity with non-specific IgE, and (iii) scalable yields for commercial diagnostic applications.

1.6.2 Pet allergy: *Canis familiaris* allergen Can f 6

Research indicates an increasing prevalence of pet allergies in developed nations, emphasizing the concerning association between these allergies and the development of asthma and rhinitis (Konradsen et al. 2015). In the USA,

up to 15.7% of those aged six years and older report experiencing allergic reactions to cats and dogs (Salo et al. 2014). Similarly, Europe demonstrates a comparable pattern, with dog allergen sensitization averaging 27.2% and reaching as high as 56.0% in Denmark (LM Heinzerling et al. 2009). Despite these concerns, dog ownership remains widespread, with recent surveys indicating that approximately 25% of households in Europe and nearly 45.5% in the USA own at least one dog (FEDIAF 2024; AMVA 2024).

The domestic dog, *Canis familiaris*, is a significant source of indoor allergens, with allergens sequentially designated from Can f 1 to Can f 8 (Allergen Nomeclature, accessed on 2025-08-23). These allergens are found in various biological materials, including dander, sweat, saliva, and urine. Among them, Can f 1, Can f 2, Can f 4, and Can f 6 belong to the lipocalin protein family, a major group of mammalian aeroallergens ubiquitously present in animals, plants, and bacteria (Konieczny et al. 1997; Kamata et al. 2007; Mattsson, Lundgren, Olsson, et al. 2010; Nilsson et al. 2012; Hilger et al. 2012). In contrast, Can f 3, Can f 5, Can f 7, and Can f 8 have been identified as serum albumin, prostatic kallikrein, Niemann pick type C2 protein, and cystatin A, respectively (Pandjaitan et al. 2000; Mattsson, Lundgren, Everberg, et al. 2009; Khurana et al. 2016; Roesner et al. 2022).

In allergology, particularly regarding dog allergies, lipocalin family members are significant as they represent cross-reactive molecules that could facilitate allergic reactions across various species (Hilger et al. 2012). Lipocalins are small extracellular molecules characterized by a common tertiary structure composed of eight anti-parallel β-strands forming a β-barrel structure with one α-helix (Yamamoto et al. 2019). Previous observations show that within the β-barrel is a hydrophobic cavity able to bind small hydrophobic molecules such as retinol, steroids, odorants, and pheromones. These small molecules are secreted into the environment, gradually releasing odorants and extending their longevity (Tegoni et al. 2000; Hilger et al. 2012). The prolonged environmental persistence of these molecules enhances their allergenic potential, as extended-release increases the exposure period for sensitized individuals.

Can f 6 is the most recently characterized lipocalin allergen, identified in dog saliva and dander (Nilsson et al. 2012; Polovic et al. 2013). Its crystalline structure classifies it as a polyvalent allergen with highly structure-dependent epitopes, displaying significant variations in IgE reactivity between its structured and denatured forms (Yamamoto et al. 2019). Conformational epitopes, rather than primary sequence identity, are recognized as the leading cause

of cross-reactivity among allergens (Jimenez-Lopez et al. 2012). Can f 6 sIgE appear in 38-56% of dog-sensitized subjects; however, that figure can rise to 61% when considering cross-reactivities with other lipocalin allergens like the cat allergen Fel d 4 and horse allergen Equ c 1 (Nilsson et al. 2012; Hilger et al. 2012; YJ Wang et al. 2017; Yamamoto et al. 2019). These findings underscore the complex role of Can f 6 in pet allergies and its relevance in cross-species allergenic responses.

Although the structural significance of lipocalin allergens like Can f 6 is well-established, until now, it has only been synthesized in *E. coli* (Nilsson et al. 2012). Despite its diagnostic potential, recombinant Can f 6 (rCan f 6) is not universally used for dog allergy diagnosis as allergen extracts from natural sources – with their inherent limitations – remain the primary basis for allergy diagnosis and AIT (Valenta, Karaulov, et al. 2018). Progress in allergy diagnosis and treatment has been notable for allergens related to cats, grasses, and dust mites. However, advancements in addressing dog allergies have lagged. Despite the widespread prevalence of dog allergies, research outcomes on their diagnosis and treatment remain varied and challenging to standardize (Chan and Leung 2018). Research indicates that the allergen content in commercially available dog allergen extracts can be highly inconsistent, with some, like Can f 6, ranging from 3.3% to non-existent. This inconsistency is especially troubling for Can f 6, which is already one of the least prevalent allergens in these extracts. In comparison, Can f 1 allergen content averages 8.7% and Can f 3 at 47.3%, while Can f 6 stands at 1.3%, highlighting its marked scarcity (Wintersand et al. 2019). Monosensitization to Can f 6 remains an underexplored area in allergy research, with a singular reported case highlighting its significance. This particular instance underscores the pivotal role of CRD and molecular-based methodologies. Previous studies indicate that multi-molecular sensitization in dog and horse allergies is associated with increased severity of rhinitis and asthma. Thus, molecular diagnosis serves not just as an allergy marker but can also aid clinicians in forecasting the nature and intensity of clinical symptoms (Uriarte and Sastre 2016; Wintersand et al. 2019).

Allergen component rCan f 6 can be found in ImmunoCAP® singleplex and ALEX® multiplex assays; however, the manufacturers behind these assays did not disclose the full details of their production, so the expression system used to produce rCan f 6 cannot be determined. In contrast, other systems – such as ImmunoCAP® ISAC, Immulite, NOVEOS (HYCOR Biomedical), and various

indirect enzyme-linked immunosorbent assays or Lateral Flow immunoassays – do not include Can f 6 in their analysis. Studies comparing allergen components with extracts for predicting dog allergy severity consistently demonstrate that measuring sIgE using components provides greater diagnostic accuracy (Schoos et al. 2021). Substituting traditional allergen extracts with recombinant or synthetic ones thus represents advancement toward CRD (Curin et al. 2017).

1.6.3 Insect venom allergy: *Vespula vulgaris* allergen Ves v 5

The majority of insect stings are caused by members of the order Hymenoptera, particularly the families Apidae (bees), Formicidae (ants), and Vespidae (wasps). Exposure is widespread, as 56.6-94.5% of the general population has been stung at least once in their lifetime (Carli et al. 2022). Among venom-sensitized individuals, local allergic reactions are reported in 4.4-4.8%, whereas life-threatening systemic reactions occur in 2.8% (Blank, Haemmerle, et al. 2019). Upon re-stings, 52% report a second large local reaction and 24% experience systemic reactions (Bilò et al. 2019). Insect venom accounts for 20.2-48.2% of anaphylaxis cases across European cohorts (Worm et al. 2014).

Members of the order Hymenoptera can cause allergic reactions, with clinically relevant stings most often caused by honeybees (*Apis mellifera*) and yellow jackets/common wasps (*Vespula vulgaris*) (Blank, Haemmerle, et al. 2019). Given the risk of repeated stings, especially in rural or occupationally exposed populations, high-risk patients are recommended VIT. In Europe and the USA, VIT is performed with allergen extracts derived from natural venoms, supplied as purified or non-purified preparations and as single-venom or mixed-venom formulations (Floyd et al. 2023). However, as noted in 1.5.1, natural allergen extracts are heterogeneous: key allergens may be missing or underrepresented, preparations can contain cross-reactive proteins, and CCDs.

Venom procurement is also technically demanding. Insect colonies must be collected from swarms or nests in the field and transported alive before freezing. After freezing, collectors must identify the species, separate male and female insects, since only females can sting, and deliver characterized insects to specialized laboratories. There, raw venom is collected, refined, and formulated for subsequent applications; for example, wasp venom is frequently obtained by hand from each female wasp (Fig. 1.13).



Collection of live stinging insects → Removal of venom sac by hand → Extraction and preparation of allergen extracts

Figure 1.13: ALK-Abelló A/S workflow for the collection, extraction, and processing of wasp venom. Stills adapted from the company's promotional video on wasp and bee venom production (alk.net). Image created in Inkscape (see 2.4).

Naturally sourced venoms show seasonal and annual variability, and their labor-intensive collection and processing contribute to periodic extract shortages. In response to 1–2-year supply interruptions, the American Academy of Allergy, Asthma & Immunology and the American College of Allergy, Asthma & Immunology issued recommendations to modify VIT protocols and dosing and to restrict treatment in lower-risk patients (Floyd et al. 2023; Golden et al. 2017). Recombinant allergens could mitigate these vulnerabilities by providing standardized composition, scalable production, and independence from seasonal venom collection.

Yellow jacket venom sacs contain three major allergens: phospholipase A, hyaluronidase, and antigen 5. These are the principal targets for molecular allergy diagnostics and AIT. Antigen 5, known as Ves v 5, is the most abundant venom allergen, representing approximately 8.1% of the venom sac content (Borodina et al. 2010). Recombinant Ves v 5 has been produced in several expression systems; however, expression in *E. coli* typically yields insoluble aggregates that require *in vitro* refolding, whereas eukaryotic hosts produce properly folded variants resembling the native protein (Kischnick et al. 2006; Schiener et al. 2017).

The allergen component rVes v 5 can be found in ImmunoCAP® singleplex, ImmunoCAP ISAC, and ALEX® multiplex, as well as in Immulite and other commercial assays; however, the expression systems used to produce rVes v 5 cannot be determined. Despite its broad availability across platforms, rVes v 5 belongs to a highly conserved Antigen 5 family, and this sequence conservation underpins clinically relevant cross-reactivity patterns. Antigen 5 family proteins show 95% sequence identity within the genus *Vespula* and 58-67% identity to *Dolichovespula* and *Polistes*. This explains the high cross-reactivity within the *Vespula* genus and only partial cross-reactivity with other *Vespidae* family members (UR Müller 2002). Consequently, Ves v 5 sIgE do

not specify yellow jacket allergy, but *in vitro* assay result interpretation of paired *V. vulgaris* allergen components can distinguish genuine sensitization. Mono-sensitization to yellow jacket venom is differentiated by sIgE to Ves v 5 or Ves v 1, as tests with positive sIgE to other *V. vespula* allergens without Ves v 1/Ves v 5 sIgE mean that the patient does not have a primary yellow jacket allergy (U Müller et al. 2012).

Due to its abundance, clinical relevance, and structural requirements, Ves v 5 was selected in this study to further evaluate the suitability of *P. pastoris* PichiaPink™ as a system for recombinant allergen production.

2. MATERIALS AND METHODS

2.1. Materials and Resources

2.1.1 Oligonucleotides and plasmid vectors

Table 2.1: List of oligonucleotides

Primer ID	Sequence (5'-3')	RE site
pET28a(+) vector based primers		
T7_FR	TAATACGACTCACTATAGGG	
T7_RV	GCTAGTTATTGCTCAGCGG	
pET28F_P4	ATGGGATCCATGGCTTACTCTT-GGGATAAC	BamHI
pET28R_P4	CTTCTCGAGTTATTGGACAACC-TTCAATGGAC	XhoI
pET28F_C6	ATGGGATCCCACGAAGAAGAA-AACGACGTTG	BamHI
pET28R_C6	CTTCTCGAGTTATTCAGCAGAA-GAAACTTGAGC	XhoI
pET28F_V5	CGGGATCCATGAACAACTACTG-CAAGATC	BamHI
pET28R_V5	GGCACTCGAGTTACTTAGTCTG-GTACAG	XhoI
MBP-TEV DNA sequence based primers		
MBP-TEV-dir	GATACTAGTATGGGCAGCAGCC-ATCATCATC	SpeI (BcuI)
MBP-TEV-rev	GATTCTAGATGATCCTTGGAAG-TATAGATTTTC	XbaI
MBP-TEV-AvrII-rev	GATCCTAGGTGATCCTTGGAAG-TATAGATTTTC	AvrII (XmaJI)

Table 2.1 (cont.) — List of oligonucleotides

Primer ID	Sequence (5'–3')	RE site
MBP-TEV-Seq-dir	CATGCCGAACATCCCGCAGAT	
MBP-StuI-dir	ATGAGGCCTATGGGCAGCAGCCATCATCATC	StuI
pFX7 vector based primers		
PYK5	TATTCATTCTTTTTCATCCTTTGG	
PGK3	TCCTTACCTTCCAATAATTCCAAAC	
pPink-αF-HC vector based primers		
AOX1-Fr	GACTGGTTCCAATTGACAAGC	
CYC1-Rv	GCGTGAATGTAAGCGTGAC	
pPink_oligo_FR	CATCACCATCACCATCACCCCGGGATCGGTAC	SmaI, KpnI
pPink_oligo_RV	CGATCCCGGGGTGATGGTGATGGTGATG	SmaI
pFG-PGK1-αF vector based primers		
6xHis-dir	CCTAGGCATCACCATCACCATCACTAA	AvrII
6xHis-rev	CTAGTTAGTGATGGTGATGGTGATGCCTAGG	AvrII
pPink-αF-6xHis-MBP-TEV vector based primers		
MBP-Pink-Fr	GATAGGCCTATGGGCAGCAGCCATCATCATC	StuI
MBP-Pink-Rv	GATGGTACCGATCCCGGGTGATCCTTGGAAGTATAGATTTTC	KpnI, SmaI
Gene sequence based primers		

Table 2.1 (cont.) — List of oligonucleotides

Primer ID	Sequence (5'–3')	RE site
pPinkMBP_RV_P4	TACGGTACCTTATTGGACAACCTTCAATGG	KpnI
pPinkMBP_RV_C6	TACGGTACCTTATTCAGCAGAAGAAACTTGAG	KpnI
pPinkF_P4	TACGAGTCTAGTCATGGCTTACTCTTGGGATAAC	MlyI
pPinkR_P4	TACGGTACCTTAATGGTGATGGTGATGGTGCCTAGGTTGGACAACCTTCA	KpnI
pPinkF_C6	TACGAGTCTAGTCCACGAAGAAGAAAACGACGT	MlyI
pPinkR_C6	TACGGTACCTTAATGGTGATGGTGATGGTGTTCAGCAGAAGAAACTTGAG	KpnI
pSF_MBP_FR_P4	GATCAATTGCATGGCTTACTCTTGGGATAAC	MunI
pSF_MBP_RV_P4	GATCAATTGTTATTGGACAACCTTCAATGGACC	MunI
pSF_MBP_FR_C6	GATGAATTCGCACGAAGAAGAAAACGACGT	EcoRI
pSF_MBP_RV_C6	GATGAATTCTTATTCAGCAGAAGAAACTTGAGC	EcoRI

RE sites are underlined, "ATG" start codons are colored green, "TTA" end codons are colored red, and 6xHis sequences are colored blue. Forward primers are marked by: F, FR, dir. Reverse primers: R, RV, rev. Oligonucleotides are synthesized by Metabion International AG (Planegg, Germany) or Invitrogen (Glasgow, UK).

Table 2.2: List of base vectors

Vector	Info
pET28a(+)	Novagen, Worcestershire, UK
pET28a(+)-MBP-TEV	A gift from Zita Balklava & Thomas Wassmer (Addgene plasmid #69929; RID:Addgene_69929)
pFX7-6xHis-C	Life Sciences Center, Institute of Biotechnology, Eukaryote Gene Engineering Department collection
pFX7-6xHis-MBP-TEV-6xHis	pFX7-6xHis-C with 6xHis-MBP-TEV fragment, amplified from pET28a(+)-MBP-TEV with MBP-TEV-dir and MBP-TEV-rev primers
pFG-PGK1-αF	Life Sciences Center, Institute of Biotechnology, Eukaryote Gene Engineering Department collection
pFG-PGK1-αF-6xHis	pFG-PGK1-αF with a 6xHis oligo, AvrII RE site, created with 6xHis-dir and 6xHis-rev oligonucleotides
pFG-PGK1-αF-6xHis-MBP-TEV-6xHis	pFG-PGK1-αF-6xHis with 6xHis-MBP-TEV fragment, amplified from pET28a(+)-MBP-TEV with MBP-TEV-dir and MBP-TEV-rev primers
pSF-TEF1-NH2-6xHis-MBP-TEV	Oxford Genetics, Oxford, UK
pPink-αF-HC	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
pPink-αF-6xHis	pPink-αF-HC with a 6xHis oligo, SmaI and KpnI RE sites, created with pPink_oligo_FR and pPink_oligo_RV oligonucleotides
pPink-αF-6xHis-MBP-TEV	pPink-αF-HC with 6xHis-MBP-TEV fragment with StuI and KpnI RE sites, amplified from pFX7-6xHis-MBP-TEV-6xHis with MBP-Pink-Fr and MBP-Pink-Rv primers

2.1.2 Bacteria and yeast strains

Table 2.3: List of bacterial and yeast strains

Strain	Genotype	Source
E. coli DH5α	F- φ80*lac*ZΔM15Δ(*lac*ZYA-*arg*F)U169 *rec*A1 *end*A1 *hsd*R17 (r_{K^-}, m_{K^+}) *pho*A *sup*E44 λ^- *thi*-1 *gyr*A96 *rel*A1	1
E. coli BL21(DE3)	F- *omp*T *hsd*S_B (r_{B^-}, m_{B^-}) *gal dcm ara*B::T7RNAP-*tet*A	1
E. coli Rosetta(DE3)	F- *omp*T *hsd*S_B (r_{B^-}, m_{B^-}) *gal dcm* (DE3) pRARE (Cam^R)	1
S. cerevisiae AH22-214	*MAT*a *leu*2-3 *leu*2-112 *his*4-519 *can*1 [*KIL*-o] (*ATCC*™ 38626)	1
S. cerevisiae AH22-214Δmnn10	*MAT*a *leu*2-3 *leu*2-112 *his*4-519 *can*1 [*KIL*-o]) (*ATCC*™ 38626) *mnn*10	1
K. lactis CBS2359*ura*3Δ-WT (WGI)	wild-type *ura*3-GGI *MAT*a (Ziogiene et al. 2019)	1
K. lactis CBS2359*ura*3Δ-WSS	*ura*3Δ::*KlSEC59*^{G405S,I419S} (Ziogiene et al. 2019)	1
K. lactis CBS2359*ura*3Δ-WSSΔpep4	relevant genotype *pep*4 (Ziogiene et al. 2019)	1
P. pastoris PichiaPink™ Strain 1	relevant genotype *ade*2	2
P. pastoris PichiaPink™ Strain 2	relevant genotype *ade*2 *pep*4	2
P. pastoris PichiaPink™ Strain 3	relevant genotype *ade*2 *prb*1	2
P. pastoris PichiaPink™ Strain 4	relevant genotype *ade*2 *pep*4 *prb*1	2

1 – Life Sciences Center, Institute of Biotechnology, Eukaryote Gene Engineering Department collection; 2 – Thermo Fisher Scientific Baltics, Vilnius, Lithuania.

2.1.3 Protein coding sequences

Protein sequence optimization was performed according to OptimumGene™ Codon Optimization Analysis for efficient gene expression in yeast. Optimized sequences were chemically synthesized by GenScript USA Inc. (Piscataway, NJ, USA) or General Biosystems (Durham, NC, USA).

- *Penaeus monodon* Pen m 4 gene coding sequence was obtained from GenBank (accession no. HM034315.1). The codon optimized Pen m 4 gene sequence (accession no. OP485557) was supplied in pUC57 cloning vector and designated pUC57/ Pen m 4.

- *Canis familiaris* Can f 6 gene coding sequence (accession no. HE653774.1) was obtained from GenBank. The codon-optimized sequence (accession no. PQ679325) was supplied in a pUC57 cloning vector as pUC57/Can f 6.

- *Vespula vulgaris* Ves v 5 gene coding sequence (acession no. M98858.1) was obtained from GenBank. The codon-optimized Ves v 5 gene was supplied in pMA-RQ cloning vector and designated pMA-RQ/Ves v 5.

  Optimized sequence (5'→3'):

  AACAACTACTGCAAGATCAAGTGCCTGAAAGGTGGTGTTCACACCGCTTGTAAGTACGGTT
  CTTTGAAGCCAAACTGCGGTAACAAGGTCGTTGTCTCCTACGGTTTGACCAAGCAAGAGAA
  GCAGGACATCCTGAAAGAGCACAACGACTTCAGACAGAAGATCGCCAGAGGTTTGGAGACT
  AGAGGTAACCCAGGTCCACAACCACCAGCTAAGAACATGAAGAACTTGGTCTGGAACGACG
  AGTTGGCTTACGTTGCTCAAGTTTGGGCTAACCAGTGTCAATACGGTCACGACACCTGTAG
  AGATGTTGCCAAGTACCAAGTTGGTCAGAACGTTGCTTTGACTGGTTCCACTGCTGCTAAG
  TACGATGACCCAGTCAAGTTGGTTAAGATGTGGGAAGATGAGGTCAAGGACTACAACCCAA
  AGAAGAAGTTCTCCGGTAACGACTTCCTTAAGACCGGTCACTACACTCAGATGGTTTGGGC
  CAACACAAAAGAGGTTGGTTGCGGTTCCATCAAGTACATCCAAGAGAAGTGGCACAAGCAC
  TACCTGGTCTGTAACTACGGTCCATCCGGTAACTTCATGAACGAGGAACTGTACCAGACTA
  AG

2.1.4 Human sera

Human sera for recombinant allergen identification and immunochemical characterization were tested with either ImmunoCAP™, Phadia UniCap™ assays, where sIgE scores below 0.35 kU/L were designated negative, and sIgE $\geqslant$ 0.35 kU/L positive; or IgE ALEX^2® Allergy test, where sIgE $\geqslant$ 0.30 kU/L is positive and $\leqslant$ 0.30 kU/L is negative.

Serum pools were prepared by combining individual samples into one mixture used for experiments. Positive-pools were pools combined from sera with positive IgE for the tested allergen extract but negative for the specific

allergen; Double-positive pools were pools combined from sera with positive IgE for both the allergen extract and the specific allergen; Negative-pools were pools combined from sera negative to IgE Ab.

Samples were purchased from PlasmaLab International (Everett, WA, USA). Part of the sera used in the experiments with allergen Pen m 4 was acquired in Lithuania by Imunodiagnostika Ltd. with the permission of the Ethics Committee for Biomedical Research (Nr. 158200-17-926-430). Samples were stored at $-20\,^{\circ}\mathrm{C}$ before testing.

2.1.5 Bioinformatic and programming tools

Table 2.4: Bioinformatic and programming tools

Service	Info
Clustal Omega https://www.ebi.ac.uk/jdispatcher/msa/clustalo	Multiple sequence alignment tool
Expasy https://web.expasy.org	Bioinformatic tool for protein sequence, structure analysis (Translate, Compute pI/MW, ProtParam, GlyConnect, PROSITE)
Reverse Complement https://www.bioinformatics.org/sms/rev_comp.html	Reverse complement of DNA sequences
Multiple Primer Analyzer https://www.thermofisher.com/lt/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html	For analyzing and comparing multiple primer sequences simultaneously
Protpi | Protein Tool https://www.protpi.ch/Calculator/ProteinTool#MolecularMass	Calculates pI and net charge of proteins, as well as the exact mass and the absorption coefficient using the amino acid sequence

Table 2.4 (cont.) — Bioinformatic and programming tools

Service	Info
Glycosylation Predictor https://comp.chem.nottingham.ac.uk/glyco/	Predicts the location of N-linked and O-linked glycosylation sites from an amino acid sequence by random forest algorithm and pairwise pattern prediction tool (Hamby and Hirst 2008)
Peptide 2.0 https://www.peptide2.com/N_peptide_hydrophobicity_hydrophilicity.php	Peptide hydrophobicity and hydrophilicity analysis tool
Grammarly https://app.grammarly.com/	English language writing assistant software tool
ChatGPT https://chatgpt.com/	Used for proofreading, assisting in writing bioinformatics code for statistical analysis
Microsoft Excel version 2202 Build 16.0.14931.20648	Used for statistical analysis
Python, managed via the Anaconda distribution and executed in the PyCharm integrated development environment	Utilized libraries: pandas and numpy for data manipulation, scipy.stats and statsmodels for statistical testing (including repeated measures ANOVA and linear mixed models (LMMs), and matplotlib and seaborn for data visualization
Academic Phrasebank https://www.phrasebank.manchester.ac.uk/	Resource for academic writing
LaTeX, compiled via the TeXstudio 4.8.5 editor	Used for thesis writing
SnapGene https://www.snapgene.com/	Software tool for *in silico* simulation of various molecular biology techniques, especially DNA cloning, sequence analysis, and plasmid mapping
Inkscape https://inkscape.org/	Open-source vector graphics editor used for editing all photos in this thesis

Table 2.4 (cont.) — Bioinformatic and programming tools

Service	Info
ImageJ https://imagej.net/ij/	Open-source image processing program designed for scientific and biomedical image analysis
BioRender https://www.biorender.com/	Illustration tool used for creating scientific figures
Diagrams.net (formerly draw.io) https://app.diagrams.net/	Open-source diagram and flowchart tool
Allergen Nomenclature database https://allergen.org/	Used as a WHO and International union of Immunological Societies approved allergen nomenclature database
UniProt https://www.uniprot.org/	Database for protein sequence and functional information
RCSB PDB https://www.rcsb.org/	Protein data bank, an open-access repository of 3D structural data for biological macromolecules
FishStat https://www.fao.org/fishery/en/topic/fishstat	FAO FishStat is a open-access source of global fishery and aquaculture statistics. Datasets used: Food balance sheets of aquatic products v2023.1.0, Global fishery and aquaculture production statistics v2025.1.0
NCBI NIH https://www.ncbi.nlm.nih.gov/	Open-access database for various bioinformatic tools and datasets (Gene, Nucleotide/Protein, BLAST, Assembly/RefSeq)

2.2. Methods with associated reagents

2.2.1 DNA restriction and dephosphorylation

Buffer solutions and incubation temperature were selected according to the manufacturer's recommendations. One unit of RE was used to digest 1 µg of substrate DNA, with a corresponding 10x reaction buffer. The reaction mixtures were incubated for 1-2 h at the optimal temperature for the RE or overnight at 4 °C. When dephosphorylation was required, one unit of shrimp alkaline phosphatase (SAP) was added 30 min before the end of

incubation. After incubation, DNA dye was added to the sample for electrophoretic separation. If the buffer solution was compatible with subsequent enzymatic reactions and not hindered by degradation products, RE and SAP were inactivated as recommended by the manufacturer.

Reagents	
Restriction enzymes, Shrimp Alkaline Phosphatase (SAP), DNA Gel Loading Dye (6X)	Thermo Fisher Scientific Baltics, Vilnius, Lithuania

2.2.2 Oligonucleotide hybridization

Oligonucleotide hybridization mixture was prepared by combining equal 20 µl volumes of two complementary oligonucleotide DNA sequences (100 pmol/µl), 5 M NaCl and H_2O. After gentle mixing, the solution was heated to 95 °C for 10 min and then cooled gradually to RT in a paper towel-insulated glass beaker. This slow temperature decrease facilitated strand annealing.

Reagents	
NaCl	Merck, Darmstadt, Germany
Solutions	
NaCl sol.	5 M NaCl, stored at 4 °C

2.2.3 DNA electrophoresis

DNA electrophoresis was performed on 1% agarose gels containing 0.1-0.2 µg/ml ethidium bromide. Gels were run horizontally in a TAE buffer at 5 V/cm voltage. DNA bands were visualized using an E.A.S.Y ® Doc Plus transilluminator under ultraviolet light.

Reagents	
Agarose, TAE buffer	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
Ethidium bromide	Sigma Aldrich, Hamburg, Germany
Solutions	
Agarose sol.	1-2% agarose (w/v) in TAE buffer
Ethidium bromide stock	10 mg/ml ethidium bromide solution

2.2.4 DNA fragment purification from an electrophoresis gel

A clean agarose gel was prepared using a broad-track forming comb, and the DNA samples were dyed with the SYBR™ Green I Nucleic Acid Gel Stain. Following electrophoresis, the desired site was marked under blue LED light

(470 nm; IO Rodeo Mini Transilluminator). According to the manufacturer's recommendations, an agarose strip containing a DNA fragment was excised and purified with the GeneJET Gel Extraction or PCR Purification kits.

Reagents	
Agarose, GeneJET Gel Extraction Kit, GeneJET PCR Purification Kit, SYBR™ Green I Nucleic Acid Gel Stain	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
Solutions	
Agarose sol.	1-2% agarose (w/v) in TAE buffer

2.2.5 DNA ligation

The DNA ligation reaction was performed in a ligase buffer solution for 1 h at RT or overnight at $4\,^{\circ}\mathrm{C}$ using 5 units of T4 DNA ligase per 1 µg of ligated DNA. The molar ratio of vector to insert ends in the ligation mixture ranged from 1:3 to 1:7. When the reaction occurred, the T4 DNA ligase was inactivated by heating at $65\,^{\circ}\mathrm{C}$ for 10 min. If ligation was unsuccessful, 1 µl dithiothreitol (DTT) was added to stabilize the T4 DNA ligase. If one of the components was a hybridized oligonucleotide, or in the case of blunt ends – $\leqslant$ 5% PEG 4000 was added to the ligation mixture.

Reagents	
T4 DNA Ligase, PEG 4000, DTT	Thermo Fisher Scientific Baltics, Vilnius, Lithuania

2.2.6 Polymerase chain reaction (PCR)

PCR was used for colony screening and amplification of DNA fragments using primers listed in Table 2.1 under conditions listed in Table 2.9.

- **Colony PCR**: bacterial colonies were picked directly from Petri plates and suspended in the reaction mixture containing buffer, dNTPs, $MgCl_2$, *Taq* DNA polymerase, and primers. Yeast colonies were lysed before PCR by suspending the colony in 10 µl of H_2O, adding 5 µl lyticase (5 U/µl) or zymolyase (0.5 U/µl), incubating for 10 min at $30\,^{\circ}\mathrm{C}$, and freezing for 10 min at $-80\,^{\circ}\mathrm{C}$. The lysed cells were used for PCR.
- **Fragment amplification PCR**: DNA fragments were amplified using Phusion™ HF Buffer, dNTPs, Phusion™ High-Fidelity DNA Polymerase, primers, and the target vector.

Table 2.9: Thermal cycling parameters used for colony PCR screening (Conditions 1) and DNA fragment amplification (Conditions 2). Primer extension time was adjusted according to target length.

Step	Temperature	Conditions 1	Conditions 2
Initial denaturation	98 °C	2 min	30 s
Denaturation	94 °C / 98 °C	30 s	10 s
Primer annealing	55 °C	30 s	30 s
Primer extension	72 °C	30–100 s	15–30 s
DNA polymerization	72 °C	10 min	10 min
Final hold	4 °C	∞	∞
Cycles	–	32	32

Reagents	
Taq DNA polymerase kit, dNTP mixture, $MgCl_2$, Phusion™ High-Fidelity DNA Polymerase kit	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
Lyticase	Sigma Aldrich, Hamburg, Germany
Zymolyase	Seikagaku Corporation, Tokyo, Japan
Solutions	
Lyticase sol. (5 U/ml)	5000 U/ml lyticase mixed with PBS + 5% glucose
Zymolyase sol. (0.5 U/ml)	20000 U/g zymolyase mixed with PBS + 5% glucose

2.2.7 Plasmid DNA purification from *E. coli*

The target bacterial colony was grown in an LB medium containing the appropriate antibiotic at 37 °C for 16 h with shaking. Cells were harvested by centrifugation for 7 min at 3000 rpm (Eppendorf Centrifuge, model 5810R); the supernatant was decanted. Plasmid DNA was purified using one of the following methods:

- the pellet was processed with the GeneJET Plasmid MiniPrep Kit following the manufacturer's protocol.
- the pellet was resuspended in the residual medium and lysed by adding 10 volumes of NaOH-SDS solution. The mixture was gently inverted several times and neutralized with an equal volume of NaAc (pH 4.8), mixed, and incubated at −20 °C for 8 min. Chloroform (0.5 volumes) was added, and the sample was mixed and centrifuged for 10 min at 13200 rpm (Eppendorf Centrifuge, model 5415D). The aqueous phase was recovered, mixed with 0.8-1 volume of isopropyl alcohol, and incubated for 5 min at −20 °C. The sample was centrifuged for 5 min at 13200

rpm. The precipitate was washed with cold 70% ethanol, which was then carefully decanted. The pellet was air-dried at 37 °C or RT and dissolved in nuclease-free water containing RNase A (5 mg/ml).

- the pellet was resuspended in the residual solution and lysed by the addition of 1 µl glycogen, 7.5 M NH_4OAc at half the sample volume, and 100% ethanol at 2.5 × (sample + NH_4OAc) volume. The sample was mixed gently and incubated for 1 h (or overnight) at −20 °C. The sample was centrifuged at 13200 rpm in 4 °C for 30 min (Eppendorf Centrifuge, model 5810R). The supernatant was discarded without disturbing the pellet. The pellet was washed twice with 150 µl 70% ethanol, each followed by centrifugation for 2 min at 13200 rpm in 4 °C. The remaining ethanol was removed, and the pellet was air-dried at RT for 5-10 min, then resuspended in 50-100 µl of water by pipetting 30-40 times.

DNA concentration was determined by spectrophotometry (NanoDrop 2000 UV Visible Spectrophotometer).

For pPink-αF-HC vector-based plasmids, plasmid purification after colony PCR (see 2.2.6) was performed directly from liquid cultures. Colonies were not pre-cultured on agar plates but were instead inoculated directly into the LB media containing the appropriate antibiotic and incubated overnight, as pre-culturing prevented efficient plasmid purification.

Reagents	
Luria Bertani (LB) media, ampicillin, kanamycin sulphate, GeneJET Plasmid MiniPrep Kit, glycogen, 2-propanol, RNase A	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
Sodium dodecyl sulphate (SDS)	Carl Roth, Karlsruhe, Germany
NaAc, NH_4OAc, chloroform	Sigma Aldrich, Hamburg, Germany
Ethanol	Honeywell, Charlotte, NC, USA
Solutions	
LB media (pH 7.0)	25 g/L LB media, autoclaved at 1 atm for 20 min
NaOH-SDS sol.	1% SDS (v/v), 0.2 M NaOH
NaAc sol. (pH 4.8)	3 M NaAc, stored at 4 °C
NH_4OAc sol.	7.5 M NH_4OAc, stored at 4 °C
Ethanol sol.	70% ethanol (v/v), stored at −20 °C
Kanamycin sol.	50 mg/ml, stored at −20 °C
Ampicillin sol.	50 mg/ml, stored at −20 °C
RNase sol. (5 mg/ml)	10 mg/ml RNase, stored at −20 °C

2.2.8 Preparation of chemically competent *E. coli* cells and transformation via heat shock

- **Large batch preparation**: 25 ml LB media was inoculated with the desired strain and incubated overnight at $37\,^{\circ}\mathrm{C}$ with shaking. The overnight culture (25 mL) was then used to inoculate 500 ml of fresh LB medium in a 2 L flask, which was incubated at $37\,^{\circ}\mathrm{C}$ with shaking until an OD of 0.6-0.7 was reached. The cells were harvested by centrifugation at 3000 rpm for 10 min in $4\,^{\circ}\mathrm{C}$ (Beckman J-6B centrifuge). All subsequent steps were carried out on ice. The supernatant was discarded, and the pellet was gently resuspended in 25 ml of ice-cold 100 mM $CaCl_2$. The suspension was transferred to a 50 ml Falcon tube and centrifuged at 3000 rpm for 10 min at $4\,^{\circ}\mathrm{C}$ (Eppendorf Centrifuge, model 5810R). The supernatant was discarded, and the cells were resuspended in 25 ml ice-cold $CaCl_2$ and incubated overnight at $4\,^{\circ}\mathrm{C}$. After incubation, the cells were collected under the same conditions, the supernatant was removed, and the cells were gently resuspended in 5 ml of 4:1 $CaCl_2$:Glycerol solution. Chemically competent cells were aliquoted into sterile Eppendorf tubes and stored in $-80\,^{\circ}\mathrm{C}$.

- **Small batch preparation**: the overnight culture was diluted in a ratio of 1:20-100 in fresh selective LB media and incubated at $37\,^{\circ}\mathrm{C}$ with shaking for 2-3 h. All subsequent steps were carried out on ice. Cells were collected by centrifugation at $4\,^{\circ}\mathrm{C}$ for 8 min at 2000 rpm (Eppendorf Centrifuge, model 5810R) and washed with ice-cold NaCl solution. After centrifugation under the same conditions, cells were resuspended in a three-fold volume of cold $CaCl_2$ solution and incubated for 1 h in an ice bath. After incubation, the cells were pelleted by centrifugation under the same conditions, the supernatant was decanted, and the cells were resuspended and used for transformation.

- **Transformation**: a mixture of competent cells and DNA was incubated for 30 min in an ice bath, followed by a 1.5 min heat shock in a $42\,^{\circ}\mathrm{C}$ water bath. The mixture was then cooled for 10 min in an ice bath and diluted with an RT LB medium lacking antibiotics. After incubation at $37\,^{\circ}\mathrm{C}$ for 10-60 min, the cells were collected by centrifugation for 3 min at 4500 rpm (Eppendorf Centrifuge, model 5415D). The supernatant was decanted, and the cells were resuspended and plated onto selective LB agar. Plates were incubated at $37\,^{\circ}\mathrm{C}$ for 12-18 h.

Reagents	
Luria Bertani (LB) media, ampicillin, kanamycin sulphate, glycerol	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
$CaCl_2$	Sigma Aldrich, Hamburg, Germany AppliChem, Darmstadt, Germany
Solutions, autoclaved at 1 atm for 20 min	
LB media (pH 7.0)	25 g/L LB media
Kanamycin sol.	50 mg/ml, stored at $-20\,^{\circ}C$
Ampicillin sol.	50 mg/ml, stored at $-20\,^{\circ}C$
$CaCl_2$ sol.	100 mM $CaCl_2$, stored at $4\,^{\circ}C$
$CaCl_2$:Glycerol sol.	100 mM $CaCl_2$ mixed with glycerol in a ratio of 4:1, stored at $-20\,^{\circ}C$

2.2.9 Preparation of chemically competent *S. cerevisiae* cells and transformation via heat shock

The yeast strain used for transformation was inoculated into the YEPD medium and grown overnight at $30\,^{\circ}C$ with shaking. The overnight culture was diluted 1:100 in fresh YEPD media and incubated for 3 h at $30\,^{\circ}C$ with shaking. Cells were centrifuged for 5 min at 2000 rpm (Eppendorf Centrifuge, model 5810R), and the supernatant was decanted. The pellet was resuspended in TE-LiAc buffer and incubated for 1 h at $30\,^{\circ}C$. After incubation, the sample was centrifuged under the same conditions, the supernatant was removed, and the pellet was resuspended in the residual buffer. The cell suspension was mixed with plasmid DNA at a 3:1 ratio and incubated for 30 min at RT. Two volumes of PEG were added, and the mixture was incubated for 1 h at $30\,^{\circ}C$. Following incubation, the sample was subjected to a heat shock in a t $42\,^{\circ}C$ water bath for 15 min. The cells were suspended in YEPD medium and incubated overnight at $30\,^{\circ}C$. The next day, the cells were pelleted (Eppendorf Centrifuge, model 5415D) and plated onto selective YEPD agar, followed by incubation at $30\,^{\circ}C$ for two days. Several resulting colonies were subsequently streaked onto high-selectivity YEPD-Ag plates and incubated for an additional two days at $30\,^{\circ}C$.

Reagents	
Agar, D-glucose, TRIS-HCl	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
Yeast extract, soy peptone	AppliChem, Darmstadt, Germany
LiAc, EDTA	Sigma Aldrich, Hamburg, Germany
Polyethyleneglycol 4000 (PEG-4000)	Serva, Heidelberg, Germany
Ethanol	Honeywell, Charlotte, NC, USA
37% formaldehyde sol.	Amresco, Solon, OH, USA
Solutions, autoclaved at 1 atm for 20 min	

YEPD (yeast, extract, peptone, dextrose) media	1% yeast extract, 2% dextrose/glucose, 2% peptone (all w/v); for YEPD-Ag add 1.5-2% agar (w/v); autoclaved both at 0.8 atm for 30 min
EDTA stock (ph 8.0)	0.5 M EDTA
TE buffer (ph 7.5)	10 mM TRIS-HCl, 1 mM EDTA
TE-LiAc sol. (ph 7.5)	100 mM LiAc in TE buffer
PEG sol.	50% (w/w) PEG-4ooo in TE buffer

2.2.10 Preparation of electrocompetent *P. pastoris* cells and transformation via electroporation

The yeast strain used for transformation was inoculated into the YEPD medium and grown for 2 days at 30 °C with shaking. The overnight culture was diluted 1:100 in fresh YEPD media and grown overnight until an OD of 1.3-1.5 was reached. Cells were pelleted at 1500 rpm for 5 min at 4 °C (Eppendorf Centrifuge, model 5810R), and the supernatant was discarded. The pellet was washed twice with ice-cold sterile water, with centrifugation performed under the same conditions. After removal of the residual medium, the pellet was resuspended in 10 ml 1 M D-sorbitol. Cells were centrifuged again, and the supernatant was removed. The final pellet was resuspended in fresh 1 M D-sorbitol to a final volume of 500 µl.

Electroporation was done by mixing 80 µl of electrocompetent yeast cells with 5-10 µg of linearized plasmid DNA. The mixture was transferred to pre-chilled electroporation cuvettes and incubated on ice for 5 min. Electroporation was carried out at 1.5 kV, 200 Ω, 25 µF, with a pulse duration of ≈ 5 ms.

Immediately after pulsing, the cells were resuspended in YPDS medium and incubated at 30 °C for 2 h. Post-incubation, cells were pelleted by centrifugation at 3000 rpm for 1 min (Eppendorf Centrifuge, model 5415D) and plated onto PAD-Ag plates. The plates were incubated at 30 °C for 3-10 days until distinct colonies appeared. Selected white clones were restreaked onto fresh PAD-Ag plates and grown for 2 days at 30 °C.

Reagents	
Agar, D-glucose,	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
CMS-Ade (Powder)	MP Biomedicals, Irvine, CA, USA
Yeast extract, soy peptone	AppliChem, Darmstadt, Germany
D-sorbitol, yeast nitrogen base without amino acids, D-biotin	Sigma Aldrich, Hamburg, Germany
Solutions, autoclaved at 0.8 atm for 30 min	
YEPD (yeast, extract, peptone, dextrose) media	1% yeast extract, 2% dextrose/glucose, 2% peptone (all w/v); for YEPD-Ag 1.5-2% agar (w/v) was added
D-sorbitol sol.	1 M D-sorbitol, autoclaved at 1 atm for 30 min, stored at 4 °C
Biotin sol.	0.02% D-biotin (w/v), filtrated with 0.2 µm filter, stored at 4 °C

YPDS (yeast extract, peptone, dextrose, sorbitol) media	1 M D-sorbitol in YEPD media
PAD-Ag	13.4 g/L YNB-aa (w/v), 1.25 g/L CMS-Ade (Powder), 2% agar (w/v), 5 mg/ml biotin, 2% D-glucose (v/v)

2.2.11 Induction of recombinant protein expression in *E. coli*

Five colonies were scraped from the transformation plate, resuspended in 3-5 ml of selective LB medium, and grown overnight. The resulting suspension was diluted 1:20-100 times in fresh selective LB medium and incubated at 37 °C with shaking until an OD of 0.6 was reached. The cells were transferred to an ice bath for 5-10 min and then induced with the IPTG solution to a final concentration of 1 mM (0.5 mM for rVes v 5). Induction was carried out overnight (for 3 h for rMBP-Ves v 5) at 24 °C (30 °C for rVes v 5 and rMBP-Ves v 5). Cells were collected by centrifugation at 4000 rpm for 20 min at 4 °C (Beckman Coulter, j-26 xpi).

Reagents	
Luria Bertani (LB) media, ampicillin, kanamycin sulphate	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
IPTG	Invitrogen™, Carlsbad, CA, USA
Solutions	
LB media (pH 7.0)	25 g/L LB media, autoclaved at 1 atm, 20 min
Kanamycin sol.	50 mg/ml, stored at −20 °C
Ampicillin sol.	50 mg/ml, stored at −20 °C
IPTG stock	1 M IPTG, stored at −20 °C

2.2.12 Induction of recombinant protein expression in *S. cerevisiae*

The yeast *S. cerevisiae*, transformed with a plasmid carrying the gene of interest, was cultured for 16-24 h in selective YEPD media. Galactose solution (10%) was then added to one-quarter of the total culture volume to induce expression. Induction was carried out for up to 16-18 h. Following induction, the cells were pelleted by centrifugation for 5 min at 3000 rpm at 4 °C (Beckman Coulter, j-26 xpi). The supernatant was decanted, and the yeast biomass was collected and stored at −20 °C.

Reagents	
D-glucose	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
Yeast extract, soy peptone	AppliChem, Darmstadt, Germany
Galactose	VWR Chemicals, London, UK
37% formaldehyde sol.	Amresco, Solon, OH, USA

Solutions, autoclaved at 0.8 atm for 30 min

YEPD (yeast, extract, peptone, dextrose) media	1% yeast extract, 2% dextrose/glucose, 2% peptone (all w/v)
YEP (yeast extract, peptone) induction media	2% yeast extract, 6% dextrose/glucose, 4% peptone (all w/v)

2.2.13 Induction of recombinant protein expression in *P. pastoris*

The *P. pastoris* expression system was designed to secrete recombinant proteins into the culture media. Therefore, the supernatant was carefully saved and stored at $4\,^{\circ}\mathrm{C}$ following the last centrifugation step. Yeast biomass was also collected when a more detailed analysis was required.

- **For small-scale induction**: single colonies were used to inoculate 10 ml of BMGY medium in 125 ml flasks and incubated at $30\,^{\circ}\mathrm{C}$ with shaking at 250-300 rpm for 1-2 days. The culture was transferred to 50 ml conical tubes, and the cells were harvested by centrifugation for 5 min at 1500 rpm (RT) (Eppendorf Centrifuge, model 5810R). After removal of the supernatant, the pellet was resuspended in 3 ml of BMMY media and transferred to a 50 ml Falcon tube. The suspension was incubated overnight at $30\,^{\circ}\mathrm{C}$ with shaking. A 100 µl aliquot was withdrawn the following day for analysis. Methanol was added to a final induction concentration of 0.5%, and the culture was returned to $30\,^{\circ}\mathrm{C}$ for an additional 24 h. Cells were then harvested by centrifugation for 10 min at 1500 rpm. Additional 100 µl samples were collected at defined intervals to monitor induction kinetics.

- **For large-scale induction**: single colonies were used to inoculate 12 ml of BMGY media in 100 ml baffled flasks and incubated at $30\,^{\circ}\mathrm{C}$ with shaking for 2 days. The 12 ml culture was used to inoculate 500 ml of BMGY media in 2 L baffled flasks and incubated at $30\,^{\circ}\mathrm{C}$ until an OD of 5-6 was reached (approximately 9 h). The cells were harvested by centrifugation at 3000 rpm for 5 min in sterile bottles (Beckman J-6B centrifuge). The supernatant was removed, and the pellet was resuspended in 100 mL of BMMY medium in 1 L baffled flasks to initiate induction. The culture was incubated at $30\,^{\circ}\mathrm{C}$ with shaking. After 24 h, 100% methanol was added to a final concentration of 0.5%, and the cultures were incubated for an additional 24 h. The cells were harvested

by centrifugation for 20 min at 3000 rpm (RT) (Eppendorf Centrifuge, model 5810R).

Reagents	
D-glucose, glycerol	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
CMS-Ade (Powder)	MP Biomedicals™, Irvine, CA, USA
Yeast extract, soy peptone	AppliChem, Darmstadt, Germany
Methanol	Honeywell, Charlotte, NC, USA
D-sorbitol, yeast nitrogen base without amino acids, D-biotin	Sigma Aldrich, Hamburg, Germany
Solutions	
Biotin sol.	0.02% (w/v) D-biotin, filtered with 0.2 µm filter, stored at 4 °C
BMGY (buffered glycerol-complex medium)	1% yeast extract (w/v), 2% peptone (w/v), autoclaved at 0.8 atm for 30 min; when RT 100 mM potassium phosphate buffer was added, 1.34% YNB-aa (v/v), 0.00004% biotin (v/v), and 1% glycerol (v/v)
BMMY (buffered methanol-complex medium)	1% yeast extract (w/v), 2% peptone (w/v), autoclaved at 0.8 atm for 30 min; when RT 100 mM potassium phosphate buffer was added, 1.34% YNB-aa (v/v), 0.00004% biotin (v/v), and 0.5% methanol (v/v)
Potassium phosphate buffer (pH 6.0)	132 ml 1 M K_2HPO_4 + 868 ml 1 M KH_2PO_4; adjust pH; autoclaved 1 atm, 20 min
YNB-aa sol. (yeast nitrogen base without amino acids)	13.4% YNB-aa (w/v), filtered with 0.2 µm filter, stored at 4 °C

2.2.14 Protein SDS-PAGE and polyacrylamide gel staining

The separation matrix consisted of a 14% or 12% polyacrylamide gel overlaid with a lower-concentration stacking gel in a vertical electrophoresis chamber (Hoefer™ Dual Gel Caster) containing Tris-glycine/SDS running buffer (Rotiphorese®). Electrophoresis was carried out at a constant current of 60 mA (170 V) for approximately 55 min. Upon completion, the gel was immersed in *Coomassie* staining solution with gentle agitation for 1 h, then washed in 5% acetic acid for at least 2 h under the same conditions.

Note: for 14% SDS-PAGE analyses performed with Pierce™ Unstained Protein MW Marker #26610, protein size assignment was based on apparent migration and was interpreted in conjunction with the theoretical MW of the target proteins, WB results, and additional comparative SDS-PAGE analyses using other MW markers.

Reagents	
Acrylamide/Bis-acrylamide (29:1), Sodium dodecyl sulphate (SDS), TEMED, ammonium persulfate (APS), Rotiphorese® 10x SDS-PAGE buffer	Carl Roth, Karlsruhe, Germany
TRIS-HCl, 2-propanol	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
Coomassie Brilliant Blue R-250	Sigma Aldrich, Hamburg, Germany

Acetic acid	Chem Pur, Piekary Śląskie, Poland
Ethanol	Honeywell, Charlotte, NC, USA
Solutions	
Tris-HCl/SDS sol. (pH 8.8)	1.5 M Tris-HCl, 0.4% SDS (w/v)
Tris-HCl/SDS sol. (pH 6.8)	0.5 M Tris-HCl, 0.4% SDS (w/v)
APS sol.	10% ammonium persulfate (w/v)
12% separation gel sol.	For one gel: 2.2 ml Acrylamide/Bis acrylamide sol., 1.38 ml Tris-HCl/SDS sol. (pH 8.8), 1.93 ml H_2O 18.4 µl APS sol., 3.7 µl TEMED
14% separation gel sol.	For one gel: 3 ml Acrylamide/Bis acrylamide sol., 1.67 ml Tris-HCl/SDS sol. (pH 8.8), 1.68 ml H_2O 64.3 µl APS sol., 6.4 µl TEMED
Stacking gel sol.	For one gel: 0.24 ml Acrylamide/Bis acrylamide sol., 0.46 ml Tris-HCl/SDS sol. (pH 6.8), 1.12 ml H_2O 9.2 µl APS sol., 1.84 µl TEMED
Coomassie staining sol.	50% ethanol (v/v), 0.07% *Coomassie* brilliant blue R-250 (w/v), 10% acetic acid (v/v)

2.2.15 *E. coli* cell lysis for protein purification

All procedures in this protocol were performed at $4\,^{\circ}\mathrm{C}$. Following the final centrifugation, the clarified lysate was collected for downstream analysis and stored at $4\,^{\circ}\mathrm{C}$ for short-term use or $-20\,^{\circ}\mathrm{C}$ for long-term storage. The pellet was resuspended in 300 µl lysis buffer and stored under the same conditions.

- **Preparation of cleared lysates under native conditions.** *E. coli* biomass was suspended in Lysis buffer at a 1:4 (w/v) ratio, and PMSF was added to the final concentration of 1 mM. The lysate was sonicated using a probe sonicator (Bandelin, model UW3100) operating at 20 kHz and 60% amplitude in 15-second pulses with 15-second rest intervals for 2-5 min. Cells were pelleted by centrifugation at 13200 rpm for 30 min (Eppendorf Centrifuge, model 5425R).

  Lysis buffer No. 2 was used for rPen m 4 and rCan f 6 variants, while lysis buffer No. 1 was used for rMBP-Ves v 5.

- **Preparation of cleared lysates under denaturing conditions.** *E. coli* biomass was suspended in Lysis buffer B at 8 ml per gram wet weight, and PMSF was added to the final concentration of 1 mM. Lysis was carried out for 4 h on a rocking shaker (MR-1 Rocker-Shaker, Biosan). The mixture was centrifuged at 10000 rpm for 30 min (Eppendorf Centrifuge, model 5425R).

Reagents	
$NaH_2PO_4 \times 2\,H_2O$, NaCl	Merck, Darmstadt, Germany
2-propanol, TRIS-HCl	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
HEPES	Carl Roth, Karlsruhe, Germany
PMSF, imidazole, urea	Sigma Aldrich, Hamburg, Germany
Solutions	
Lysis buffer (pH 8.0) under native conditions, No 1	50 mM $NaH_2PO_4 \times 2\,H_2O$, 300 mM NaCl, 10 mM imidazole
Lysis buffer (pH 8.0) under native conditions, No 2	50 mM HEPES, 300 mM NaCl, 10 mM imidazole
Lysis buffer B (pH 8.0) under denaturing conditions	100 mM $NaH_2PO_4 \times 2\,H_2O$, 10 mM Tris-HCl, 8 M urea, 10 mM imidazole
PMSF stock (pH 7.2)	20 mM PMSF in isopropyl alcohol, stored at $-20\,^{\circ}C$

2.2.16 Mechanical tearing of yeast cells

All procedures in this protocol were performed at $4\,^{\circ}C$. To the collected yeast biomass, an equal volume of glass beads and twice the volume of ice-cold lysis buffer were added. Yeast cells were lysed by vigorous shaking on an orbital shaker (Vortex-Genie 2, Scientific Industries Inc.) in 30-second intervals with 30-second rest periods for 10 min (10 cycles). The whole procedure lasts 10 min (10 cycles). The lysate was centrifuged at 2000 rpm for 10 min (Eppendorf Centrifuge, model 5810R). The clarified lysate was collected for downstream studies and stored at $4\,^{\circ}C$ for short-term use or $-20\,^{\circ}C$ for long-term storage.

Reagents	
$NaH_2PO_4 \times 2\,H_2O$, NaCl	Merck, Darmstadt, Germany
Na_2HPO_4	Carl Roth, Karlsruhe, Germany
2-propanol	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
PMSF, EDTA	Sigma Aldrich, Hamburg, Germany
Solutions	
Lysis sol.	2 mM PMSF, 2 mM EDTA in PBS
PMSF stock (pH 7.2)	20 mM PMSF in isopropyl alcohol, stored at $-20\,^{\circ}C$
EDTA stock (pH 8.0)	0.5 M EDTA, autoclaved at 1 atm for 20 min
PBS buffer (pH 7.5)	0.08 M Na_2HPO_4, 0.024 M $NaH_2PO_4 \times 2\,H_2O$, 0.099 M NaCl, filtered with 0.2 µm filter, autoclaved at 1 atm for 20 min

2.2.17 Rapid total protein extraction from yeast by alkaline lysis

A 100 µl aliquot of overnight (or extended) yeast suspension was collected, and the cells were pelleted by centrifugation for 1 min at 13200 rpm (Eppendorf Centrifuge, model 5415D). The supernatant was decanted, and the pellet was resuspended in 200 µl of the lysis solution. After incubating cells for 5-10 min,

cells were centrifuged for 1 min at 4000 rpm. The supernatant was removed, and the cells were gently resuspended in 100 µl of 1x protein-dye solution. Samples were boiled for 5-10 min at 100 °C, and cell debris was pelleted by centrifugation for 1 min at 8000 rpm. A 2-4 µl aliquot of the supernatant was carefully collected without disturbing the pellet and loaded onto a PAA gel.

Reagents	
TRIS-HCl, glycerol	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
Bromphenol blue	Merck, Darmstadt, Germany
NaOH	Standard, Lublin, Poland
Sodium dodecyl sulphate (SDS)	Carl Roth, Karlsruhe, Germany
Mercaptoethanol	Sigma Aldrich, Hamburg, Germany
Solutions	
Lysis sol.	0.2 M NaOH
Tris-HCl sol. (ph 7.5)	1 M TRIS-HCl, stored at 4 °C
2x protein dye	100 mM Tris-HCl, 4% SDS (w/v), 0.1-0.2% Bromphenol blue (w/v), 20% glycerol (w/w), and 10% mercaptoethanol (v/v) added just before use

2.2.18 Manual Ni-NTA batch purifications

Before use, the Ni-NTA resin was equilibrated by centrifugation at 1000 rpm for 30-60 s (Eppendorf Centrifuge, model 5810R). The supernatant was carefully removed, and the resin was washed twice with two volumes of lysis buffer. After each wash, the resin was pelleted under the same conditions. Following equilibration, the agarose was immediately mixed with the clarified lysate.

- **Ni-NTA purification under native conditions**. The resin-lysate suspension was incubated in 4 °C for 2 h on a rocker shaker (MR-1 Rocker-Shaker, Biosan). The solution was centrifuged at 1000 rpm for 30-60 s (Eppendorf Centrifuge, model 5810R) and loaded to a gravity flow column with the bottom outlet capped. After removing the cap, the column flow-through was collected for subsequent analysis. The resin was washed three times with two volumes of Wash Buffer per resin volume: the first wash included a 10-minute incubation, while the second and third were performed without incubation. Bound protein was eluted by adding one resin volume of Elution buffer, incubating for 10 min, and collecting the eluate; this step was repeated once. Ten additional elution fractions were collected with 2 min incubations each.

- **Ni-NTA purification under denaturing conditions**. The resin-lysate suspension was incubated overnight in 4 °C on a rocker shaker (MR-1

Rocker-Shaker, Biosan). The solution was centrifuged at 1000 rpm for 30-60 s (Eppendorf Centrifuge, model 5810R) and loaded to a capped column. After uncapping, the column flow-through was collected for subsequent analysis. The resin was washed sequentially with denaturing Wash buffers C, C1, D, D1, and D2, each applied at two resin volumes. Four sequential washes were performed for each buffer: the first wash without incubation, the second with a 5 min incubation, and the third and fourth without incubation. Bound protein was eluted using Elution buffer at half the resin volume per fraction. The first eluate was collected immediately without incubation; the subsequent eight to fifteen fractions were each incubated for 5-15 min before collection.

Reagents	
$NaH_2PO_4 \times 2\,H_2O$, NaCl	Merck, Darmstadt, Germany
TRIS-HCl	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
Ni-NTA agarose	QIAGEN, Hilden, Germany
Imidazole, urea	Sigma Aldrich, Hamburg, Germany
Solutions	
Wash buffer (pH 8.0) under native conditions	50 mM $NaH_2PO_4 \times 2\,H_2O$, 300 mM NaCl, 10 mM imidazole
Elution buffer (pH 8.0) under native conditions	50 mM $NaH_2PO_4 \times 2\,H_2O$, 300 mM NaCl, 250 mM imidazole
Lysis buffer B (pH 8.0) under denaturing conditions	100 mM $NaH_2PO_4 \times 2\,H_2O$, 10 mM Tris-HCl, 8 M urea, 10 mM imidazole
Wash buffer C (pH 6.3) under denaturing conditions	100 mM $NaH_2PO_4 \times 2\,H_2O$, 10 mM Tris-HCl, 8 M urea, 20 mM imidazole
Wash buffer C1 (pH 6) under denaturing conditions	100 mM $NaH_2PO_4 \times 2\,H_2O$, 10 mM Tris-HCl, 8 M urea, 20 mM imidazole
Wash buffer D (pH 5.9) under denaturing conditions	100 mM $NaH_2PO_4 \times 2\,H_2O$, 10 mM Tris-HCl, 8 M urea, 50 mM imidazole
Wash buffer D1 (pH 5.5) under denaturing conditions	100 mM $NaH_2PO_4 \times 2\,H_2O$, 10 mM Tris-HCl, 8 M urea, 50 mM imidazole
Wash buffer D2 (pH 5.0) under denaturing conditions	100 mM $NaH_2PO_4 \times 2\,H_2O$, 10 mM Tris-HCl, 8 M urea, 100 mM imidazole
Elution buffer E (pH 4.5) under denaturing conditions	100 mM $NaH_2PO_4 \times 2\,H_2O$, 10 mM Tris-HCl, 8 M urea, 250 mM imidazole

2.2.19 Purification of proteins using ZYMO Reasearch columns

Protein purification using the His-Spin Protein Miniprep™ and MBP-Spin Protein Miniprep Kit™ was done according to the manufacturer's instructions. Purification was conducted from *E. coli* clarified lysate or supernatant, as well as from yeast clarified lysate, supernatant, or culture medium.

Reagents	
His-Spin Protein Miniprep™ MBP-Spin Protein Miniprep Kit™	Zymo Research, Irvine, CA, USA

2.2.20 Protein dialysis

The dialysis procedure involved several preparatory steps, including assessing the protein's isoelectric point (pI), molecular weight (MW), and other factors that could impact protein stability. An appropriate dialysis buffer was selected based on the protein's pI and the intended purpose of dialysis. A dialysis membrane with a molecular weight cut-off (MWCO) matching the protein's size was also chosen. The dialysis buffer volume was typically 200-500 times that of the protein sample.

Buffer selection was dependent on both the target protein and the intended application:

- **Ion exchange purification**: dialysis buffer was selected to ensure the protein acquired a charge opposite to the chromatography resins. Adjusting the pH below or above the protein pI induces a positive charge when pH < pI and a negative charge when pH > pI, thereby enabling binding to the oppositely charged resin.
- **Ni-NTA purification**: for effective binding of His-tagged proteins, dialysis was done in buffers with pH values ranging from 7.6 to 9.0.
- **Protein refolding after purification under denaturing conditions**: stepwise dialysis, usually three steps. The sample was dialyzed sequentially against refolding buffers with progressively lower denaturant concentrations to gradually restore the protein to its native state.

All dialysis buffers were chosen so that their pH differed by at least one unit from the protein's pI and, where possible, remained within a physiologically relevant range to avoid pH extremes.

Dialysis procedure: before use, dialysis membranes were pretreated by immersion in deionized water and microwaved until boiling. The membranes were left to sit in the hot water for 1 min, then rinsed with deionized water to cool. Once one membrane end was sealed, the protein sample was carefully transferred inside. For *P. pastoris* culture media, samples were filtered through a 0.45 μM pore PVDF membrane before dialysis. After transferring, the membrane was sealed and submerged in a dialysis buffer. Dialysis was conducted at

4 °C with slow stirring using a magnetic mixer (Rotilabo®-heating magnetic stirrer MH15). The buffer was replaced after 2-3 h, again after 2-5 h, and left overnight. Depending on the protein, dialysis was either completed or continued for an additional day with another buffer change after 3-5 h.

If the protein concentration is too low, dialysis fractions were concentrated with Pierce™ protein concentrators of specific MWCO.

Reagents	
Stericup Quick Release-HV sterile vacuum filtration system, $NaH_2PO_4 \times 2\,H_2O$, NaCl	Merck, Darmstadt, Germany
TRIS-HCl, glacial acetic acid, Pierce™ protein concentrators PES	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
Imidazole	Sigma Aldrich, Hamburg, Germany
Solutions	
CEX dialysis buffer (pH 4.3) for rMBP-Pen m 4 and rMBP-Can f 6	50 mM acetic acid-NaOH
CEX dialysis buffer (pH 6.0) for rMBP-Ves v 5	25 mM acetic acid-NaOH
AEX dialysis buffer (pH 7.5) for rMBP-Pen m 4 and rMBP-Can f 6	20 mM Tris-HCl
Ni-NTA dialysis buffer (pH 8.0) for rPen m 4, rCan f 6 and rVes v 5	50 mM $NaH_2PO_4 \times 2\,H_2O$, 300 mM NaCl, 10 mM imidazole

2.2.21 Protein concentration from yeast media with methanol

1 mL of yeast culture medium was transferred to a 1.5 mL microcentrifuge tube and centrifuged at 2000 rpm for 3 min (Eppendorf centrifuge, model 5415D) to pellet the cells. From the resulting supernatant, 400 µl was transferred to a fresh tube, mixed with 400 µl of methanol by vortexing, followed by the addition of 200 µl of chloroform. The mixture was briefly vortexed and centrifuged at 13200 rpm for 10 min to partition proteins at the interphase. The upper aqueous phase was carefully removed and discarded. To the remaining interphase and organic phase, 600 µl of methanol was added, and the mixture was vortexed to resuspend the proteins. After centrifugation at 13200 rpm for 10 min, the supernatant was discarded, retaining the protein pellet. The pellet was washed twice by adding 600 µl of methanol, briefly vortexed, and centrifuged at 13200 rpm for 5 min, and the supernatant was discarded after each wash. The pellet was air-dried for 5-10 min at RT. The dried pellet was dissolved by adding 22 µl PBS and 22 µl 2x protein-dye solution, gently vortexed, and incubated at 100 °C for 10 min to ensure complete solubilization.

Reagents	
TRIS-HCl, glycerol	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
Methanol	Honeywell, Charlotte, NC, USA
Bromphenol blue	Merck, Darmstadt, Germany
Sodium dodecyl sulphate (SDS)	Carl Roth, Karlsruhe, Germany
Mercaptoethanol, chloroform	Sigma Aldrich, Hamburg, Germany
Solutions	
Tris-HCl sol. (ph 7.5)	1 M TRIS-HCl, stored at 4 °C
2x protein dye	100 mM Tris-HCl, 4% SDS (w/v), 0.1-0.2% Bromphenol blue (w/v), 20% glycerol (w/w), and 10% mercaptoethanol (v/v) added just before use

2.2.22 Ion chromatography

Purifications using the *Ä*KTApurifier 100 chromatography system were performed by Dr. Mindaugas Zaveckas.

The collected *P. pastoris* culture media was filtered through a 0.45 µM pore PVDF membrane with the Stericup Quick Release-HV sterile vacuum filtration system. The purifications were done at 4 °C using the *Ä*KTApurifier 100 chromatography system equipped with the sample pump P-960 and the fraction collector Frac-920. Cation exchanger SP Sepharose Fast Flow (GE Healthcare Bio-Sciences AB) and anion exchanger Q Sepharose Fast Flow (GE Healthcare BioSciences AB) were selected for the purification of the fusion protein rMBP-Pen m 4 and rMBP-Can f 6, while rMBP-Ves v 5 was purified using only the cation exchanger.

- **Cation exchange chromatography**: the clarified media was dialyzed (see 2.2.20) against CEX dialysis buffer. The C 10/10 column (GE Healthcare Bio-Sciences AB) was packed with 2 ml of the SP Sepharose Fast Flow sorbent and equilibrated with 10 column volumes (CV) of CEX dialysis buffer. The clarified media was loaded on the column at the flow rate of 2 ml/min. The column was washed with 7 CV (8 CV for rMBP-Ves v 5) of CEX dialysis buffer. Next steps were conducted at the flow rate of 1.5 ml/min (1.3 ml/min for rMBP-Ves v 5). The target protein was eluted with a linear gradient from 0-100% CEX elution buffer over 10 CV.

- **Anion exchange chromatography**: the clarified media was dialyzed (see 2.2.20) against AEX dialysis buffer. The XK 16/20 column (GE Healthcare Bio-Sciences AB) was packed with 5 ml of the Q Sepharose Fast Flow sorbent and equilibrated with 10 CV of AEX dialysis buffer at

the flow rate of 5 ml/min. Further steps were done at the flow rate of 3.5 ml/min. The clarified media was loaded onto the column and washed with 12 CV of AEX dialysis buffer. The target protein was eluted in a linear gradient from 0-100% AEX elution buffer over 10 CV.

Reagents	
Stericup Quick Release-HV sterile vacuum filtration system, $NaH_2PO_4 \times 2\,H_2O$, NaCl	Merck, Darmstadt, Germany
TRIS-HCl, glacial acetic acid	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
Imidazole	Sigma Aldrich, Hamburg, Germany
Solutions	
CEX dialysis buffer (pH 4.3) for rMBP-Pen m 4 and rMBP-Can f 6	50 mM acetic acid-NaOH
CEX elution buffer (pH 4.3) for rMBP-Pen m 4 and rMBP-Can f 6	50 mM acetic acid-NaOH, 1 M NaCl
CEX dialysis buffer (pH 6.0) for rMBP-Ves v 5	25 mM acetic acid-NaOH
CEX elution buffer (pH 6.0) for rMBP-Ves v 5	25 mM acetic acid-NaOH , 1 M NaCl
AEX dialysis buffer (pH 7.5) for rMBP-Pen m 4 and rMBP-Can f 6	20 mM Tris-HCl
AEX elution buffer (pH 7.5) for rMBP-Pen m 4 and rMBP-Can f 6	20 mM Tris-HCl, 1 M NaCl

2.2.23 Ni-NTA chromatography

Purifications using the *Ä*KTApurifier 100 chromatography system were performed by Dr. Mindaugas Zaveckas.

The gathered *P. pastoris* media was filtered through a 0.45 µM pore PVDF membrane with the Stericup Quick Release-HV sterile vacuum filtration system and dialyzed (see 2.2.20) against Ni-NTA dialysis buffer (pH 8.0). Before chromatographic purification, the dialyzed media was again filtered through a 0.45 µM pore PVDF membrane. The purification was performed at $4\,^{\circ}C$ using an *Ä*KTApurifier 100 chromatography system equipped with a sample pump P-960 and a fraction collector Frac-920 (GE Healthcare Bio-Sciences AB, Uppsala, Sweden).

- **rPen m 4**. The clarified media was directly loaded (1.4 ml/min) onto the Ni-NTA Superflow column (1 ml, Qiagen, Hilden, Germany) equilibrated with 10 CV of the Ni-NTA dialysis buffer. The column was washed with 11 CV of Ni-NTA wash buffer at the flow rate of 1 ml/min. After washing, a linear gradient from 0-100% Ni-NTA elution buffer was applied to elute the target protein at the flowrate of 1 ml/min over 10 CV.

- **rCan f 6 and rVes v 5**. The clarified media was directed onto a Ni-NTA Superflow column (1 ml, Qiagen, Hilden, Germany), equilibrated with the Ni-NTA dialysis buffer at a 1 ml/min flow rate. The column was subsequently rinsed with 10 CV of Ni-NTA wash buffer, and a linear imidazole gradient from 20-500 mM was used to elute the target protein over 10 CV (12 CV for rVes v 5).

Reagents	
Stericup Quick Release-HV sterile vacuum filtration system, $NaH_2PO_4 \times 2\,H_2O$, NaCl	Merck, Darmstadt, Germany
Imidazole	Sigma Aldrich, Hamburg, Germany
Solutions	
Ni-NTA dialysis buffer (pH 8.0)	50 mM $NaH_2PO_4 \times 2\,H_2O$, 300 mM NaCl, 10 mM imidazole
Ni-NTA wash buffer (pH 8.0)	50 mM $NaH_2PO_4 \times 2\,H_2O$, 300 mM NaCl, 20 mM imidazole
Ni-NTA elution buffer (pH 8.0)	50 mM $NaH_2PO_4 \times 2\,H_2O$, 300 mM NaCl, 500 mM imidazole

2.2.24 Protein concentration determination

Purified protein concentration was determined using one of the following methods:

- **Bradford assay**. Measurements were performed in a 96-well, flat-bottom polystyrene plate. Standards of bovine serum albumin (BSA; 1-10 µg per well) and a series of dilutions of the target protein were arrayed such that each well contained the appropriate volume of water (or, for urea-dissolved proteins, an 8 M urea solution) to bring the pre-reagent volume to 50 µl. Then, a 1x Bradford working solution was aliquoted in 200 µl fractions into each well; plates were incubated at RT for 5 min, then read on a Tecan plate reader by measuring the absorbance ratio at 590/450 nm. The OD values for BSA standards were plotted in an X-Y scatter plot and a linear trendline (displaying the equation and R^2 value; accept $R^2 \geqslant 0.95$) was fitted to generate the calibration curve. Sample OD values falling within the standard range were converted to absolute protein amounts using the trendline equation, and concentrations were calculated based on sample volumes. Samples with OD values outside the calibration range were appropriately diluted and reassayed.

- **Densitometric analysis via ImageJ software**. Relative protein concentration was determined by comparing band intensities to BSA standards on SDS-PAGE gel using ImageJ software, developed for analyzing

digitized images (Schneider et al. 2012). After electrophoresis, gel images were processed by converting them to grayscale and correcting for background. Regions of interest (ROIs) corresponding to protein bands were defined, and intensity profiles were extracted. ROI profiles are shown as peaks that are integrated to obtain qualitative values. These values were compared to the calibration curve generated from qualitative measurements of BSA standards, and the corresponding protein concentrations were estimated based on signal intensity, applying scaling or dilution when necessary.

Reagents	
Pierce™ Bovine Serum Albumin Standard Pre-Diluted Set	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
Bradford Plus Protein Assay Reagent	Carl Roth, Karlsruhe, Germany

2.2.25 Protein long term storage

After the last step, each tube was labeled with the protein identity, amount, and date and stored at $-20\,^{\circ}\mathrm{C}$. Prior to downstream use, protein integrity was assessed via SDS-PAGE.

- **Lyophillization.** Aliquots of the protein sample (10-200 µg in $\geqslant$ 0.5 ml volume) were dispensed into screw-cap microcentrifuge tubes wrapped with Parafilm rather than capped. The tubes were transferred to a $-70\,^{\circ}\mathrm{C}$ freezer overnight. Immediately before lyophilization, the Parafilm seals were punctured with a fine needle to permit vapor escape, and the tubes, which were kept on ice, were transferred into a pre-chilled to $-55\,^{\circ}\mathrm{C}$ lyophilizer (Labonco™ FreeZone™ 230 V Model) under vacuum. Depending on sample volume, drying was carried out for 24-48 h. Upon completion, the screw caps were securely fastened.
- **Storage in glycerol.** Aliquots of the protein sample (10-200 µg in $\geqslant$ 0.5 ml volume) were mixed with an equal volume of pure 20-50% glycerol.

Reagents	
Glycerol	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
Solutions	
Glycerol sol.	20-50% glycerol (w/v), filtered with 0.2 µm filter

2.2.26 Western blot (WB)

Protein samples were boiled in a reducing sample buffer and fractioned in 12% and 14% SDS-PAGE gels. After fractioning, the gels, WB filter paper, and the activated polyvinylidene difluoride (PVDF) membrane (pretreated for 15 s with ethanol) were immersed in protein transfer buffer (PTB) and gently agitated for 5-15 min. Following assembly of the transfer stack, proteins were transferred from the SDS-PAGE gels to PVDF membranes via semi-dry electroblotting at 20 V, 700 mA for 40 min (Invitrogen™ Novex™ Semi-Dry Blotter).

The membrane was blocked with Roti®Block protein-free blocking solution for 1 h, washed twice with the protein buffer solution supplemented with Tween-20 (PBS-T) for 5 min each. Incubation with primary antibodies diluted with PBS-T to working concentrations was performed overnight. After incubation, the membrane was washed twice with PBS-T for 5 min each and subsequently incubated with secondary antibodies diluted with PBS-T to working concentrations. After 2 h incubation, subsequent wash steps were performed (15 min, 5 min, and 5 min), and the chemiluminescence signal from the enzymatic reaction was developed by adding 4-chloro-1-napthol and H_2O_2 or 1-Step™ Ultra TMB-Blotting solution.

Primary antibodies used were 6xHis Tag Monoclonal Antibody (mAb) (MA121315, Invitrogen, Carlsbad, CA, USA) at a 1:3000 dilution in PBS-T and an in-house produced murine anti-MBP mAb 19C19 obtained from VU LSC IBT (Sližiene et al. 2023) at a 1:3000 dilution in PBS-T. The secondary antibodies were goat anti-mouse IgG (H + L) HRP conjugate (Cat. No. 170-6516, Bio-Rad, Hercules, CA, USA) at a 1:3000 dilution in PBS-T.

For immunodetection with human sera (see 2.1.4), membranes prepared as above were incubated with sera samples diluted 50-fold in PBS-T. The secondary antibody used was a mouse anti-human IgE Fc-HRP conjugate (B3102E8, SouthernBiotech, Birmingham, AI, USA) at a 1:1000 dilution in PBS-T.

Reagents	
$NaH_2PO_4 \times 2\,H_2O$, NaCl	Merck, Darmstadt, Germany
4-chloro-1-napthol	Sigma Aldrich, Hamburg, Germany
Na_2HPO_4, Roti®Block protein free blocking solution	Carl Roth, Karlsruhe, Germany
TRIS-HCl, 1-Step™ Ultra TMB-Blotting solution	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
H_2O_2	Fluka, Buchs, Switzerland
Polyvinylidene difluoride (PVDF) membrane	Cytiva, Amersham Pl, UK

Tween-20	AppliChem, Darmstadt, Germany
Ethanol, methanol	Honeywell, Charlotte, NC, USA
Solutions	
PBS buffer (pH 7.5)	0.08 M Na_2HPO_4, 0.024 M $NaH_2PO_4 \times 2\,H_2O$, 0.099 M NaCl, filtered with 0.2 µm filter, autoclaved at 1 atm for 20 min
PBS-T	0.1% Tween-20 (w/v) in PBS
PTB	25 mM Tris-HCl, 149 mM glycine, 10% methanol (v/v)
Blocking sol.	1x Roti®Block protein-free blocking solution (v/v)

2.2.27 Glycosylation assay

The glycosylation of the recombinant protein was assessed through a lectin-probed WB analysis, a technique that reveals glycan structures of glycoproteins by exploiting the carbohydrate-binding specificity of the chosen lectin (Sato 2014). Concanavalin A (Con A) lectin was selected for its high binding affinity toward mannose residues (Maupin et al. 2012). The process was done by modifying the previously referenced WB technique (see 2.2.26). Following membrane blocking and washing, incubation was performed overnight at RT with *Canavalia ensiformis* (Jack bean) lectin Concanavalin A peroxidase conjugate (Sigma-Aldrich, MO, USA), diluted to 0.1 µg/ml in PBS-T. The membrane was subsequently washed three times with PBS-T (15 min, 5 min, and 5 min), and the chromogenic signal was detected using 1-Step™ Ultra TMB-Blotting solution.

Reagents	
$NaH_2PO_4 \times 2\,H_2O$, NaCl	Merck, Darmstadt, Germany
Na_2HPO_4, Roti®Block protein free blocking solution	Carl Roth, Karlsruhe, Germany
TRIS-HCl, 1-Step™ Ultra TMB-Blotting solution	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
Polyvinylidene difluoride (PVDF) membrane	Cytiva, Amersham Pl, UK
Tween-20	AppliChem, Darmstadt, Germany
Ethanol, methanol	Honeywell, Charlotte, NC, USA
Solutions	
PBS buffer (pH 7.5)	0.08 M Na_2HPO_4, 0.024 M $NaH_2PO_4 \times 2\,H_2O$, 0.099 M NaCl, filtered with 0.2 µm filter, autoclaved at 1 atm for 20 min
PBS-T	0.1% Tween-20 (w/v) in PBS
PTB	25 mM Tris-HCl, 149 mM glycine, 10% methanol (v/v)
Blocking sol.	1x Roti®Block protein-free blocking solution (v/v)

2.2.28 Deglycosylation assay

The deglycosylation process was performed using Protein Deglycosylation Mix II, following the guidelines provided by the manufacturer. For the reaction, 100 µg of purified protein was combined with water to a final volume of 40 µl. Then, 5 µl of the Deglycosylation Mix Buffer 2 was added, and the mixture was incubated at 75 °C for 10 min to promote protein denaturation, followed by cooling to RT. Next, 5 µl of Protein Deglycosylation Mix II was added. After gentle mixing, the reaction was incubated at 25 °C for 30 min before being shifted to 37 °C for an additional hour to ensure complete deglycosylation. Results were analyzed by SDS-PAGE (see 2.2.14) and WB (see 2.2.26) methods.

Reagents	
Protein Deglycosylation Mix II	New England Biolabs, Ipswich, MA, USA

2.2.29 Electrospray ionization mass spectometry (ESI-MS)

Electrospray ionization mass spectrometry (ESI-MS) is one of the gentlest ionization techniques, producing predominantly intact molecular ions, as analytes are desolvated rather than fragmented. Before the analysis, glycerol was removed from the purified protein samples by diluting the sample with 1x PBS, followed by concentration using Pierce™ Protein Concentrators PES. Sample equilibration with formic acid and acetonitrile, as well as all subsequent mass spectrometric procedures, were carried out by Dr. Audronė Rukšėnaitė. Juta Dvareckienė interpreted the resulting data.

Reagents	
$NaH_2PO_4 \times 2\,H_2O$, NaCl	Merck, Darmstadt, Germany
Na_2HPO_4	Carl Roth, Karlsruhe, Germany
Pierce™ Protein Concentrators PES	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
Solutions	
PBS buffer (pH 7.5)	0.08 M Na_2HPO_4, 0.024 M $NaH_2PO_4 \times 2\,H_2O$, 0.099 M NaCl, filtered with 0.2 µm filter, autoclaved at 1 atm for 20 min

2.2.30 Indirect enzyme-linked immunosorbent assay (ELISA)

Multiwell Nunc MaxiSorp™ polystyrene plates were coated with 50 µl/well of antigen solution prepared in Coating buffer at the concentration of 5 µg/mL and incubated overnight at 4 °C. The coated plates were blocked with 300 µl/well Blocking solution for 2 h at RT. After blocking, the plates were rinsed twice with water. Human sera were diluted 1:10 in PBS-T in a separate microplate

and transferred at 100 µl/well to the coated plate. The plates were incubated at RT with shaking at 400 rpm for 2 h. After incubation, the plates were washed four times with 1:2 diluted PBS-T buffer. Mouse anti-human IgE Fc-HRP conjugate (SouthernBiotech, Birmingham, AL, USA) 1:1000 diluted in PBS-T was added at 50 µl/well and incubated for 1 h under the same conditions. Plates were washed four times with 1:2 diluted PBS-T buffer and twice with water. Antigen-specific antibodies were detected by adding 50 µl/well of TMB chromogen solution. The reaction was stopped by adding 25 µl/well of 3.6% H_2SO_4 solution. The OD was measured at 450 nm (reference filter 620 nm) in a microplate reader (Sunrise Tecan, Männedorf, Switzerland).

Ca^{2+} ion chelation analysis was carried out as written in the above procedures with the addition of EDTA at a concentration of 5 mM to the serum dilution buffer. All ELISA experiments were performed in triplicate, and the data obtained were expressed as a mean (OD) ± SD. The cut-off was calculated as the mean OD of the negative controls plus three SDs.

Experimental controls: the maltose binding protein (MBP), expressed and purified from *E. coli* (Sližienė et al. 2022); the *Penaeus monodon* allergen extract (DST-Code f24, Allergome 1893, Ef 24-7, 20052), obtained from Diagnostische Systeme Technologien GmbH (Schwerin, Germany); the *Canis familiaris* allergen extract was obtained from dr. Vytautas Rudokas (commercial source not specified); the horseradish peroxidase (HRP) (Fluka, Buchs, Switzerland).

Reagents	
$NaH_2PO_4 \times 2\,H_2O$, NaCl	Merck, Darmstadt, Germany
Na_2HPO_4, Roti®Block protein free blocking sol.	Carl Roth, Karlsruhe, Germany
Tween-20	AppliChem, Darmstadt, Germany
$NaHCO_3$	Reachim, Russia
EDTA	Sigma aldrich, Hamburg, Germany
H_2SO_4	Fluka, Buchs, Switzerland
NeA-Blue Tetramethylbenzidine (TMB) substrate	Clinical Science Products, Mansfield, MA, USA
Multiwell Nunc MaxiSorp™ polystyrene plates, TMB chromogen solution	Invitrogen, Carlsbad, CA, USA
Solutions	
PBS buffer (pH 7.5)	0.08 M Na_2HPO_4, 0.024 M $NaH_2PO_4 \times 2\,H_2O$, 0.099 M NaCl, filtered with 0.2 µm filter, autoclaved at 1 atm for 20 min
PBS-T	0.1% Tween-20 (w/v) in PBS
Blocking sol.	1x Roti®Block protein-free blocking solution (v/v)
Coating buffer (pH 9.5)	50 mM $NaHCO_3$
EDTA stock (pH 8.0)	0.5 M EDTA, autoclaved at 1 atm for 20 min
H_2SO_4 sol.	3.6% H_2SO_4 (v/v)

2.2.31 Reverse-phase protein microarray

The concentrations of recombinant protein samples were normalized to 1 mg/ml, and a series of dilutions (2, 4, 8, 16, and 32 times) were printed onto 2D-Epoxy glass slides (PolyAn GmbH, Berlin, Germany) using the sciFLEXARRAYER SX microarray printer (SCIENION GmbH, Berlin, Germany). Each allergen was printed as a single droplet (420 pL/drop) in triplicate at a spot-to-spot distance of 320 µm. Streptavidin-Cy5 (SouthernBiotech, Birmingham, AL, USA) were printed as guide dots for grid alignment. For quality control purposes, several proteins were spotted at their standard concentrations: allergen extracts – 0.2 µg/ml, control recombinant allergens – 0.2 µg/ml, HRP – 0.05 µg/ml, HSA (human serum albumin) – 0.1 µg/ml, MBP – 0.1 µg/ml. Before use, the arrays were stored overnight at $4\,^{\circ}\mathrm{C}$. Before the experiment, the slides were dried at $37\,^{\circ}\mathrm{C}$ for 15 min. Blocking was done with 2% BSA in PBS-T (dilution buffer) while gently stirring the slides for 30 min at RT. After washing once for 30 s with deionized H_2O, the slides were dried via centrifugation for 10 s. Slides were fitted into a 24-Well Hybridization Cassette (Arrayit Corporation, CA, USA), and 80 µl of prepared sera diluted 1:5 in PBS-T was added to each well except for control wells filled with dilution buffer. The fitted slides were sealed with an adhesive seal strip and incubated at $37\,^{\circ}\mathrm{C}$ for 2 h. After washing four times with 100 µl/well of PBS-T, the fitted slides were probed with the mouse anti-human IgE Alexa Fluor 647 conjugate (internal mouse antibody produced by VU LSC IBT) at 80 µl/well volumes for 1 h at RT. After incubation, the slides were washed four times with 100 µl/well of PBS-T. After removing the 24-Well Hybridization Cassette, the slides were incubated with PBS-T with gentle agitation for 10 min. After the final wash for 30 s with deionized H_2O, the slides were dried via centrifugation for 10 s.

Processed slides were scanned using the InnoScan 710 AL microarray scanner (Innopsys, Carbonne, France) with photomultiplier tube (PMT) settings 40 at 635 nm. The images were analyzed using the MAPIX software (Innopsys, Carbonne, France). Each protein's average mean $\pm$ SD value was calculated from the mean fluorescence intensity (MFI) values of the triplicate spots.

A parallel experiment was performed to assess potential CCD interference in the obtained values, adding HRP at a concentration of 100 µg/ml to the serum dilution buffer. The sera were incubated for 10 min at RT prior to use in the assay. 100 µg/ml of HRP inhibits CCD interference by approximately 94%. Spotted HRP served as a positive control for the unaltered human sera samples and a negative control for the inhibited human sera samples.

Reverse-phase protein microarray experiment was done by Laimis Silimavičius in Imunodiagnostika Ltd. (Vilnius, Lithuania). Juta Dvareckienė did data procession and further analysis.

Reagents	
$NaH_2PO_4 \times 2\,H_2O$, NaCl	Merck, Darmstadt, Germany
Na_2HPO_4, Roti®Block protein free blocking solution	Carl Roth, Karlsruhe, Germany
Tween-20	AppliChem, Darmstadt, Germany
Bovine serum albumin (BSA)	Thermo Fisher Scientific Baltics, Vilnius, Lithuania
Solutions	
PBS buffer (pH 7.5)	0.08 M Na_2HPO_4, 0.024 M $NaH_2PO_4 \times 2\,H_2O$, 0.099 M NaCl, filtered with 0.2 µm filter, autoclaved at 1 atm for 20 min
Microarray PBS-T	0.05% Tween-20 (w/v) in PBS
Dillution buffer	2% BSA (w/v) in PBS-T

2.2.32 Statistical analysis for ELISA datasets

The triplicate ELISA datasets were normalized across multiple plates to account for inter-plate variability, ensuring that the results are comparable regardless of the plate-to-plate differences. Normalization was done by subtracting the mean OD of blank wells accounting for background OD. Then, each sample's OD was divided by the OD of its respective control (e.g., the OD value of *C. familiaris* allergen extract reaction with the positive pool plasma) on the same plate. That removed plate-to-plate variability by scaling the sample's values relative to the control of that plate. Once all samples were normalized relative to their plate controls, each normalized sample value was multiplied by the mean control OD (the average of all the plate's positive controls). This step ensured that all plates were adjusted to the same baseline (the total mean), making the data consistent across plates. The triplicate results were presented as the normalized mean OD $\pm$ SD. The cut-off value was determined as the normalized mean OD of the negative controls plus three standard deviations.

For statistical analysis, a repeated-measures ANOVA was performed to assess differences in reactivity between recombinant tested protein variants. The analysis was conducted in Python (via the PyCharm IDE) using pandas and numpy libraries for data manipulation; scipy.stats and statsmodels for statistical testing; and matplotlib and seaborn for data visualization. Tukey's Honest Significant Difference (HSD) test or Bonferroni correction was performed for post hoc pairwise comparisons. Additionally, paired t-tests (t-Test: Paired Two Sample for Means) were performed in Microsoft Excel version

2504 (Build 16.0.18730.20122) to compare specific proteins tested on the same serum samples. Statistical differences were considered significant at $p < 0.05$.

2.2.33 Concentration selection after reverse-phase microarray experiment

The net signal (Δ) for each protein was calculated at the tested concentrations (1, 0.5, 0.25, 0.125, and 0.0625 mg/mL) using the following equation:

$$\Delta = S_{\text{pos}} - S_{\text{neg}},$$

$$\text{where } S_{\text{pos}} = \text{Signal}_{\text{positive sera}}, \quad S_{\text{neg}} = \text{Signal}_{\text{negative sera}}.$$

The bigger Δ signal indicated a higher degree of separation between positive and negative values. Alongside the Δ signal, precision was analyzed by computing the within-group variability (coefficient of variation, CV) for each positive-sera group at each concentration:

$$\text{CV} = \frac{\sigma}{\mu} \times 100\%,$$

$$\text{where } \mu = \text{mean}, \quad \sigma = \text{standard deviation}.$$

These values were used to identify the optimal concentration that maximized signal-to-background difference while maintaining acceptable precision, typically defined as CV below < 2-3%.

2.2.34 Statistical analysis for microarray datasets

The triplicate microarray datasets (mean fluorescence intensities, MFI) were processed and normalized to remove technical bias and make measurements comparable across slides. First, probes flagged as "-100FLAG" or with signal-to-noise ratio (SNR) $\leqslant 2$ were excluded from further analysis. The remaining triplicate MFI values were $\log_2$-transformed to stabilize the variance and reduce right-skewness. Next, the median background fluorescence was subtracted from the $\log_2$-transformed foreground signal for each spot, yielding normalized $\log_2$ values.

Because the experiment was performed on two slides, systematic slide-to-slide variation was assessed by comparing the raw $\log_2$ and background-subtracted $\log_2$ values between slides. Measurements on Slide 2 were found to be, on average, 0.16 units lower on the $\log_2$ scale than those on Slide 1

(corresponding to ≈12 % lower on the original scale). This slight offset was attributed to technical rather than biological factors. Slide 2 $\log_2$-values were each increased by 0.16, thereby aligning both datasets to a common baseline.

After this adjustment, triplicate measurements for the serum pool and control spots (MBP, HRP, allergen extracts, total IgE, HSA, and control proteins) were averaged across both slides. Because the slide-equalization rendered the two technical replicates equivalent, these averages represented a single composite measurement rather than two independent triplicates. Concentration selection for further analysis was done according to the 2.2.33 method.

A linear mixed-effects model (LMM) was fitted to the normalized $\log_2$ MFI values for statistical analysis of variant reactivity. The protein variant was treated as a fixed effect, while slide and serum donor were included as random effects to account for residual inter-slide and inter-individual variability. LMMs were chosen because they appropriately model the correlation structure of repeated measurements (triplicates within slides and multiple antigens within donors) and can handle unbalanced data resulting from probe exclusion. The model was implemented in Python (PyCharm IDE) using pandas and numpy libraries for data manipulation; scipy.stats and statsmodels for statistical testing; and matplotlib and seaborn for data visualization. Post hoc pairwise tests between variants were adjusted using the Bonferroni correction to control the family-wise error rate when testing multiple hypotheses.

For the CCD-interference analysis, normalized $\log_2$ MFI values from uninhibited and inhibited sera were paired for each protein. Normality of the paired differences was assessed using the Shapiro-Wilk test: if $p \geqslant 0.05$, indicating no significant deviation from normality, a paired Student's t-test was applied; otherwise, a nonparametric Wilcoxon signed-rank test was used. Raw p-values across all proteins were then adjusted to control the false discovery rate (FDR = 5%) using the Benjamini-Hochberg procedure - this balances the discovery of true effects with the limitation of type I error when testing multiple proteins. Finally, to evaluate the overall effect of CCD inhibition, a two-way repeated-measures ANOVA was fitted with "protein" and "HRP inhibition" as within-subject factors and "plasma donor" as the subject variable. This test allows assessment of both main effects and the protein $\times$ inhibitor interaction while accounting for the repeated-measures design. Significance thresholds were set at $\alpha = 0.05$ throughout.

3. RESULTS AND DISCUSSION

This thesis aimed to evaluate strategies for producing recombinant allergens in yeast and compare them with counterparts expressed in the prokaryotic host *E. coli*. The comparison was used to identify the optimal candidate for *in vitro* allergy diagnostics, focusing on three key criteria: specific recognition by sIgE antibodies, lack of cross-reactivity with non-specific IgE, and feasibility for high-yield, large-scale production.

This chapter is organized into four thematic sections. The first examines available expression platforms for bioactive recombinant protein production, with particular attention to approaches that enhance yield and improve downstream applications. One such approach is the fusion of allergens to maltose-binding protein (MBP), a strategy known to improve solubility and folding in *E. coli* but with less predictable outcomes in yeast expression systems (Moua et al. 2016).

The second and third sections analyze the production of different allergens, the effects of posttranslational modifications on the allergen's ability to be recognized by sIgE, and the importance of production in a eukaryotic system for the recombinant allergen. Each section addresses the production of recombinant allergens in the most efficient manner, as determined by previous experiments. The fourth section tests the workflow created from the analysis of the first two sections.

3.1. Selection of expression platforms for the production of recombinant allergens

Selecting an appropriate expression system is critical for therapeutic protein production. Key considerations include cultivation and purification costs, the need for PTMs, production time, and the protein's intended application (detailed discussion in 1.5.1). To establish an optimal production platform, systematic evaluation of prokaryotic and eukaryotic expression systems was performed using representative allergens across diverse structural families, focusing on how system-specific limitations impact yield and antigenicity.

The *E. coli* BL21 (DE3) strain provided a prokaryotic baseline using pET28a(+) and pET28a(+)-MBP-TEV vectors (see 2.2) with maltose-binding protein (MBP) fusion to address the well-documented solubility challenges of recombinant allergen expression in bacteria. While MBP has been proven to be effective, the theoretical advantage of amylose-affinity purification often fails in practice, necessitating hexahistidine (6xHis) tagging for reliable nickel-affinity chromatography (Nallamsetty and Waugh 2006; Nallamsetty, Austin, et al. 2005; Bell et al. 2013). This dual-tag strategy established the baseline for subsequent yield and purification efficiency comparisons.

Initial eukaryotic trials focused on *Saccharomyces cerevisiae*, testing both intracellular (pFX7-based) and secreted (pFG-PGK1-αF-based) expression strategies. Expression vectors were similarly designed to include MBP as a fusion partner. Although MBP has been shown to enhance recombinant protein expression in various systems, its performance in eukaryotic hosts remains inconsistent. MBP may increase the yield of specific target proteins, but this effect is not universal, and the molecular determinants underlying such variability are not fully understood (Z Li et al. 2010; Moua et al. 2016; Jiang et al. 2016; J Gao et al. 2023). Notably, to our knowledge, MBP fusion has not been previously employed for recombinant allergen production in yeast systems, representing an unexplored strategy that could potentially address the dual challenges of yield enhancement and native-like folding required for diagnostic applications.

To enable efficient allergen cloning, standardized vector templates were constructed:

- pFX7-MBP-TEV-6xHis vector was constructed by amplifying the 6xHis-MBP-TEV fragment from the pET28a(+)-MBP-TEV plasmid with primers MBP-TEV-dir and MBP-TEV-rev (see 2.1), introducing the restriction

sites of XbaI and SpeI (BcuI). PCR was performed with *Pfu* DNA polymerase according to the manufacturer's recommendations (see 2.2.6). The amplified fragment was cloned into the XbaI site of the linearized pFX7-6xHis-C vector (see 2.2.1, 2.2.5).

- pFG-PGK1-αF-6xHis vector was created by inserting a hybridized oligonucleotide into the pFG-PGK1-αF vector, which was digested by StuI and XbaI RE (see 2.2.1, 2.2.5). The oligonucleotide duplex was prepared by annealing equimolar amounts of complementary 6xHis-dir and 6xHis-rev (see 2.1, 2.2.2) oligos, which incorporated the AvrII restriction site, the 6xHis-tag, and the STOP codon.

- pFG-PGK1-αF-MBP-TEV-6xHis vector was assembled similarly to the pFX7-MBP-TEV-6xHis construct, with the exception that the 6xHis-MBP-TEV fragment was amplified with MBP-TEV-dir and MBP-TEV-AvrII-rev primers (see 2.1), introducing the restriction sites of XbaI and AvrII (XmaJI). The pFG-PGK1-αF-6xHis vector was digested with AvrII prior to ligation with the prepared 6xHis-MBP-TEV fragment (see 2.2.1, 2.2.5).

All constructs were sequence-verified at Vilnius University Life Sciences Center Institute of Biotechnology (VU LSC IBT) sequencing center and analyzed using SnapGene Viewer software (see 2.4).

Pilot experiments were conducted with *S. cerevisiae* (AH22-derived strain 214) using two allergen panels. The initial set comprised dust mite allergens (*Dermatophagoides farinae* Der f 1 and Der f 2, and *D. pteronyssinus* Der p 21 and Der p 36) chosen solely to guide platform selection. A second set with constructs kindly provided by Dr. Rasa Petraitytė-Burneikienė, was comprised of *Penaeus monodon* allergen Pen m 4, *Canis familiaris* allergens Can f 5 and Can f 6, and *Apis mellifera* allergen Api m 4, was included for broader evaluation. Results showed that expression levels were low or undetectable across both intracellular and secreted formats, and target proteins showed marked degradation (Fig. 3.1). Heterologous protein expression into the culture media also showed that *S. cerevisiae* released substantial amounts of host proteins into the medium, complicating protein purification.

Because *S. cerevisiae* is prone to hyperglycosylation of recombinant proteins, a glycosylation-deficient strain, AH22-214$\Delta mn10$, was evaluated as an alternative host. Although target proteins were detectable, Con A lectin assays (see 2.2.27) revealed variability in glycosylation patterns that depended on

colony selection, induction temperature, and induction duration. For example, rApi m 4 was more glycosylated in AH22-214Δ*mn*10 than in the parental 214 strain (Fig. 3.1). Together, these results indicated that *S. cerevisiae* was unsuitable for high-yield, large-scale production for recombinant allergens.

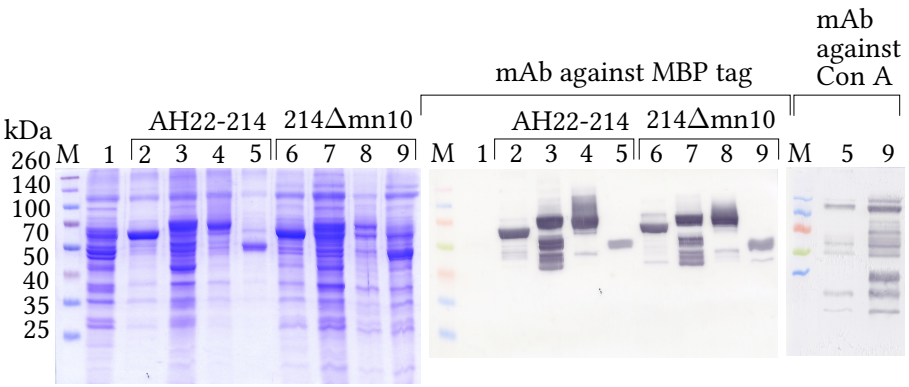


Figure 3.1: Analysis of extracelullar expression in *S. cerevisiae* AH22-derived strain 214 and AH22-214Δ*mn*10 strain by SDS-PAGE and WB (annotated by the used mAb designation). Strains are shown above the corresponding lanes. The analyzed yeast media fractions were concentrated 23-fold with methanol (see 2.2.21). Growth conditions are written in brackets in lane descriptions. Lanes: [1], negative control – AH22-214 untransformed (30 °C, 48 h); [2], pFG-PGK1-αF-MBP-TEV-Pen m 4 (30 °C, 40 h); [3], pFG-PGK1-αF-MBP-TEV-Can f 5 (30 °C, 48 h); [4], pFG-PGK1-αF-MBP-TEV-Can f 6 (24 °C, 48 h); [5], pFG-PGK1-αF-MBP-TEV-Api m 4 (24 °C, 48 h); [6], pFG-PGK1-αF-MBP-TEV-Pen m 4 (30 °C, 40 h); [7], pFG-PGK1-αF-MBP-TEV-Can f 5 (24 °C, 48 h); [8], pFG-PGK1-αF-MBP-TEV-Can f 6 (24 °C, 48 h); [9], pFG-PGK1-αF-MBP-TEV-Api m 4 (30 °C, 40 h). M – Spectra™ Multicolor Broad Range Protein Ladder. (*unpublished data*)

Given the limitations observed in *S. cerevisiae*, subsequent expression experiments were conducted in *P. pastoris*. Numerous recombinant allergens have been expressed successfully in *P. pastoris*, and secretion from this host typically yields media with reduced host background (Brondyk 2009). Therefore, the PichiaPink™ expression system was selected, utilizing the pPink-HC vector with the *S. cerevisiae* α-factor signal peptide under the AOX1 promoter control for secreted production (see 2.2). While the PichiaPink™ expression system was evaluated using allergens from diverse sources, including shrimp, dog, honeybee, and wasp; this study focused specifically on the immunogenicity and antigenicity of *P. monodon* shrimp allergen Pen m 4, *C. familiaris* dog allergen Can f 6, and the *Vespula vulgaris* yellow jacket allergen Ves v 5.

3.2. Recombinant allergen Pen m 4 production and immunological characterization

3.2.1 Generation of codon-optimized Pen m 4 allergen sequence

The gene encoding Pen m 4 was optimized using the OptimumGene™ algorithm software from GenScript (see 2.1.3). The native Pen m 4 codon sequence was modified to achieve better protein expression in yeast, considering various factors: codon adaptability, mRNA structure, and cis-elements in transcription and translation stages. The codon adaptation index (CAI), GC content, and the codon frequency distribution (CFD) predictions were used to compare the expression level evaluation of the native Pen m 4 (nPen m 4) gene expressed in the chosen host system with the codon-optimized Pen m 4 (rPen m 4) gene expression. As shown in Table 3.1, GC content and CFD values are in the ideal range in native and optimized gene sequences. However, the CAI value was higher in the optimized gene and closer to the ideal value range; ergo, the expression level is expected to be higher in both expression systems with the optimized Pen m 4 gene sequence.

Table 3.1: The CAI, GC content, and CFD *in silico* predictions for native and recombinant Pen m 4 constructs in *E. coli* and yeast expression hosts. The ideal values of the predicted elements are shown.

Host / Parameters	***E. coli***		**Yeast**		**Ideal value**
	nPen m 4	*rPen m 4*	*nPen m 4*	*rPen m 4*	
Codon adaptation index (CAI)	0.64	0.72	0.59	0.85	0.8–1.0
GC content (%)	54.48	42.27	54.48	39.83	30–70
Codon frequency distribution (CFD)	9	8	3	0	<30

The numerical values were generated by using the GenScript Rare Codon Analysis Tool (https://www.genscript.com/tools/rare-codon-analysis)

3.2.2 Cloning, expression, and purification of rPen m 4 from *E. coli*

P. monodon Pen m 4 gene encodes a 193-aa-long protein. The 582 bp Pen m 4 gene was subcloned into two expression vectors (see 2.2): (i) pET28a(+) and (ii) pET28a(+)-MBP-TEV, to assess whether an MBP moiety will serve to improve the yield of the passenger protein (Nallamsetty, Austin, et al. 2005). Firstly, the Pen m 4 gene was amplified by PCR (see 2.2.6) using pUC57/ Pen m 4

plasmid as a template with primers pET28F_P4 and pET28R_P4 (see 2.1), which introduced the restriction sites of BamHI and XhoI enzymes, respectively. PCR amplification with the *Pfu* DNA polymerase was done according to the manufacturer's recommendations. The resulting DNA fragment was cloned (see 2.2.1, 2.2.5) into the BamHI and XhoI sites of pET28a(+) and pET28a(+)-MBP-TEV vectors. The presence of Pen m 4 and MBP-TEV-Pen m 4 gene constructs, encoding an N-terminal 6xHis tag, was verified by sequencing at the Vilnius university Life Sciences Center Insitute of Biotechnology (VU LSC IBT) sequencing center, and the resulting sequences were analyzed with the "SnapGene Viewer" software. In the (i) expression vector, the Pen m 4 gene was tagged with an N-terminal 6xHis tag (Fig. 3.2a). In the (ii) vector, the Pen m 4 gene was subcloned in frame with the N-terminally 6xHis tagged MBP protein coding sequence (Fig. 3.2b).

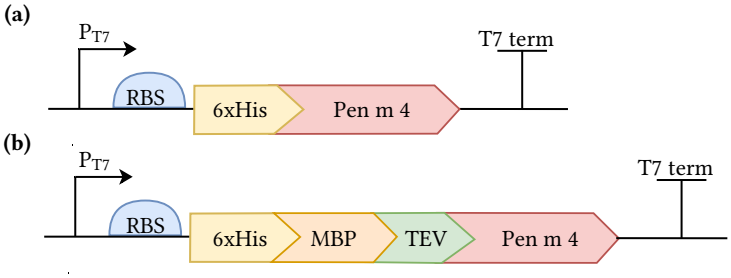


Figure 3.2: Schematic representation of recombinant Pen m 4 gene expression cassettes for production in *E. coli*. **(a)** pET28a(+)-Pen m 4 expression vector, where Pen m 4 gene is N-terminally tagged with a 6xHis tag. **(b)** pET28a(+)-MBP-TEV-Pen m 4 expression vector, where Pen m 4 is fused to the C-terminal end of a 6xHis-MBP fusion protein coding gene with a TEV protease recognition site between the fusion and passenger protein coding sequences.

Expression of rPen m 4 constructs was demonstrated in *E. coli* strain BL21 (DE3) (see 2.2.8, 2.2.15, 2.2.11). As analyzed by SDS-PAGE (see 2.2.14), prominent 25 kDa (Fig. 3.3a) and 66 kDa (Fig. 3.3b) protein bands, corresponding to rPen m 4 and rMBP-Pen m 4, respectively, were observed in the soluble fractions of IPTG-induced bacteria (lane 3 in Fig. 3.3a and 3.3b). Meanwhile, no additional bands of corresponding molecular size were seen in the lysate fractions before induction with IPTG (lane 1 in Fig. 3.3a and 3.3b). Following batch purification of 6xHis-tagged proteins from *E. coli* under native conditions (see 2.2.18), rPen m 4 and rMBP-Pen m 4 were obtained in electrophoretically pure states (lanes 11-14 in Fig. 3.3a and lanes 10-14 in Fig. 3.3b).

Closer inspection of Fig. 3.3b shows that for rMBP-Pen m 4 protein purification, slight modifications could be made to QIAGEN's original purification protocol. The leftover rMBP-Pen m 4 protein in lane 4 indicates that not all protein was successfully bound to the agarose beads; wash fractions in lanes 5-9 show that the target protein eluted too early. Exploring higher amounts of Ni^{2+}-NTA agarose and lower amounts of imidazole in the wash buffer for a successful rMBP-Pen m 4 purification is advised. The yield of *E. coli*-produced rPen m 4 and rMBP-Pen m 4 proteins after nickel-chelate chromatography was approximately 52 mg/L and 66 mg/L, respectively.

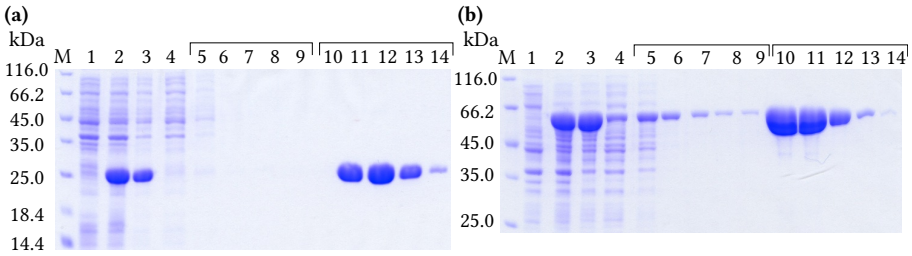


Figure 3.3: Analysis of *E. coli* expression and purification of rPen m 4 **(a)** and rMBP-Pen m 4 **(b)** by SDS-PAGE. Lanes: [1], lysate fraction from non-induced *E. coli*; [2], lysate fraction from IPTG-induced *E. coli*; [3], soluble fraction before purification with nickel-chelate chromatography; [4], soluble fraction after incubation with Ni-NTA agarose; [5-9], wash fractions; [10-14], elution fractions with purified rPen m 4 and rMBP-Pen m 4 proteins, respectively. M – Pierce™ Unstained Protein MW Marker #26610 (see note in 2.2.14). Proteins were separated in (a) 12% and (b) 14% SDS-PAGE gels.

These results are consistent with the literature, as variants of rPen m 4 have been synthesized in bacteria as soluble proteins and could be successfully purified using Ni-NTA affinity chromatography. According to Mita et al. 2013, rPen m 4 expression in *E. coli* using the pFN6A (HQ) Flexi vector system yielded approximately 60 mg/L of purified rPen m 4. In this study, *E. coli*-produced rPen m 4 yielded 52 mg/L. This discrepancy could be attributed to a codon-optimized gene sequence or a different expression strain. However, the comparison could not be made as Mita et al. 2013 did not specify which *E. coli* strain was used for their experiment. Fusion with the dual 6xHis-MBP tag increased target protein yield approximately 1.3-fold (66 mg/L), making this the most successful strategy for synthesizing rPen m 4 in bacteria.

3.2.3 Characterization of rPen m 4 variants purified from *E. coli*

To confirm the purified protein's identity and to determine their molecular weights (MW), mass spectrometry (MS) (see 2.2.29) and WB (see 2.2.26) analyses were performed. MS analysis revealed that MW of *E. coli*-produced rPen m 4 and rMBP-Pen m 4 were approximately 25 and 66 kDa, respectively, which aligned with the target proteins' calculated theoretical weights. WB analysis with mAb against the 6xHis and MBP tags confirmed the expression of the recombinant fusion protein (Fig. 3.4). After one month of storage (see 2.2.25) in different conditions, the WB analysis was performed to assess protein stability. Unexpectedly, protein sample lyophilization as an option for long-term storage appeared to impact both negatively – rPen m 4 (lanes 2 in Fig. 3.4a) and rMBP-Pen m 4 (lanes 2 in Fig. 3.4b) proteins. No significant differences were found between other storage options ($4\,^{\circ}\mathrm{C}$, $20\,^{\circ}\mathrm{C}$ and $-20\,^{\circ}\mathrm{C}$ with 40% glycerol); however, the repeated experiment after 24 months (data not shown) showed that storage at $4\,^{\circ}\mathrm{C}$ and $-20\,^{\circ}\mathrm{C}$ without glycerol severely degraded both protein samples while storage at$-20\,^{\circ}\mathrm{C}$ with 40% glycerol showed only slight degradation.

As rPen m 4 is expected to be used as a pharmaceutical product, storage conditions were evaluated for acceptable shelf life. An unexpected finding was that Pen m 4 is susceptible to freezing and/or drying stress arising from the lyophilization process (W Wang 2000). This finding should be taken into account when preparing allergen extracts. It may explain why the Pen m 4 molecule is absent in 2/5 of commercially available shrimp extracts, mostly made from frozen shrimp (Asero, Scala, et al. 2017).

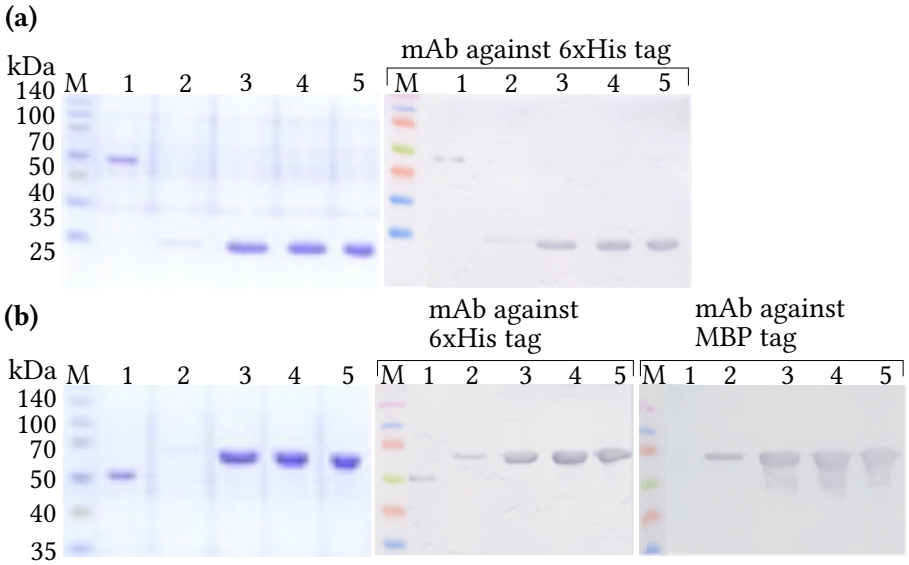


Figure 3.4: Analysis of purified *E. coli*-produced rPen m 4 **(a)** and rMBP-Pen m 4 **(b)** proteins by SDS-PAGE and WB (annotated by the used mAb designation). Lanes: [1], *S. cerevisiae*-produced 6xHis-tagged Hantavirus Sin Nombre nucleocapsid protein (Kucinskaite-Kodze et al. 2011), which acts as the WB positive control; [2], protein samples after lyophilization; [3], protein samples after storage at $4\,^{\circ}\mathrm{C}$; [4], protein samples after storage at $-20\,^{\circ}\mathrm{C}$ without glycerol; [5], protein samples after storage at $-20\,^{\circ}\mathrm{C}$ with 40% glycerol. Protein samples were analyzed after one month of storage. Lane [1] protein acts as a positive control for WB with mAb against the 6xHis tag and a negative control for WB with mAb against the MBP tag. M – Spectra™ Multicolor Broad Range Protein Ladder.

3.2.4 Molecular cloning and expression of rPen m 4 in *P. pastoris*

The PichiaPink™ expression system for secreted protein production was exploited to express Pen m 4 variants in *P. pastoris*. The codon-optimized Pen m 4 gene was amplified by PCR (see 2.2.6) using pUC57/Pen m 4 plasmid as a template with primers pPinkF_P4 and pPinkR_6xHis_P4 (see 2.1), which incorporated the restriction sites of MlyI and KpnI enzymes, respectively, and the 6xHis tag at the C-terminal end of the Pen m 4 gene. The resulting Pen m 4 gene sequence with a C-terminal 6xHis tag was subcloned (see 2.2.1, 2.2.5) in frame with the *S. cerevisiae* alpha-mating factor pre-sequence in the pPink-αF-HC expression vector, resulting in the expression vector pPink-αF-Pen m 4.

To produce a fusion protein MBP-Pen m 4, PCR was performed using pFX7-

6xHis-MBP-TEV-Pen m 4 plasmid (from earlier experiments in *S. cerevisiae*) as a template, with primers pPinkMBPF_P4 and pPinkMBPR_P4 (see 2.1), which incorporated the restriction sites of StuI and KpnI enzymes, respectively. The resulting MBP-TEV-Pen m 4 gene with an N-terminal 6xHis tag was cloned into the StuI and KpnI restriction sites of pPink-αF-HC vector. The pPink-αF-MBP-TEV-Pen m 4 (Fig. 3.5a) and pPink-αF-Pen m 4 (Fig. 3.5b) plasmids were verified by sequencing in the VU LSC IBT sequencing center, and the resulting sequences were checked with the "SnapGene Viewer" software.

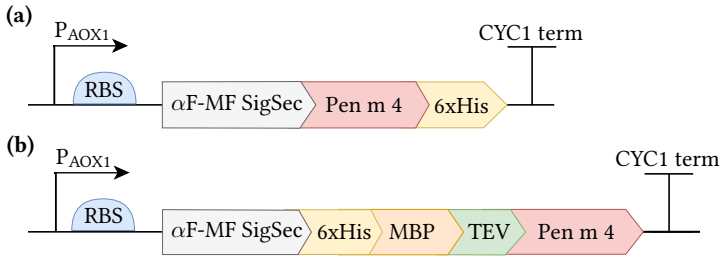


Figure 3.5: Schematic representation of recombinant Pen m 4 gene expression cassettes for production in *P. pastoris*. In frame with the *S. cerevisiae* alpha-mating factor signal pre-sequence (αF-MF-SigSec) in **(a)** pPink-αF-Pen m 4 expression vector, is the Pen m 4 gene C-terminally tagged with a 6xHis tag; in **(b)** pPink-αF-MBP-TEV-Pen m 4 expression vector, is Pen m 4 fused to the C-terminal end of a 6xHis-MBP fusion protein with a TEV protease recognition site between the fusion and passenger protein.

The constructed vectors were transformed into four different *ade2* strains from the PichiaPink™ expression system (see 2.3). These strains differ in their protease knockout genotypes: Strain 1 is the "protease wild-type", Strain 2 is a *pep*4 knockout, Strain 3 is a *prb*1 knockout, and Strain 4 is a double knockout of *pep*4 and *prb*1. Pilot experiments were used to determine which strain produced the highest amount of our target proteins before scaling up expression (see 2.2.13). *P. pastoris*-produced rMBP-Pen m 4 has a theoretical MW of approximately 67 kDa. Analysis by SDS-PAGE (see 2.2.14) and WB (see 2.2.26) with mAb against the MBP tag showed prominent, similar-sized protein bands in all four strains transformed with the expression vector pPink-αF-MBP-TEV-Pen m 4 after heterologous protein induction with MeOH (Fig. 3.6a). Strain 2 was selected for scale-up expression (Fig. 3.6a lane 2).

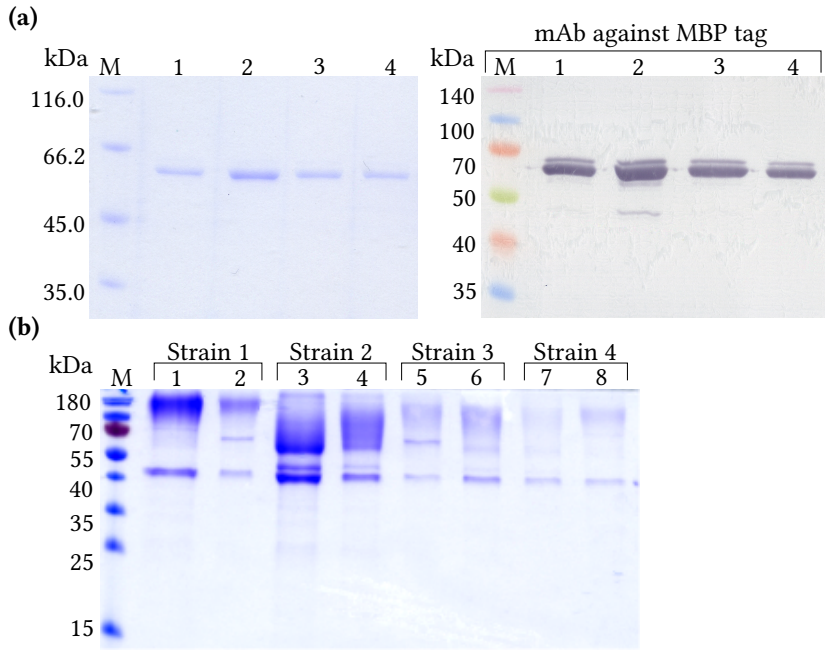


Figure 3.6: Analysis of media collected after heterologous protein induction in PichiaPink™ yeast strains. **(a)** Yeast transformed with pPink-αF-MBP-TEV-Pen m 4 expression vector by SDS-PAGE and WB (annotated by the used mAb designation). Lanes: [1], Strain 1; [2], Strain 2; [3], Strain 3; [4], Strain 4. M – Pierce™ Unstained Protein MW Marker #26610 (see note in 2.2.14) and Spectra™ Multicolor Broad Range Protein Ladder. **(b)** Analysis of yeast, transformed with pPink-αF-Pen m 4 expression vector, culture media samples after induction with MeOH. Culture samples were concentrated with MeOH before being fractioned in SDS-PAGE gel. The names of the strains are written above the corresponding lane numbers. Two colonies from each strain are shown. M – Pierce™ Unstained Protein MW Marker #26617.

rPen m 4 production in *P. pastoris* gave unexpected results. SDS-PAGE analysis of all four strains containing the expression vector pPink-αF-Pen m 4 after induction with MeOH showed protein band sizes from 45-65 kDa and no band corresponding to the theoretical rPen m 4 MW of 23 kDa (Fig. 3.6b).

3.2.5 Purification and characterization of *P. pastoris*-produced rPen m 4

Several tests were performed to elucidate the protein identity in Fig. 3.6b. Dialyzed extracellular *P. pastoris* Strain 2 medium containing the secreted rPen m 4 protein was subjected to affinity FPLC on a Ni-NTA SuperFlow

column (see 2.2.23). Purifications using the ÄKTApurifier 100 chromatography system were performed by Dr. Mindaugas Zaveckas. The yield of purified *P. pastoris*-produced rPen m 4 was approximately 17 mg/L (Fig. 3.7a lane 3).

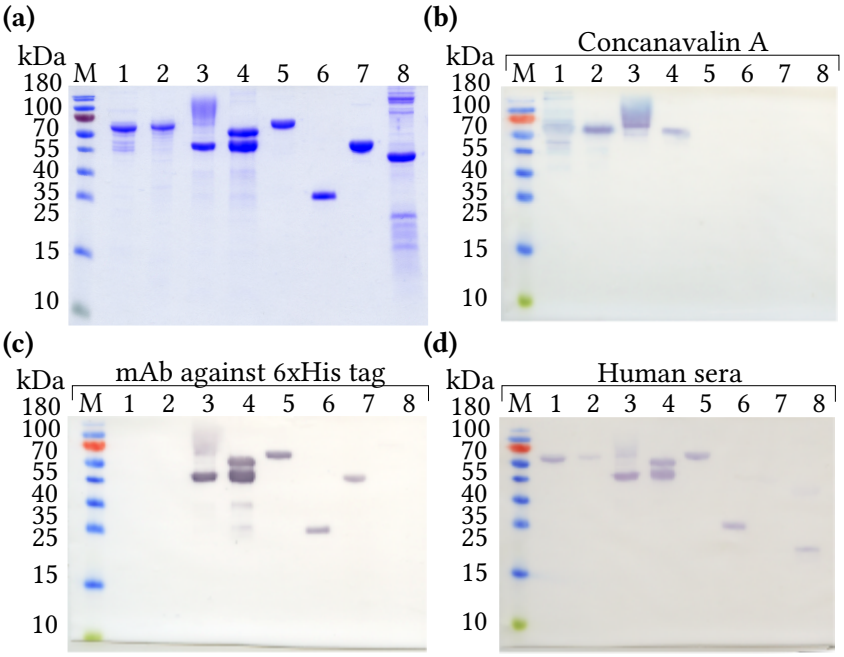


Figure 3.7: Analysis of rPen m 4 variants by **(a)** SDS-PAGE and **(b, c, d)** WB (annotated by the used lectin (b) mAb (c) or positive human sera (d) designation). Lanes: *P. pastoris*-produced rMBP-Pen m 4 purified by [1] anion and [2] cation exchange chromatography; [3], *P. pastoris*-produced rPen m 4; [4], deglycosylated *P. pastoris*-produced rPen m 4; [5], *E. coli*-produced rMBP-Pen m 4; [6], *E. coli*-produced rPen m 4; [7], *E. coli*-produced rMBP with a 6xHis tag; [8], *P. monodon* allergen extract. M – PageRuler™ Prestained Protein Ladder #26617. Proteins were separated in 14% SDS-PAGE gels. (b) Glycosylation analysis was performed with the Con A lectin; (d) IgE analysis was performed with double-positive human sera, which was positive to *P. monodon* allergen extract and the Pen m 4 allergen.

WB with mAb against the 6xHis tag verified the existence of a His-tagged protein (Fig. 3.7c lane 3). Specific reactivity was observed with 45 kDa to 60 kDa protein bands, which did not correspond to the theoretical rPen m 4 MW of 23 kDa. The higher MW suggests that *P. pastoris*-produced rPen m 4 could have undergone hyperglycosylation (Singh and Bhalla 2006; Ahmad et al. 2014). Plasmid verification by sequencing, new transformations, and testing different colonies and growth conditions did not influence the obtained

results.

We evaluated whether these findings would apply to recombinant Pen m 4 produced in other yeast systems. However, rPen m 4 purified from *Kluveromyces lactis* WGI (wild-type), WSS (dolichol kinase mutant), WSSΔ*pep*4 (*pep*4 peptidase mutant) strains (Ziogiene et al. 2019) had an apparent MW of approximately 23 kDa – consistent with the predicted MW of rPen m 4 (Fig. 3.8), which suggested that the observed effects are specific to *P. pastoris*. All experiments with *K. lactis*-produced rPen m 4 prior to purification were done by Dr. Danguolė Žiogienė and Andrius Burdulis, purifications with ZYMO Research columns (see 2.2.19) were performed by Juta Dvareckienė.

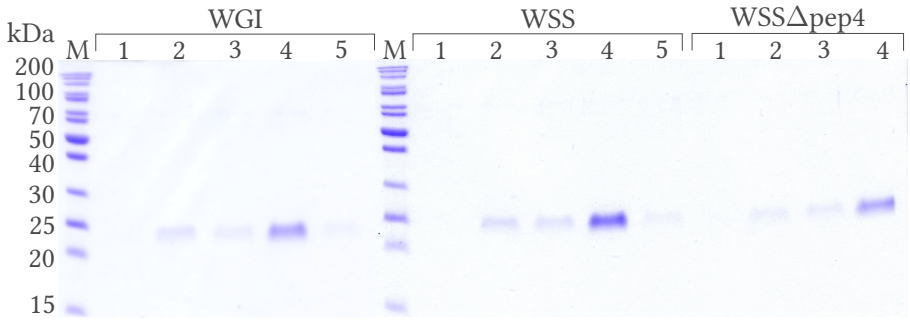


Figure 3.8: Analysis of *K. lactis*-produced rPen m 4 by SDS-PAGE. *K. lactis* strains are indicated above the respective lanes. Lanes: [1], wash fraction; [2-5], elution fractions. M – PageRuler™ Unstained Protein Ladder #26614. Proteins were separated in 14% SDS-PAGE gels. The pKDSU-PGK1-Pen m 4 plasmid construction and the following *K. lactis* strain transformation were performed by Dr. Danguolė Žiogienė, and the transformed cultures were grown by Andrius Burdulis. Purification was carried out via ZYMO His-Spin Protein MiniPrep™.

The Pen m 4 gene is predicted to have two O-glycosylation sites and one N-glycosylation site, as determined using a random forest algorithm and pairwise pattern prediction tool (Hamby and Hirst 2008). To evaluate whether the increased molecular mass of *P. pastoris*-produced rPen m 4 could be related to glycosylation, the protein was treated with PNGase F and analyzed using a Con A lectin assay (Fig. 3.7b lane 4) (see 2.2.27, 2.2.28).

The glycosylation assay confirmed that the protein was mostly N-glycosylated. Residual Con A activity after PNGase F treatment suggested that not all carbohydrate-associated modifications were removed. The residual signal may reflect incomplete deglycosylation with PNGase F, O-glycosylation, or

glycation (de Oliveira et al. 2016). WB analysis of the deglycosylated sample with mAb against the 6xHis tag revealed four distinct protein bands of approximately 55 kDa, 47 kDa, 35 kDa, and 23 kDa sizes (Fig. 3.7c lane 4). The smallest band corresponded to the calculated MW of rPen m 4.

These results suggest that the broad 45-65 kDa signal observed for rPen m 4 is at least partly explained by heterogeneous glycosylation in the *P. pastoris* expression system. However, glycosylation alone does not fully explain the persistence of higher molecular mass forms after deglycosylation. Therefore, stable dimer formation is another likely contributing factor.

Natural shrimp SCPs are known to occur as dimers composed of two polypeptide chains ($\alpha\alpha$, $\alpha\beta$, and $\beta\beta$) and Pen m 4 is a member of the α chain group (Ayuso et al. 2009). In addition, transient dimer formation appears to be common among allergens; one study reported that approximately 80% of analyzed allergens can form transient dimers at high local protein concentrations reached through cellular colocalization (Rouvinen et al. 2010). ESI-MS analysis (see 2.2.29) of purified rPen m 4 revealed the dominant protein peak of 47 kDa MW (Fig. 3.9), which is close to the calculated MW of a Pen m 4 homodimer and approximately 1.3 kDa higher than expected. The difference may be explained by the residual carbohydrate modifications after PNGase F treatment.

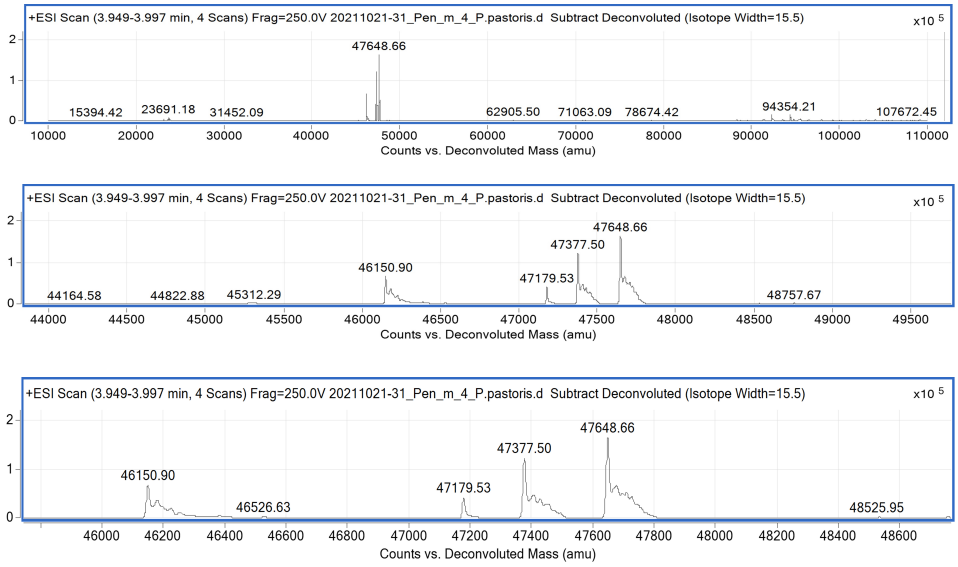


Figure 3.9: ESI-MS analysis spectrum of *P. pastoris*-produced rPen m 4. MS analysis was conducted by Dr. Audronė Rukšėnaitė.

Native SCP dimers dissociate in the presence of SDS or urea to monomers (Wnuk and Jauregui-Adell 1983; Ayuso et al. 2009). In contrast, the 47 kDa band remained detectable after incubation under several denaturing or reducing conditions, including 4-6 M urea, 4-6 M GuaHCl, 50-250 mM DTT, 100 mM EDTA, and heating at 100 °C for 30 min. This suggests that *P. pastoris*-produced rPen m 4 may form unusually stable dimers. Glycosylation or other carbohydrate-associated modifications may contribute to the altered migration and/or stability of these forms. Similar dimerization has been reported for other proteins produced in *P. pastoris*, for example the FKBP12 and PDIA2 proteins (Z Li et al. 2010; AK Walker et al. 2013).

Glycation was also considered as a possible contributor to the observed electrophoretic pattern. The Pen m 4 sequence contains a relatively high proportion of lysine residues (7.2% of the total amino acid content), which may favor non-enzymatic glycation reactions. In addition, protein smearing above the main bands and residual Con A reactivity after PNGase F treatment are consistent with the possibility of covalently attached carbohydrate residues (Lima and Baynes 2013; de Oliveira et al. 2016). However, the current study could not confirm the exact PTM responsible for the higher MW forms.

Despite this heterogeneity, *P. pastoris*-produced rPen m 4 retained sIgE reactivity. Both glycosylated and deglycosylated rPen m 4 samples reacted with sera from *P. monodon*-allergic patients positive for Pen m 4, while no non-specific cross-reactivity with unrelated antibodies was observed (Fig. 3.7d lanes 3 and 4). Therefore, strain 2, shown in Fig. 3.6b lane 3, was selected for further experiments. Overall, these findings indicate that *P. pastoris*-produced rPen m 4 is a stable, IgE-reactive recombinant allergen, although glycosylation controls should be included when using this protein in diagnostic applications.

3.2.6 Purification and characterization of *P. pastoris*-produced rMBP-Pen m 4

Comparative purification of rMBP-Pen m 4 fusion protein was performed using two methods: cation exchange (CEX) and anion exchange (AEX) chromatography (see 2.2.22). In both cases, the extracellular medium containing the secreted rMBP-Pen m 4 was dialyzed (see 2.2.20) with the appropriate buffer before being subjected to ion-exchange FPLC on either SP Sepharose Fast Flow (CEX) or Q Sepharose Fast Flow (AEX) columns. Fractions eluted with NaCl were pooled and analyzed by SDS-PAGE and ImageJ 1.50b software (see 2.2.24). Purifications using the ÄKTApurifier 100 chromatography system

were performed by Dr. Mindaugas Zaveckas.

Interestingly, while target protein purity was ascertained to be the same after both purifications (Fig. 3.10a lanes 2 and 3; Fig. 3.7a lanes 1 and 2), glycosylation analysis (see 2.2.27, 2.2.28) showed completely different glycosylation profiles. The main band of rMBP-Pen m 4 purified via AEX (Fig. 3.10b lane 2; Fig. 3.7b lane 1) was mostly non-glycosylated. Lower MW glycosylated bands comprised the splintered glycosylated MBP (Fig. 3.10c lane 2). Contrary to these results, analysis of rMBP-Pen m 4 purified by CEX exposed the glycosylation of only the main protein band (Fig. 3.10b lane 3; Fig. 3.7b lane 2). The yield of *P. pastoris*-produced rMBP-Pen m 4 purified via AEX and CEX was 39.5 mg/L and 1.5 mg/L, respectively.

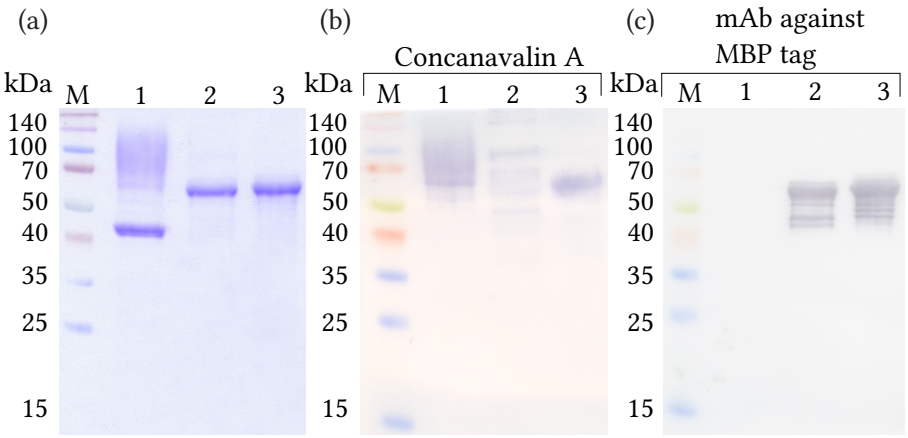


Figure 3.10: Analysis of purified *P. pastoris*-produced rPen m 4 and rMBP-Pen m 4 proteins by **(a)** SDS-PAGE and **(b, c)** WB (annotated by the used mAb or lectin designation). Lane: [1], rPen m 4; [2], rMBP-Pen m 4 purified via AEX; [3], rMBP-Pen m 4 purified via CEX. M – Spectra™ Multicolor Broad Range Protein Ladder. **(b)** WB glycosylation analysis was performed with Con A lectin; **(c)** WB analysis was performed using mAb against the MBP tag. Proteins were separated in 14% SDS-PAGE gels.

ESI-MS analysis (see 2.2.29) of purified rMBP-Pen m 4 proteins revealed that the MW of the proteins was 2.28-2.29 kDa lower than the calculated theoretical MW (Fig. 3.11). This distinction was also confirmed by WB analysis with mAb against the 6xHis-tag, as *P. pastoris*-produced rMBP-Pen m 4 proteins did not react (Fig. 3.7c lanes 1 and 2). In contrast, specific reactivity was observed in WB with mAb against the MBP tag (Fig. 3.10c lanes 2 and 3; Fig. 3.6) and *P. monodon* positive human sera (Fig. 3.7d lanes 1 and 2). Taken together, these results suggest proteolytic removal of the N-terminal 6xHis-containing

region of the 6xHis-MBP-TEV-Pen m 4 construct, while the MBP and Pen m 4 portions remained detectable.

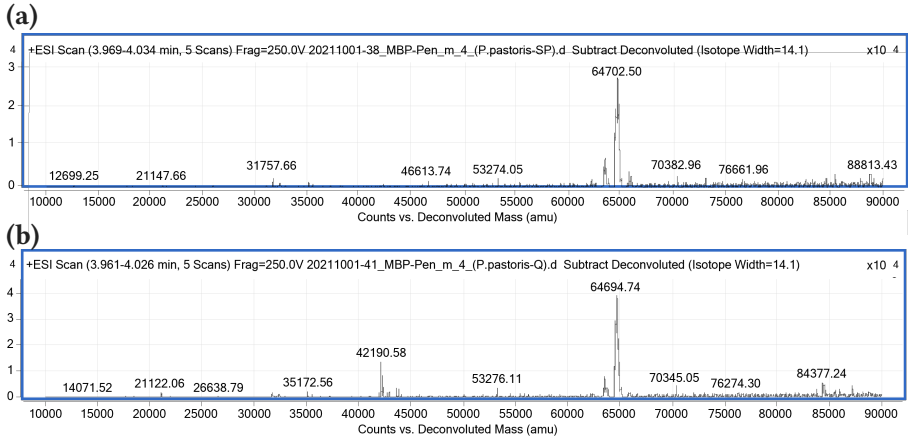


Figure 3.11: ESI-MS analysis spectrum of *P. pastoris*-produced rMBP-Pen m 4, purified by **(a)** CEX and **(b)** AEX chromatography. MS analysis was conducted by Dr. Audronė Rukšėnaitė.

In addition to this N-terminal processing, an MBP-reactive proteolysis product ($\sim$ 45 kDa) was detected by WB with mAb against the MBP tag (Fig. 3.6a lane 2). Because this fragment retained MBP reactivity but was smaller than the intact rMBP-Pen m 4 fusion protein, it most likely represents a second cleavage pattern affecting regions C-terminal to the MBP domain, such as the linker/spacer region or the fused Pen m 4 cargo sequence. This observation is consistent with previous reports of MBP-fusion protein degradation in *P. pastoris*, where a putative MBP-associated protease was proposed to recognize the three-dimensional structure of the fusion protein rather than a specific amino acid sequence (Z Li et al. 2010; Moua et al. 2016). The exact cleavage site was not mapped in the present study; therefore, this interpretation is based on the apparent molecular mass of the fragment, MBP-specific WB reactivity, and comparison with published MBP-fusion protein degradation patterns.

Although these results indicate partial proteolytic processing of rMBP-Pen m 4, the MBP-fusion strategy was still evaluated as a production approach because the fusion protein remained detectable, purifiable, and immunoreactive. Previous studies have shown that MBP fusion can improve the yield and secretion efficiency of recombinant proteins in *P. pastoris*; however, this effect is target-dependent, and empirical optimization is required for each protein

(Moua et al. 2016).

P. pastoris-produced rMBP-Pen m 4 was purified using cation and anion exchange chromatography, yielding 1.5 mg/L and 39 mg/L, respectively. The significant difference between protein yields highlights the importance of comparative studies when designing production and purification strategies. The protein remained detectable after purification under both acidic and alkaline conditions; however, samples purified using different ion-exchange resins showed distinct glycosylation profiles. Despite the observed proteolytic processing, fusion with MBP increased the yield of rPen m 4 2.3-fold, making it the most successful strategy tested for producing rPen m 4 in the eukaryotic expression system.

3.2.7 Immunoreactivity of rPen m 4 and rMBP-Pen m 4

Four groups of patient plasma were used to test recombinant Pen m 4 proteins (Table 3.2): (i) twelve individual human plasma samples tested positive for Pen m 4 sIgE; (ii) Pool-positive, pool composed of eight human plasma samples, tested positive for *P. monodon* allergen extract but negative for Pen m 4 sIgE; (iii) Pool-negative, pool composed of eight human plasma samples, tested negative for IgE, and (iv) one hundred individual *P. monodon* allergy negative human plasma samples. Two standard allergy diagnostic platforms were used to accurately compare the recombinant proteins: indirect ELISA and reverse-phase protein microarray.

Table 3.2: Characteristics of human plasma and serum samples used in the assays.

	No. of samples	*P. monodon* sIgE	Pen m 4 sIgE	Other allergens	Serum pool
i.	12	Positive[1,2]	Positive[1,2]	Positive[1,2]	Double-positive pool
ii.	8	Positive[1,2]	Negative[1,2]	Positive[1,2]	Positive-pool
iii.	8	Negative[1]	Negative[1]	Negative[1]	Negative-pool
iv.	100	Negative[1,2]	Negative[1,2]	Positive[1,2]	

[1] IgE ALEX[2]® Allergy test — sIgE ≥ 0.30 kU/L = positive; < 0.30 kU/L = negative.
[2] ImmunoCAP™ or Phadia UniCap™ assays — sIgE ≥ 0.35 kU/L = positive; < 0.35 kU/L = negative.

First, the immunoreactivity of rMBP-Pen m 4, produced in *P. pastoris* was evaluated. Indirect ELISA (see 2.2.30, 2.2.32) results showed that rMBP-Pen m 4 purified via AEX reacted significantly more strongly with sIgE than rMBP-Pen m 4 purified by CEX (Fig. 3.12). A paired Student's t-test with 12

observations (each measured in triplicate) supported the statistical significance with $p = 4.3684 \times 10^{-5}$. These results imply that purification at pH 7.5 (AEX conditions), close to the physiological pH, helps the protein retain its native conformation, preserving the conformational epitopes recognized by sIgE.

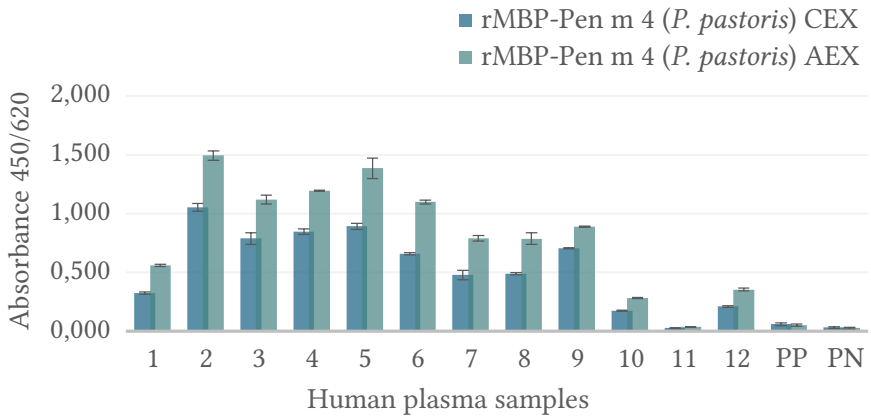


Figure 3.12: Immunoreactivity analysis of *P. pastoris*-produced rMBP-Pen m 4 proteins in an indirect ELISA platform. CEX – represents rMBP-Pen m 4 purified through cation exchange chromatography, and AEX – represents rMBP-Pen m 4 purified through anion exchange chromatography. [1-12] correspond to individual plasma from *P. monodon*-sensitized patients positive for Pen m 4; PP – pool-positive (*P. monodon* sIgE-positive, but Pen m 4 sIgE-negative, n = 8); PN – pool-negative (n = 8).

Analysis of the indirect ELISA results (Fig. 3.13) showed that all synthesized rPen m 4 proteins specifically bound Pen m 4 sIgE in human plasma samples with no evidence of non-specific interactions. One-way repeated measures ANOVA was performed to determine whether recombinant Pen m 4 proteins exhibited significant differences in IgE reactivity across sera. Positive and negative controls were excluded from the statistical analysis and pairwise comparison to avoid artificially inflating variability. The ANOVA analysis revealed significant differences among proteins $F(4,\ 44) = 41.28,\ p < 0.001$, indicating variability in reactivity. A one-way ANOVA was also performed on the technical replicates to assess the consistency of all measurements. The analysis yielded a low F-value and high p-value $F(2,\ 213) = 0.014,\ p = 0.986$, indicating no significant variability between the technical replicates and confirming high measurement consistency. Fig. 3.13 shows that *E. coli*-produced proteins exhibited higher sensitivity when tested with serum 11, which had a very low sIgE count. However, *E. coli*-produced rPen m 4 also showed higher

3,500

non-specific interaction with the Pool-positive sera (PP sera).

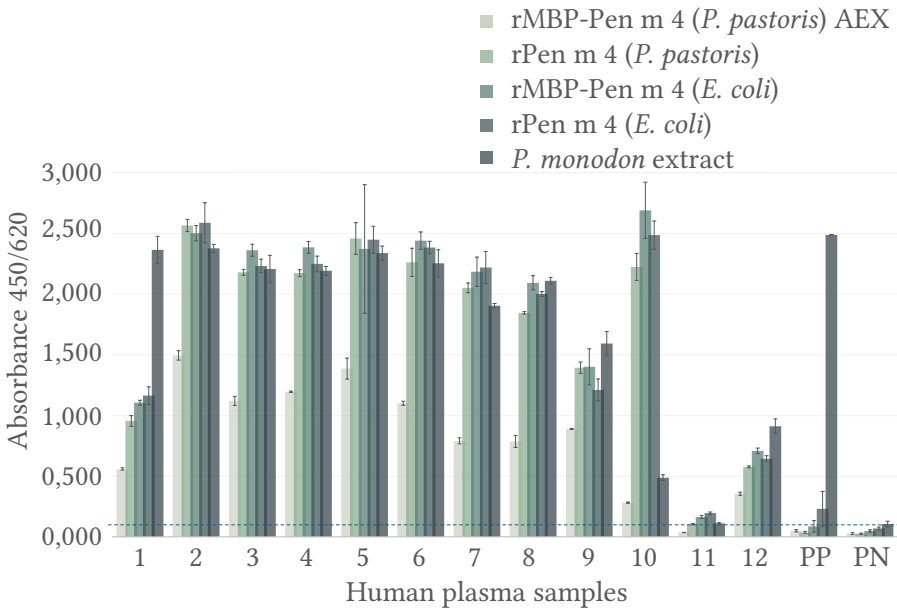


Figure 3.13: Immunoreactivity analysis of rPen m 4 variants in an indirect ELISA platform. AEX – represents rMBP-Pen m 4 purified through anion exchange chromatography. [1-12] correspond to individual plasma from *P. monodon*-sensitized patients positive for Pen m 4; PP – pool-positive (*P. monodon* sIgE-positive, but Pen m 4 sIgE-negative, n = 8); PN – pool-negative (n = 8). The dotted line represents the cut off.

Post hoc comparisons were conducted using the Bonferroni correction, which revealed specific protein pairs with significant differences in their reactivity (Fig. 3.14, Table 3.3). *E. coli*-produced rMBP-Pen m 4 and rPen m 4, and *P. pastoris*-produced rPen m 4 differed significantly from both *P. pastoris*-produced rMBP-Pen m 4 (CEX and AEX). Additionally, rMBP-Pen m 4 purified via AEX differed significantly from rMBP-Pen m 4 purified via CEX. *P. pastoris*-produced rMBP-Pen m 4 proteins had the lowest median reactivity, which is a disadvantage for diagnostic test sensitivity but an advantage if such recombinant proteins were to be used in AIT, as they would elicit a lower immune response. Other protein variants reacted with similar intensity, with *E. coli*-produced rMBP-Pen m 4 having the highest median reactivity, but not high enough to be significantly different.

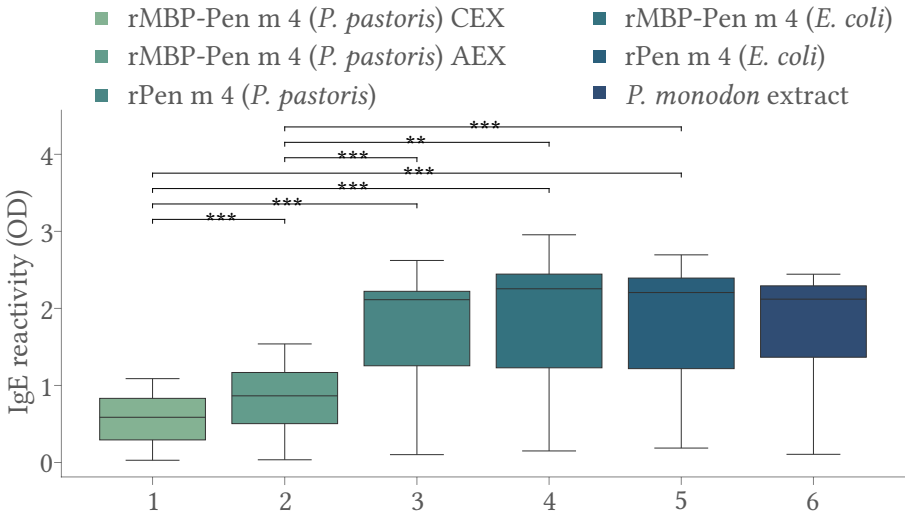


Figure 3.14: IgE reactivity of patient plasma with recombinant Pen m 4 proteins visualized in a boxplot. Lane: [1], *P. pastoris*-produced rMBP-Pen m 4 purified through cation exchange chromatography (CEX) and [2] – through anion exchange chromatography (AEX); [3], *P. pastoris*-produced rPen m protein; [4], *E. coli*-produced rMBP-Pen m 4; [5], *E. coli*-produced rPen m 4; [6], *P. monodon* allergen extract. The data represents normalized IgE reactivity across 13 individual plasma from *P. monodon*-sensitized patients positive for Pen m 4. Reactivity is represented as normalized OD values, and significant differences between proteins are annotated with asterisks ($** = p < 0.01$, $*** = p < 0.001$).

Table 3.3: Pairwise comparisons after repeated measures ANOVA and Bonferroni-corrected post hoc analysis.

A	System A	B	System B	p-unc	p-corr	Reject H_0
rMBP Pen m 4	*E. coli*	rPen m 4	*E. coli*	0.1305	1.0000	False
rMBP Pen m 4	*E. coli*	rPen m 4	*P. pastoris*	0.0093	0.0932	False
rMBP Pen m 4	*E. coli*	rMBP Pen m 4, AEX	*P. pastoris*	0.0001	0.0010	True
rMBP Pen m 4	*E. coli*	rMBP Pen m 4, CEX	*P. pastoris*	0.0000	0.0003	True
rPen m 4	*E. coli*	rPen m 4	*P. pastoris*	0.0256	0.2557	False
rPen m 4	*E. coli*	rMBP Pen m 4, AEX	*P. pastoris*	0.0001	0.0010	True
rPen m 4	*E. coli*	rMBP Pen m 4, CEX	*P. pastoris*	0.0000	0.0003	True
rPen m 4	*P. pastoris*	rMBP Pen m 4, AEX	*P. pastoris*	0.0001	0.0009	True
rPen m 4	*P. pastoris*	rMBP Pen m 4, CEX	*P. pastoris*	0.0000	0.0003	True
rMBP Pen m 4, AEX	*P. pastoris*	rMBP Pen m 4, CEX	*P. pastoris*	0.0000	0.0004	True

p-unc – uncorrected p-value, tells the significance without accounting for multiple comparisons; p-corr – Bonferroni-corrected p-value, calculated by multiplying the p-unc by the number of comparisons, it controls the family-wise error rate to reduce false positives; Reject H_0 – reject the null hypothesis, this indicates whether the null hypothesis (no difference between the two groups) is rejected after correction. True if p-corr < 0.005 – significant difference between the pair, False if p-corr $\geqslant 0.005$ – no significant difference.

Reverse-phase protein microarray analysis (see 2.2.31, 2.2.34) supports the indirect ELISA findings, which show that *E. coli*-produced Pen m 4 variants bind more readily to sIgE Ab. Interestingly, ELISA results did not show high differences between the *E. coli*-produced Pen m 4 variants' sensitivity; however, the protein microarray revealed that *E. coli*-produced rMBP-Pen m 4 outperformed *E. coli*-produced rPen m 4 MFI levels.

All MFI values were $\log_2$-transformed to stabilize variance and reduce right skew. This compression of high intensities made the overall distribution more Gaussian-like and evened the spread of values. The $\log_2$ transformation also makes fold-change interpretation easier: each one-unit increase in the $\log_2$ scale translates to a doubling of MFI intensity values.

Reverse-phase protein microarray experiment not only tested which Pen m 4 variant was more suited for *P. monodon* Pen m 4 sIgE detection, but also at which concentration the highest sensitivity was achieved without losing specificity (see 2.2.33). The decision of which concentration to choose was also impacted by our protein microarray optimization experiment with over 100 sera samples from individuals with a negative *P. monodon* allergen extract sIgE test result (sIgE $\leqslant$ 0.3 kU/L) but positive to other allergies. This experiment showed that at concentrations greater than 0.25 mg/mL, false-positive results were observed from all four recombinant proteins. Lower concentrations than 0.125 mg/mL showed drastically reduced signals. Also, a case in which a patient was only sensitized to the Pen m 4 allergen but not sensitized (sIgE $\leqslant$ 0.1 kU/L) to the Pen m 1 and Pen m 2 allergens was noted, reflecting previously published data, that showed that Pen m 4 is capable of mono-sensitization (Sogo et al. 2018). Also, some false-positive results were observed with the *P. pastoris*-produced rPen m 4 variants. These false positives could indicate that yeast glycosylation patterns could interfere with the correct diagnosis if the diagnostic array does not contain CCD markers (Kozlov et al. 2025).

Considering all factors, 0.25 mg/mL concentration was chosen to illustrate the microarray results. As shown in Fig. 3.15, rMBP-Pen m 4 (*E. coli*) exhibited the highest sensitivity in 66% of plasma, with rPen m 4 (*E. coli*) being a close second (33%). Interestingly, reactions with the patient plasma sample 11 remained below the cut-off threshold in the microarray experiment as opposed to the ELISA experiment (Fig. 3.13). The false negative result showed that while protein microarray shares a similar underlying design with ELISA, its higher sensitivity comes at the cost of increased background noise, which can obscure true positives.

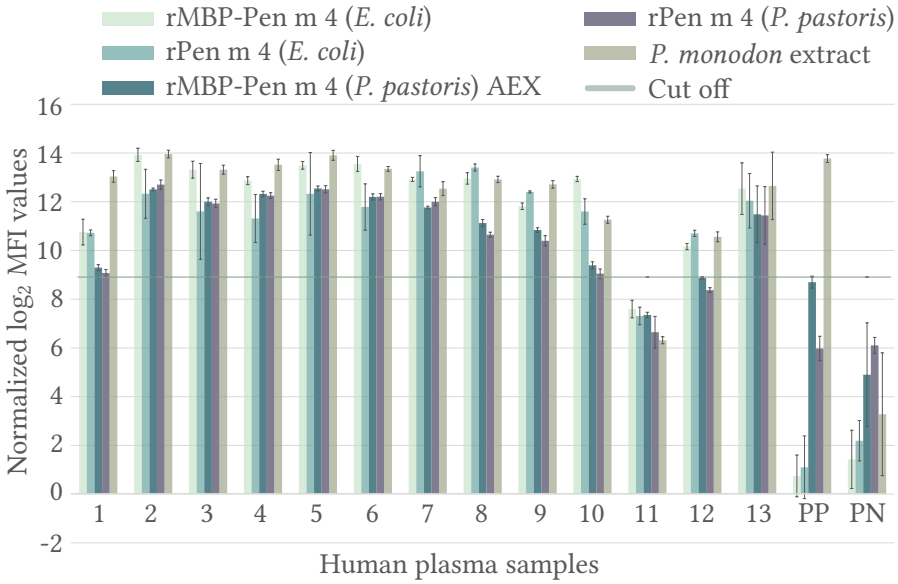


Figure 3.15: IgE reactivity of patient sera with rPen m 4 proteins and the *P. monodon* allergen extract in a reverse-phase protein microarray platform. The graph depicts the results of proteins spotted at concentrations of 0.25 mg/mL. [1-13] are individual sera from *P. monodon* sensitized patients positive for Pen m 4; PP – pool-positive (*P. monodon* sIgE-positive, but Pen m 4 sIgE-negative, n = 8); PN – pool-negative (n = 8). The y-axis displays normalized $\log_2$ MFI values; each one-unit difference on this scale corresponds to a two-fold change in raw MFI intensity. The cut off is shown in the figure.

A linear mixed-effects model (LMM) was performed on the raw normalized $\log_2$ MFI values to analyze these results further. LMM was chosen considering that some replicates were missing after spot quality screening. Controls were filtered out from the analysis but kept in for boxplot visualization. LMM analysis confirmed that protein construct significantly influenced reactivity to plasma (likelihood-ratio $\chi^2(3) = 88.78$, $p = 1.77 \times 10^{-18}$), which meant that proteins exhibited significant differences in their IgE binding. The highest mean reactivity was observed from rMBP-Pen m 4 (*E. coli*), with a value of 12.05. The following post hoc Bonferroni-corrected analysis showed significant differences between all *E. coli* and *P. pastoris* pairs, but no significant differences within the same systems (Fig. 3.16, Table 3.4). This showed that although rMBP-Pen m 4 (*E. coli*) reacted more strongly than rPen m 4 (*E. coli*), the difference was not significant. However, taking these results together with the achieved protein production rates, *E. coli*-produced rMBP-Pen m 4 proves to be the optimal choice for allergy diagnostics.

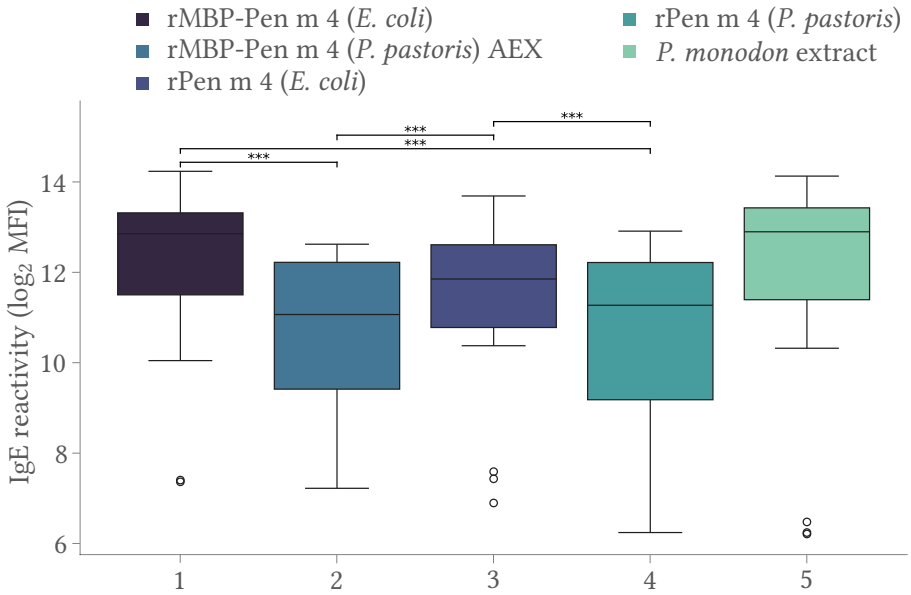


Figure 3.16: IgE reactivity of patient plasma with recombinant Pen m 4 proteins visualized in a boxplot. Lanes: [1], *E. coli*-produced rMBP-Pen m 4; [2], *P. pastoris*-produced rMBP-Pen m 4 purified via AEX; [3], *E. coli*-produced rPen m 4; [4], *P. pastoris*-produced rPen m 4 protein; [5], *P. monodon* allergen extract. The data represents normalized $\log_2$ MFI values across 13 individual plasma from *P. monodon*-sensitized patients positive for Pen m 4. Significant differences between proteins are annotated with asterisks ($*** = p < 0.001$).

Table 3.4: Bonferroni-corrected pairwise comparisons of mean serum reactivities.

A	System A	B	System B	p-unc	p-corr	Reject H_0
rMBP-Pen m 4	*E. coli*	rMBP-Pen m 4	*P. pastoris*	0.0000	0.0000	True
rMBP-Pen m 4	*E. coli*	rPen m 4	*E. coli*	0.0737	0.4420	False
rMBP-Pen m 4	*E. coli*	rPen m 4	*P. pastoris*	0.0000	0.0000	True
rMBP-Pen m 4	*P. pastoris*	rPen m 4	*E. coli*	0.0000	0.0000	True
rMBP-Pen m 4	*P. pastoris*	Pen m 4	*P. pastoris*	0.3615	1.0000	False
rPen m 4	*E. coli*	rPen m 4	*P. pastoris*	0.0000	0.0000	True

p-unc – uncorrected p-value, tells the significance without accounting for multiple comparisons; p-corr – Bonferroni-corrected p-value, calculated by multiplying the p-unc by the number of comparisons, it controls the family-wise error rate to reduce false positives; Reject H_0 – reject the null hypothesis, this indicates whether the null hypothesis (no difference between the two groups) is rejected after correction. True if p-corr < 0.005 – significant difference between the pair, False if p-corr $\geqslant 0.005$ – no significant difference.

Interestingly, while fusion with MBP increased recovered protein mass, it did not increase the calculated molar yield of allergen-containing molecules

(Table 3.5). This distinction is important because the immunological assays were performed using equal protein mass for coating, as is common in ELISA and protein microarray workflows. Under these mass-normalized conditions, MBP-fusion proteins contained fewer allergen-containing molecules than their non-fused counterparts because of their higher molecular mass.

Table 3.5: Comparison of mass and molar yields of non-fused and MBP-fused recombinant proteins.

System	Protein	MW (kDa)	Yield (mg/L)	Mass ratio**	Molar yield* (µmol/L)	Molar ratio**
E. coli	rPen m 4	23	52	1.27	2.26	0.44
E. coli	rMBP-Pen m 4	66	66		1.00	
P. pastoris	rPen m 4	23	17	2.29	0.74	0.79
P. pastoris	rMBP-Pen m 4	67	39.5		0.59	

* Molar yield was calculated as yield (mg/L) divided by calculated MW (kDa).
** Ratios indicate the effect of MBP fusion within each expression system.

For the *E. coli*-produced variants, lower molar concentration did not result in reduced IgE signal, as rPen m 4 and rMBP-Pen m 4 did not differ significantly in either ELISA or microarray analysis. This indicates that the MBP-fusion format retained strong apparent IgE reactivity despite containing fewer Pen m 4 molecules at the same coated mass. In contrast, *P. pastoris*-produced rMBP-Pen m 4 showed no significant reduction in ELISA but produced lower signals in the microarray assay. Since the same molar-loading disadvantage did not reduce the signal of the *E. coli*-produced MBP fusion, the weaker microarray reactivity of *P. pastoris*-produced rMBP-Pen m 4 is more likely related to construct- or expression system-specific protein properties. Conversely, a comparable or higher signal despite fewer allergen-containing molecules may reflect favorable properties of the MBP-fusion format, such as improved adsorption or retention on the assay surface, better preservation or exposure of IgE epitopes, altered surface orientation, or improved folding (Crowther 2000).

3.2.8 CCD interference in IgE binding to rPen m 4

Glycosylation patterns that do not include classical CCD markers, such as mannosylation seen in yeast, are unclear in their interaction with anti-CCD IgE (more information in 1.4.3). To understand whether anti-CCD IgE could have influenced the results of our assays, CCD interference analysis was

performed on the rPen m 4 variants produced by *P. pastoris*. In this experiment, HRP acted as a marker for aberrant glycosylation (Homann et al. 2017).

The normalized $\log_2$ MFI values for uninhibited and inhibited sera were paired and tested for normality using the Shapiro-Wilk test. If the Shapiro-Wilk yielded $p \geqslant 0.05$, a paired Student's t-test was applied; otherwise, a Wilcoxon signed-rank test was used. Raw p-values for the tested proteins were adjusted to control the false discovery rate (FDR = 5%) via the Benjamini-Hochberg procedure. A two-way repeated-measures ANOVA was fitted with the tested proteins and HRP as within-subject factors and plasma as the subject to evaluate the overall inhibitor effect. Significance thresholds were set at α = 0.05. In the microarray results (Fig. 3.17), a slight reduction of MFI signals in most plasma samples was observed; however, some signals intensified in HRP-inhibited sera. Statistical analysis showed no significant difference between HRP-inhibited and uninhibited plasma groups, except in the case of HRP, which was added as a positive control for glycan sIgE.

However, false positive results were observed with the *P. pastoris*-produced rPen m 4 variants when tested with sera from individuals with a negative *P. monodon* allergen extract sIgE test result (sIgE $\leqslant$ 0.3 kU/L) but positive for other allergies. Together with the findings of Kozlov et al. 2025, these results suggest that yeast-derived glycosylation patterns may contribute to false-positive IgE signals and could interfere with correct diagnosis if diagnostic arrays do not include CCD markers. In most cases, mannosylation does not significantly affect sIgE binding; false positives are seen in patients with high anti-CCD IgE antibody levels. Still, such patients can have false positive results even with ImmunoCAP assays, where anti-CCD IgE react with the ImmunoCAP cellulose matrix. All these cases can be negated by CCD blockers used before *in vitro* IgE assays (Hemmer et al. 2018; Kozlov et al. 2025). Alternatively, mutating N-glycosylation sites before expression in yeast can also be an option for *in vitro* diagnosis. However, for AIT, *P. pastoris* glycosylation could still be useful as mannosylation reflects the native glycosylation pattern of many common allergens, and it enables production of a recombinant allergen closely related to its native form (Al-Ghouleh et al. 2012).

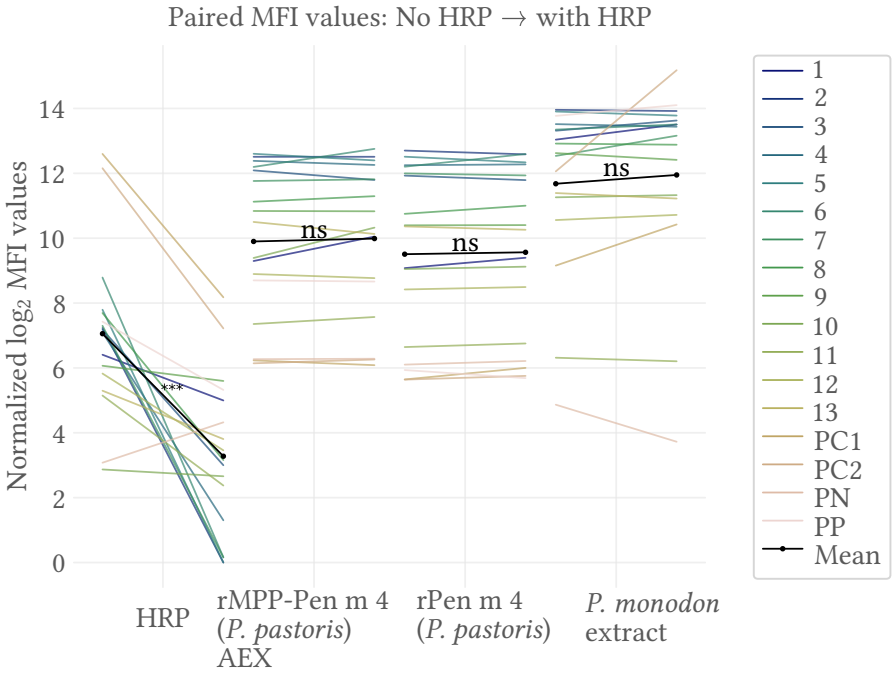


Figure 3.17: CCD interference analysis, which is visualized in a line graph. *P. pastoris*-produced rMBP-Pen m 4 and rPen m 4 binding to sIgE was analyzed in HRP-uninhibited and HRP-inhibited plasma, analyzing whether the addition of a carbohydrate determinant marker had a significant difference in MFI signals. *P. monodon* allergen extract acts as a positive control for positive sera; HRP was used as a carbohydrate determinant marker. [1-13] - individual plasma from *P. monodon*-sensitized patients positive for Pen m 4; PC1 and PC2 – plasma, positive for glycan sIgE; PP – pool-positive (n = 8); PN – pool-negative (n = 8). Protein mixtures were spotted at 0.25 mg/mL concentrations and HRP at 0.05 mg/mL. The y-axis displays normalized $\log_2$ MFI values. Significant differences between inhibited and uninhibited plasma are annotated with asterisks ($ns = p > 0.05, \quad *** = p < 0.001$)

.

3.2.9 IgE reactivity to rPen m 4 in Ca^{2+}-depleted environment

The IgE reactivity of the Pen m 4 allergen is attributed to a mix of conformational and linear IgE epitopes. Like most SCPs, some conformational epitopes are maintained by Ca^{2+}-chelation. A prior study noted that a strong Ca^{2+} chelator negatively affects the diagnostic sensitivity of Pen m 4, as is the case with other EF-hand Ca^{2+}-binding allergens (Morii et al. 2013; Tomura et al. 2008). The underlying problem is that *in vitro* allergy diagnostic tests do not specify what type of blood sample should be used. Most laboratories use

either blood serum or plasma to diagnose and develop allergy test kits. Plasma samples are collected when blood is treated with an anticoagulant, inhibiting blood clot formation. These anticoagulants work based on Ca^{2+}-chelation (EDTA, trisodium citrate, and fluoride) or by inhibition of thrombin (heparin). For example, only serum is usually used when analyzing ionized calcium (iCa) levels in blood, since anticoagulants can alter the sample. If plasma must be used for faster diagnosis, a negative bias needs to be considered; however, the calculated iCa levels correlate poorly with the iCa levels determined using serum analysis (Jafri et al. 2014). *In vitro* allergy diagnostics do not consider these inconsistencies.

This experiment aimed to analyze whether blood collected in blood collection tubes with added EDTA as an anticoagulant would affect *in vitro* rPen m 4 diagnosis. To test this, a Pen m 4-positive serum pool was diluted with buffer supplemented with 5 mM of EDTA, a standard EDTA concentration found in blood collection tubes. Fig. 3.18 shows the significant difference between the two conditions; adding 5 mM EDTA reduced IgE reactivity to rPen m 4 by 61-73%, while the analysis with the *P. monodon* extract showed no significant difference. Further analysis showed that increasing EDTA to 6 mM did not significantly alter IgE reactivity compared to 5 mM. However, increasing the number of replicates led to improved statistical power, revealing significant differences even for the *P. monodon* extract between non-supplemented and supplemented sera.

These results are preliminary and should be validated using blood samples collected directly from allergic patients in different blood collection tubes. Nevertheless, they underscore the need for deeper analysis of recombinant allergens before their implementation in allergy diagnostics.

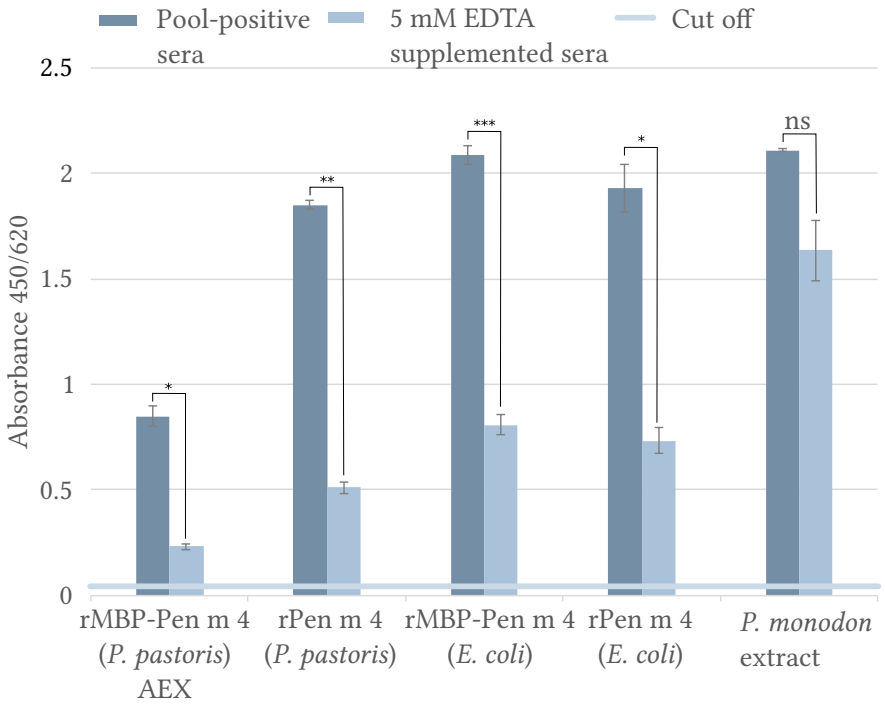


Figure 3.18: Effect of Ca^{2+}-depletion on the IgE reactivity of rPen m 4 proteins and the *P. monodon* allergen extract. ELISA analysis was performed with pooled sera from *P. monodon*-sensitized patients positive for Pen m 4 and *P. monodon* extract (n = 12). Significant differences between the 5 mM EDTA-supplemented and unaltered sera groups are annotated with asterisks ($ns = p > 0.05$, $* = p < 0.05$, $** = p < 0.01$ $* ** = p < 0.001$). The cut off is shown in the figure.

3.3. Recombinant allergen Can f 6 production and immunological characterization

3.3.1 Generation of codon-optimized Can f 6 allergen sequence

The Can f 6 gene was re-engineered with GenScript's OptimumGene™ software workflow to suit the codon-usage bias of the yeast production host. During the redesign, mRNA secondary structure was predicted based on codon compatibility, and key cis-regulatory motifs that influenced transcriptional and translational efficacy were simultaneously balanced. A side-by-side comparison of the native Can f 6 sequence (nCan f 6) with its optimized counterpart (rCan f 6) was summarized in Table 3.6. Both versions fall within the recommended windows for GC content and CFD. The decisive difference was observed in CAI and CFD values: rCan f 6 shows a marked increase and decrease, respectively, approaching the optimal CAI and CFD range for yeast, whereas nCan f 6 remains below that threshold. Based on this – and in line with the Pen m 4 results – rCan f 6 is expected to translate more efficiently and yield higher recombinant protein levels in selected expression systems.

Table 3.6: The CAI, GC content, and CFD *in silico* predictions for native and recombinant Can f 6 constructs in *E. coli* and yeast expression hosts. The ideal values of the predicted elements are shown.

Host / Parameters	***E. coli***		**Yeast**		**Ideal value**
	nPen m 4	*rCan f 6*	*nCan f 6*	*rCan f 6*	
Codon adaptation index (CAI)	0.69	0.72	0.69	0.77	0.8–1.0
GC content (%)	38.63	42.69	38.63	42.69	30–70
Codon frequency distribution (CFD)	9	6	4	0	<30

The numerical values were generated by using the GenScript Rare Codon Analysis Tool (https://www.genscript.com/tools/rare-codon-analysis)

3.3.2 Cloning, expression, and purification of rCan f 6 from *E. coli*

C. familiaris Can f 6 gene encodes a 175-aa-long protein. The 528 bp Can f 6 coding gene was subcloned into two distinct expression vectors (see 2.2): (i) pET28a(+) and (ii) pET28a(+)-MBP-TEV. The goal was to determine if an MBP segment could enhance recombinant allergen expression (Nallamsetty, Austin, et al. 2005). Consistent with Pen m 4 construction (3.2.2), the codon-optimized Can f 6 gene was amplified by PCR (see 2.2.6) using pUV57/Can f 6 plasmid as a

template with primers pET28F_C6 and pET28R_C6 (see 2.1), which introduced the restriction sites of BamHI and XhoI enzymes, respectively. In the first vector, an N-terminal 6xHis tag was added to the Can f 6 gene (Fig. 3.19a). In the second, Can f 6 was integrated with the N-terminal 6xHis-tagged MBP protein coding gene (Fig. 3.19b). Both vectors were verified by sequencing in the VU LSC IBT sequencing center.

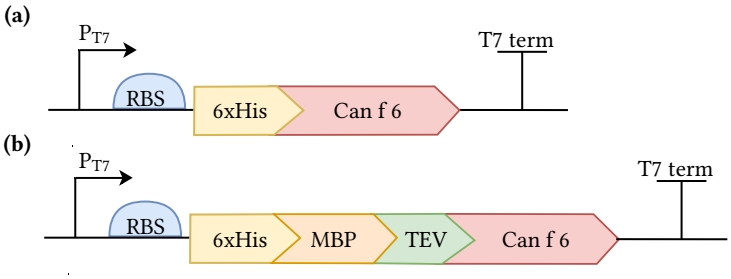


Figure 3.19: Schematic representation of recombinant Can f 6 gene expression cassettes for production in *E. coli*. **(a)** pET28a(+)-Can f 6 expression vector, in which the Can f 6 gene is N-terminally tagged with a 6xHis tag. **(b)** pET28a(+)-MBP-TEV-Can f 6 expression vector, in which the Can f 6 gene is fused to the C-terminal end of a 6xHis-MBP fusion protein coding gene with a TEV protease recognition site between the fusion and passenger protein coding sequences.

Expression of rCan f 6 constructs was demonstrated in *E. coli* strain BL21 (DE3) (see 2.2.8, 2.2.15, 2.2.11). SDS-PAGE (see 2.2.14) analysis revealed distinct bands at 23.7 kDa (Fig. 3.20a) and 64.4 kDa (Fig. 3.20b), corresponding to rCan f 6 and rMBP-Can f 6, respectively. These bands were evident in the soluble fractions after IPTG induction (lane 3 in Fig. 3.20a and 3.20b) but not before (lane 1 in Fig. 3.20a and 3.20b). Following batch purification of 6xHis-tagged proteins from *E. coli* under native conditions (see 2.2.18), both rCan f 6 and rMBP-Can f 6 were isolated in their pure forms (lanes 11-14 in Fig. 3.20a and lanes 10-14 in Fig. 3.20b). The yields post nickel-chelate chromatography were 42 mg/L for rCan f 6 and 78 mg/L for rMBP-Can f 6. rMBP-Can f 6 was detected in the wash fractions, indicating that the purification protocol may require optimization to reduce protein loss.

Consistent with the published literature, rCan f 6 expressed in *E. coli* was soluble. In the absence of MBP fusion, a yield of 42 mg/L was obtained, surpassing previously reported yields of 21 mg/L achieved with the pGEX4T-2 plasmid (Yamamoto et al. 2019) and 25 mg/L using the pET20b vector (Nilsson

et al. 2012). All studies, including this one, have employed the *E. coli* BL21 (DE3) strain. Co-expression with MBP increased the yield to 78 mg/L, positioning this method as the most efficient in *E. coli*-based expression systems.

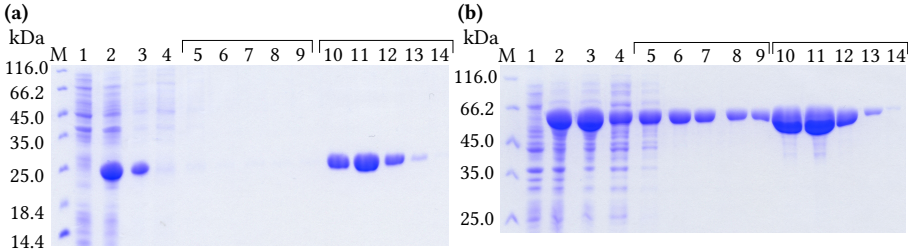


Figure 3.20: Analysis of *E. coli* expression and purification of rCan f 6 **(a)** and rMBP-Can f 6 **(b)** by SDS-PAGE. Lanes: [1], lysate fraction from non-induced *E. coli*; [2], lysate fraction from IPTG-induced *E. coli*; [3], soluble fraction before purification with nickel-chelate chromatography; [4], soluble fraction after incubation with Ni-NTA agarose; [5-9], wash fractions; [10-14], elution fractions containing purified rCan f 6 and rMBP-Can f 6 proteins, respectively. M – Pierce™ Unstained Protein MW Marker #26610 (see note in 2.2.14). Proteins were separated in (a) 12% and (b) 14% SDS-PAGE gels.

3.3.3 Characterization of rCan f 6 variants purified from *E. coli*

Recombinant protein expression was confirmed using WB (see 2.2.26) analysis with mAb targeting 6xHis and MBP tags (Fig. 3.21). Protein stability was evaluated after one month of storage under various conditions. As shown in Fig. 3.21, lyophilization adversely impacted rCan f 6 (Fig. 3.21a lane 1) and rMBP-Can f 6 (Fig. 3.21b lane 1). Short-term storage at $4\,^{\circ}\mathrm{C}$, $-20\,^{\circ}\mathrm{C}$, and $-20\,^{\circ}\mathrm{C}$ supplemented with 40% glycerol did not show significant changes (Fig. 3.21 lanes 2, 3, and 4). However, a follow-up assessment at 24 months (data not presented) showed marked degradation for samples stored at $4\,^{\circ}\mathrm{C}$ and $-20\,^{\circ}\mathrm{C}$ without glycerol. Conversely, samples stored at $-20\,^{\circ}\mathrm{C}$ with 40% glycerol exhibited only marginal degradation.

Experimental data from this study indicate that rCan f 6 undergoes degradation upon lyophilization, irrespective of its MBP fusion. Similar degradation was also observed during prolonged storage at $-20\,^{\circ}\mathrm{C}$ without stabilizing agents like glycerol (data not shown). This result ties in with studies analyzing dog allergen concentration in commercial skin prick test (SPT) solutions, which have shown that allergen profiles can vary significantly depending on the manufacturer, and even between batches from the same supplier. Can f 6

was consistently detected in low quantities across all tested SPT solutions. This is a major concern, as basophil activation testing further showed that a patient mono-sensitized to Can f 6 only reacted to SPT solutions containing the highest amount of Can f 6 (Wintersand et al. 2019; Chan and Leung 2018).

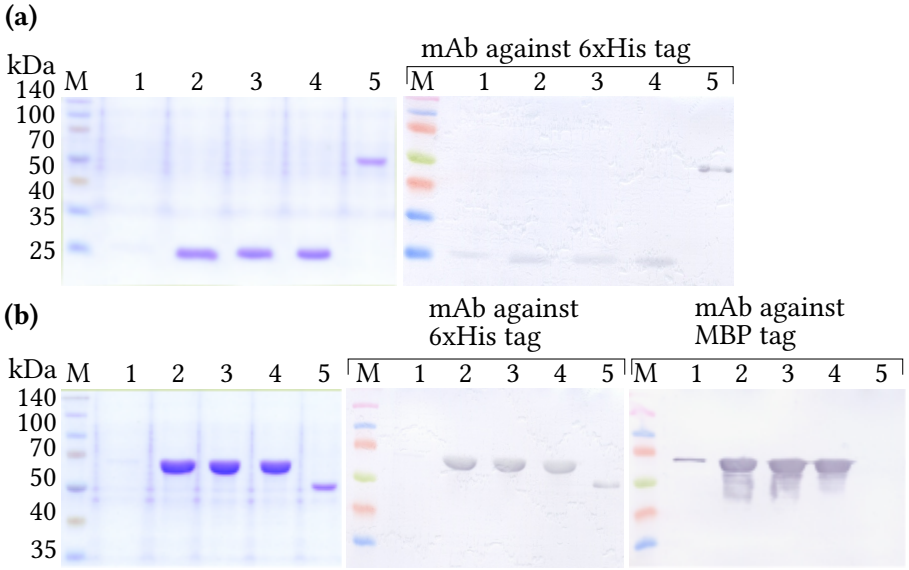


Figure 3.21: Analysis of purified *E. coli*-produced rCan f 6 **(a)** and rMBP-Can f 6 **(b)** proteins by SDS-PAGE and WB (annotated by the used mAb designation). Lanes: [1], protein samples after lyophilization; [2], protein samples after storage at $4\,^{\circ}\mathrm{C}$; [3], protein samples after storage at $-20\,^{\circ}\mathrm{C}$ without glycerol; [4], protein samples after storage at $-20\,^{\circ}\mathrm{C}$ with 40% glycerol; [5], *S. cerevisiae*-produced 6xHis-tagged Hantavirus Sin Nombre nucleocapsid protein (Kucinskaite-Kodze et al. 2011). Protein samples were analyzed after one month of storage. Lane [5] protein acts as a positive control for WB with mAb against the 6xHis tag and a negative control for WB with mAb against the MBP tag. M – Spectra™ Multicolor Broad Range Protein Ladder.

While this disparity could be attributed to using different dog breeds in SPT solution development as studies have revealed that dog allergen levels depend not only on the breed but also on grooming habits and coat-to-hair ratios. In fact, up to several-fold variation has been reported among individuals of the same breed (Vredegoor et al. 2012). However, Can f 6 is produced by the submaxillary salivary gland and is present in dander; therefore, it would be expected in all SPT solutions derived from *C. familiaris* epithelial material (Yamamoto et al. 2019; Fernández-Caldas et al. 2017). However, since studies have shown that this is often not the case, the observed disparity

likely reflects variations in in-house manufacturing methods. Given that many commercially available dog allergen extracts are provided in a lyophilized format, the susceptibility of Can f 6 to degradation during the lyophilization process may account for its variability in these extracts. This sensitivity highlights the importance of stability when developing products for clinical use.

3.3.4 Molecular cloning and expression of rCan f 6 in *P. pastoris*

Considering the inherent limitations of *E. coli*, particularly concerning PTMs and misfolding, alternative expression systems were explored (Schmidt and Hoffman 2002). In this context, the production of rCan f 6 in *P. pastoris* was pioneered. The PichiaPink™ expression system facilitated the production of rCan f 6 variants in yeast, similarly to allergen Pen m 4 (see 3.2.4). The codon-optimized Can f 6 gene was amplified by PCR (see 2.2.6) using pUC57/Can f 6 plasmid as a template with primers pPinkF_C6 and pPinkR_C6 (see 2.1). Post-amplification, the gene, which now included a C-terminal 6xHis tag sequence, was integrated into the StuI and KpnI restriction sites of the pPink-αF-HC vector (see 2.2). This resulted in a plasmid that allowed the expression of Can f 6, aligned with the *S. cerevisiae* α-mating factor pre-sequence and controlled by the methanol-activated alcohol oxidase 1 (AOX1) promoter.

The rMBP-Can f 6 fusion protein was assembled by subcloning the entire open reading frame (ORF) of 6xHis-MBP-TEV-Can f 6 sequence into the StuI and KpnI restriction sites of pPink-αF-HC expression vector. The Can f 6 ORF was created using pFX7-6xHis-MBP-TEV-Can f 6 plasmid, previously constructed in *S. cerevisiae*, as a template with primers pPinkMBPF_C6 and pPinkMBPR_C6 (see 2.1), which incorporated the restriction sites of StuI and KpnI enzymes, respectively.

The pPink-αF-Can f 6 (Fig. 3.22a) and pPink-αF-MBP-TEV-Can f 6 (Fig. 3.22b) constructs were verified by sequencing at VU LSC IBT sequencing center. Resulting sequences were checked with the “SnapGene Viewer” software.

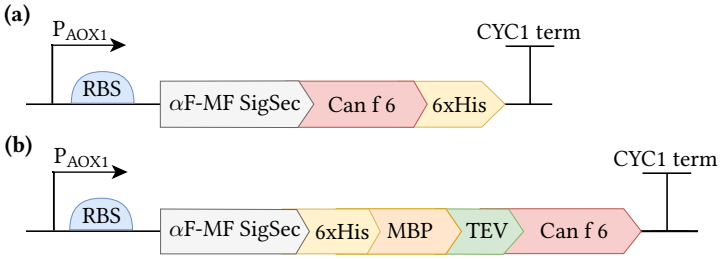


Figure 3.22: Schematic representation of recombinant Can f 6 gene expression cassettes for production in *P. pastoris*. In frame with the *S. cerevisiae* alpha-mating factor signal pre-sequence (αF-MF-SigSec) in **(a)** pPink-αF-Can f 6 expression vector, is the Can f 6 gene C-terminally tagged with a 6xHis tag; in **(b)** pPink-αF-MBP-TEV-Can f 6 expression vector, is Can f 6 fused to the C-terminal end of a 6xHis-MBP fusion protein coding sequence with a TEV protease recognition site between the fusion and passenger protein sequences.

Preliminary experiments were conducted to discern which PichiaPink™ expression system strain (see 2.3) would be most adept at producing the desired protein quantity. *P. pastoris*-produced fusion protein rMBP-Can f 6 theoretical MW was about 65 kDa. Following induction with methanol, all strains containing the pPink-αF-MBP-TEV-Can f 6 vector showed consistent protein bands, as characterized by SDS-PAGE (see 2.2.14) and WB (see 2.2.26) analyses with an MBP-targeting mAb (Fig. 3.23).

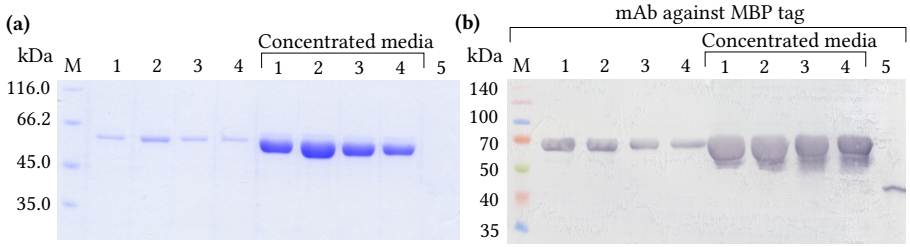


Figure 3.23: Analysis of media collected after heterologous protein induction in PichiaPink™ yeast strains. Yeast transformed with pPink-αF-MBP-TEV-Can f 6 expression vector by **(a)** SDS-PAGE and **(b)** WB (annotated by the used mAb designation). Lanes: [1], Strain 1; [2], Strain 2; [3], Strain 3; [4], Strain 4; [5], *P. pastoris*-produced rMBP (*unpublished*). M – Pierce™ Unstained Protein MW Marker #26610 (see note in 2.2.14) and Spectra™ Multicolor Broad Range Protein Ladder.

Interestingly, upon methanol induction, every strain transformed with the pPink-αF-Can f 6 vector exhibited a dual-band pattern. The first band

aligned with the anticipated theoretical 21 kDa MW of rCan f 6, while the other seemed to represent a glycosylated version of the rCan f 6 protein (Fig. 3.24). Based on the random forest algorithm and pairwise pattern prediction tool (see 2.4), allergen Can f 6 was predicted to have one O-glycosylation site, three N-glycosylation sites, and, possibly, ten additional glycosylation sites. This result was therefore expected. Strain 2 was chosen for both rCan f 6 variants for scale-up expression, based on protein quantity and stability.

3.3.5 Purification and characterization of *P. pastoris*-produced rCan f 6

The dialyzed extracellular *P. pastoris* medium containing the secreted rCan f 6 protein was processed via affinity FPLC using a Ni-NTA Superflow column (Fig. 3.24a, see 2.2.23). The isolated protein purity and quantity were assessed by SDS-PAGE (see 2.2.14) and quantified using ImageJ 1.50b software (see 2.2.24). The protein was estimated to be completely pure with a yield of 36 mg/L (lanes 7 and 8 Fig. 3.24b). The dual-band pattern was more clearly visible in concentrated elution fractions separated by 12% SDS-PAGE (Fig. 3.24a elution fractions) rather than in pooled, diluted fractions separated by 14% SDS-PAGE, where the two bands migrated closely and appeared as a single dominant band (Fig. 3.24b). The dual-band pattern was analyzed by WB (see 2.2.26) with a 6xHis-targeting mAb and a glycosylation assay (see 2.2.27) with Con A (Fig. 3.24b). Lane 1 (Fig. 3.24c), which contained the purified rCan f 6, displayed two bands in WB and the glycosylation assay. Both bands appeared with strong intensity in WB, whereas in the glycosylation assay, the upper band was much more prominent than the lower. These results indicate that both bands correspond to rCan f 6, with the upper band representing a hyperglycosylated form. Lane 2, containing the extracellular *P. pastoris* medium with the secreted rCan f 6 protein, showed results consistent with lane 1 in the WB with 6xHis-targeting mAbs. However, the glycosylation assay revealed only a prominent upper band and additional higher MW bands, suggesting minimal glycosylation of the lower band, which was mainly observed in concentrated samples. Deglycosylation (see 2.2.28) of rCan f 6 in lane 3 further confirmed these observations.

P. pastoris glycosylation patterns predominantly comprise mannose residues, as evidenced by Con A, which primarily favors mannan. Prior studies have noted the importance of allergen mannosylation, describing it as the dominant glycosylation pattern in most environmental allergens (Al-Ghouleh et al.

2012).

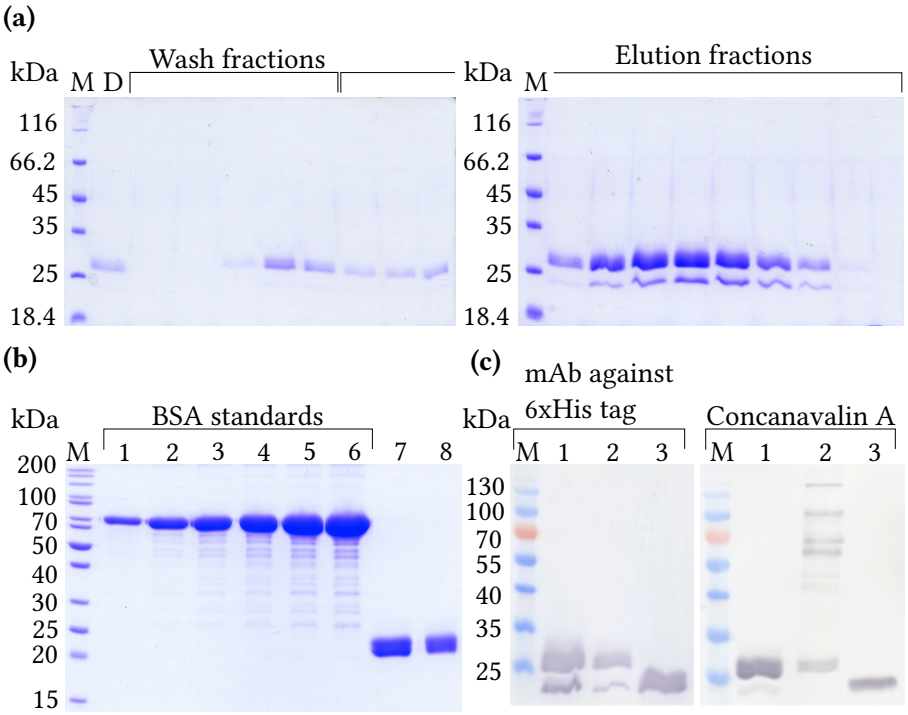


Figure 3.24: Analysis of the purified *P. pastoris* rCan f 6 protein by **(a, b)** SDS-PAGE and **(c)** WB (annotated by the used mAb designation). (a) D – dialysed media fraction, and wash, elution fractions after purification with Ni-NTA chromatography. M – Unstained Protein MW Marker #26610. (b) Lanes: [1-6], correspond to BSA standards with concentrations of 0.25, 0.5, 0.75, 1, 1.5, and 2 mg/ml, respectively; [7] rCan f 6; [8] rCan f 6 diluted by 50%. M – Pierce™ Unstained Protein MW Marker #26614. (c) Lanes: [1], purified rCan f 6; [2], dialyzed extracellular *P. pastoris* medium containing the secreted rCan f 6; [3] purified rCan f 6 deglycosylated with PNGase F. M – Spectra™ Multicolor Broad Range Protein Ladder. Proteins were fractioned in 12% (a, c) and 14% (b) SDS-PAGE gels.

3.3.6 Purification and characterization of *P. pastoris*-produced rMBP-Can f 6

The rMBP-Can f 6 fusion protein was isolated using two distinct methods: cation exchange (CEX) and anion exchange (AEX) chromatography (see 2.2.22). The collected fractions, post-purification, were evaluated using SDS-PAGE (see 2.2.14) and the ImageJ 1.50b software (see 2.2.24). The yields of *P. pastoris*-produced rMBP-Can f 6 purified by CEX and AEX were 18.8 mg/L and

57.5 mg/L, respectively. Further characterization was performed using WB (see 2.2.26) with MBP-targeting mAb for protein identification, and Con A lectin was used for glycosylation (see 2.2.27) profiling (Fig. 3.25). In comparing the results from CEX and AEX, a greater degree of protein degradation was observed in the former. SDS-PAGE gels and WB with MBP-targeting mAb showed multiple bands for the protein sample purified via CEX (Fig. 3.25 lane 3), suggesting protein fragmentation.

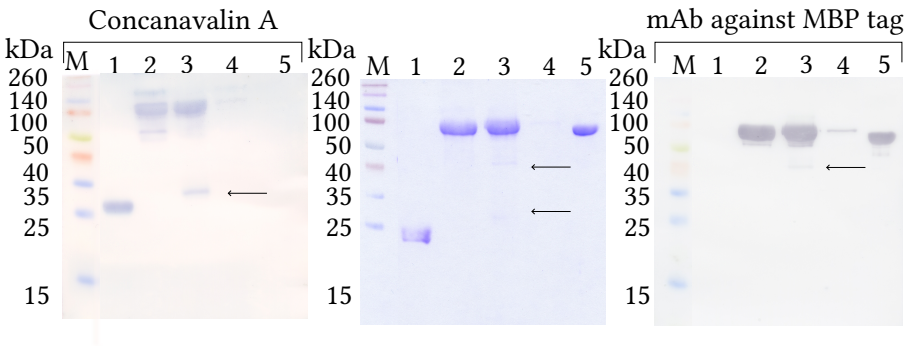


Figure 3.25: Analysis of the purified *P. pastoris* rCan f 6 protein variants by SDS-PAGE and WB (annotated by the used mAb designation). Lanes: [1], rCan f 6; [2], rMBP-Can f 6 purified via AEX; [3], rMBP-Can f 6 purified via CEX; [4], wash fraction from purification of rMBP-Can f 6 by CEX; [5], *E. coli*-produced rMBP-Pen m 4. M – Spectra™ Multicolor Broad Range Protein Ladder. Proteins were fractioned in 14% SDS-PAGE gels. Arrows designate the probable MBP-containing and Can f 6-containing fragments.

Together with the rMBP-Pen m 4 data (see 3.2.6), these findings indicate that *P. pastoris*-produced MBP-fusion proteins can undergo at least two proteolytic processing patterns. The first pattern is cleavage in regions C-terminal to the MBP domain, inferred from MBP-reactive lower MW fragments. In the rMBP-Can f 6 CEX-purified sample, the ~ 45 kDa fragment detected by WB with mAb against the MBP tag most likely represents an MBP-containing proteolysis product (Fig. 3.25 lane 3, upper arrow). The additional ~ 25 kDa band, seen in WB with Con A, is compatible with a Can f 6-containing fragment (Fig. 3.25 lane 3, lower arrow), although this assignment was based on apparent molecular mass rather than direct confirmation. These observations are consistent with previous reports of MBP-fusion protein degradation in *P. pastoris*, where MBP-associated proteolytic activity was proposed to affect linker/spacer and cargo regions located C-terminally to MBP (Z Li et al. 2010; Moua et al. 2016).

The glycosylation profiles of AEX- and CEX-purified rMBP-Can f 6 were comparable, but the CEX-purified sample contained an additional band corresponding approximately to the molecular mass of MBP (Fig. 3.25 lane 3). This supports fragmentation of the fusion protein and suggests that cleavage may have occurred in or near the spacer/cargo region. Interestingly, degradation was more evident in the CEX-purified rMBP-Can f 6 sample, whereas a comparable degradation pattern was previously observed for the AEX-purified rMBP-Pen m 4 sample (see 3.11). Therefore, the observed proteolytic processing is unlikely to be explained solely by the purification method, purification pH, protein yield, or cargo sequence. It also cannot be attributed to proteinase A activity alone, because the proteins were produced in the *P. pastoris pep*4 knockout strain.

The second processing pattern involved the loss of the N-terminal 6xHis-containing region. Neither rMBP-Can f 6 protein reacted with mAbs against the 6xHis tag (data not shown), suggesting removal or processing of the N-terminal region of the 6xHis-MBP-TEV-Can f 6 construct. Similar loss of 6xHis reactivity was observed for rMBP-Pen m 4 proteins (see 3.2.6). This pattern differs from the C-terminal tag loss reported by Moua et al. 2016, where MBP-Cargo-Tag and Cargo-MBP-Tag constructs were cleaved within spacer and cargo regions downstream of MBP. In the present constructs, the 6xHis tag was located N-terminally to MBP, a region not previously reported as susceptible to proteolysis in comparable fusion proteins (Z Li et al. 2010; Moua et al. 2016).

Similar to other higher eukaryotes *P. pastoris* expresses a diverse set of proteases, which are generally categorized into three groups: cytosolic proteosome components, vacuolar proteases, and proteases located en route and in the secretory pathway (Y Zhang, R Liu, et al. 2007). The PichiaPink™ strain 2 is a *pep*4 knockout, rendering it incapable of synthesizing the aspartic vacuolar proteinase A. Because proteinase A activates additional vacuolar proteases, *pep*4 knockout strains exhibit reduced proteinase B activity and carboxypeptidase Y activity (Gast et al. 2022).

Dipeptidyl aminopeptidases (DPP) are a class of proteolytic enzymes belonging to the S9B serine protease family. DPPs cleave N-terminal dipeptides from proteins bearing an X-Proline (P)/Alanine (A) motif. In *P. pastoris* two DPP enzymes have been identified: Ste13p and Dap2p (Hopkins et al. 2014; Prabha et al. 2009). In this study, the N-terminal end of the fusion proteins contains the *S. cerevisiae* α-mating factor presequence and pro-region, terminating with lysine (K) and arginine (R) amino acids, the recognition sequence

for the Kex2p protease. The secretory pathway endoprotease Kex2p cleaves the α-mating factor to release the protein of interest. This process is typically followed by Ste13p-mediated removal of the EAEA sequence (glutamic acid and alanine); however, unintended cleavage of additional residues by Ste13p can hinder proper protein maturation (Prabha et al. 2009). As a result, most modern expression systems omit the EAEA sequence in advance (Obst et al. 2017). Accordingly, Ste13p should not have affected the 6xHis-tag present in the current constructs. Dap2p likewise prefers a proline or alanine in the penultimate position, further supporting the expectation that the tag would remain intact (Hopkins et al. 2014).

Thus, the canonical α-mating factor processing pathway does not readily explain the loss of the 6xHis tag in these constructs. Because no published studies using identical or sufficiently similar constructs were identified, the precise mechanism remains unresolved. However, the absence of 6xHis reactivity in multiple *P. pastoris*-produced MBP-fusion proteins, together with the retained MBP signal and analogous results from rMBP-Pen m 4, supports proteolytic processing of the N-terminal 6xHis-containing region rather than simple antibody inaccessibility. It is therefore plausible that an endogenous *P. pastoris* protease cleaved the N-terminal region of these MBP-fusion proteins.

3.3.7 Immunoreactivity of rCan f 6 and rMBP-Can f 6

Three groups of patient plasma were used to test recombinant Can f 6 proteins (Table 3.7): (i) nineteen individual human plasma samples tested positive for Can f 6 sIgE; (ii) Pool-positive, pool composed of seven human plasma samples that tested positive for *C. familiaris* allergen extract but negative for Can f 6 sIgE; (iii) Pool-negative, pool composed of seven human plasma samples that tested negative for IgE.

Table 3.7: Characteristics of human plasma and serum samples used in the assays.

	No. of samples	*C. familiaris* sIgE	Can f 6 sIgE	Other allergens	Serum pool
i.	19	Positive[1,2]	Positive[1,2]	Positive[1,2]	Individual samples
ii.	7	Positive[1,2]	Negative[1,2]	Positive[1,2]	Positive-pool
iii.	7	Negative[1]	Negative[1]	Negative[1]	Negative-pool

[1] IgE ALEX[2]® Allergy test — sIgE ≥ 0.30 kU/L = positive; < 0.30 kU/L = negative.

[2] ImmunoCAP™ or Phadia UniCap™ assays — sIgE ≥ 0.35 kU/L = positive; < 0.35 kU/L = negative.

First, the immunoreactivity of rMBP-Can f 6, produced in *P. pastoris*, was evaluated. Indirect ELISA (see 2.2.30, 2.2.32) results (Fig. 3.26) showed that rMBP-Can f 6 purified by CEX elicited a significantly stronger sIgE response than the one purified via AEX. A paired student t-test with 19 observations (each observation measured in triplicate) confirmed the statistical significance with $p = 1.92 \times 10^{-16}$. This result may indicate that lipocalin allergen Can f 6 undergoes pH-dependent structural changes under low pH.

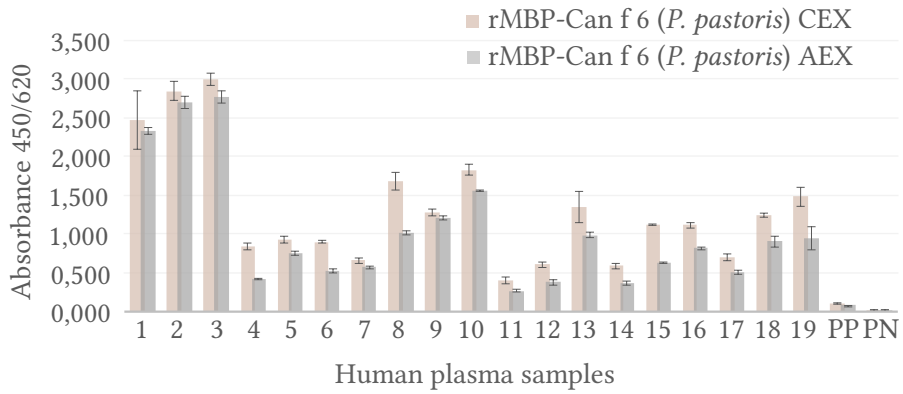


Figure 3.26: Immunoreactivity analysis of *P. pastoris*-produced rMBP-Can f 6 proteins in an indirect ELISA platform. CEX – represents rMBP-Can f 6 purified through cation exchange chromatography, and AEX – represents rMBP-Can f 6 purified through anion exchange chromatography. [1-19] correspond to individual plasma from *C. familiaris*-sensitized patients positive for Can f 6; PP – positive-pool (*C. familiaris* sIgE-positive, but Can f 6 sIgE-negative, n = 7); PN – negative-pool (n = 7).

pH-dependent conformational rearrangements have been reported in stud-

ies involving other lipocalins, such as nitrophorin (Roberts et al. 2001), tear lipocalin (Gasymov et al. 2010), plasma retinol binding protein (Newcomer and Ong 2000; Calderone et al. 2003), membrane enzyme PagP (Hwang et al. 2002), and bovine β-lactoglobulin (Fogolari et al. 2005). Despite weak overall amino acid identity – often below the 20% threshold required for reliable alignment – lipocalins share several short conserved sequence motifs. They also exhibit a common tertiary structure composed of eight anti-parallel β-strands forming a β-barrel structure with one α-helix (Yamamoto et al. 2019; Flower et al. 2000). This hydrogen-bonded β-barrel forms a hydrophobic cavity that binds small hydrophobic molecules such as retinol, steroids, odorants, and pheromones (Hilger et al. 2012). In the case of bovine β-lactoglobulin, ligand binding is controlled through the reversible loop movements at the end of the β-barrel structure. These loops act as a lid, closing the cavity and regulating access for small hydrophobic molecules. This mechanism is speculated to apply to all lipocalin proteins as a general feature of their ligand-binding behavior (Gasymov et al. 2010; Fogolari et al. 2005).

Such conformational changes would also result in the exposure of different epitopes, potentially explaining the observed differences in sIgE binding between rMBP-Can f 6, purified under different pH conditions. Alternatively, the increased immunoreactivity may stem from partial degradation of the cation-purified rMBP-Can f 6, exposing linear epitopes (YJ Wang et al. 2017). Another possibility is that the positive net charge required for CEX alters the protein's conformation or stability differently than the negative charge conditions used in AEX. Given that ion exchange chromatography is the gold standard for characterizing therapeutic proteins' charge variants, the separation process could inadvertently favor protein isoforms with exposed or intact IgE-binding epitopes (Fekete et al. 2015).

Absorbance measurements (Fig. 3.27) confirmed that all recombinant Can f 6 variants specifically bind to Can f 6 sIgE in human plasma samples without evidence of non-specific interactions. One-way repeated measures ANOVA was performed to determine whether recombinant Can f 6 proteins exhibit significant differences in IgE reactivity across patient sera. Positive and negative control proteins were included in the indirect ELISA to ensure specificity and validate the experimental setup. However, as in Pen m 4 analysis, control proteins were excluded from the statistical comparison of recombinant proteins to prevent artificial inflation of variability. The ANOVA revealed significant differences among proteins $F(4,\ 72) = 29.8,\ p < 0.001$,

indicating variability in reactivity. No significant variability was observed between replicates, confirming the consistency of the triplicate measurements. Glycosylation-related variability was assessed using a CCD control to evaluate whether anti-CCD IgE affected the observed sIgE binding to Can f 6 variants. As shown in Fig. 3.27, no significant CCD-related differences were found.

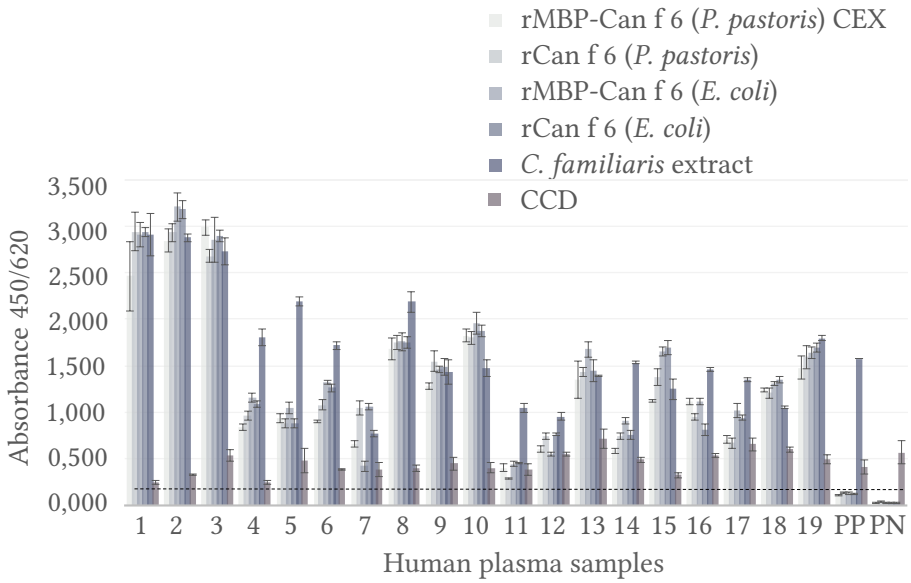


Figure 3.27: IgE reactivity of patient plasma with recombinant Can f 6 proteins and the *C. familiaris* allergen extract in an indirect ELISA platform. CEX – represents rMBP-Can f 6 purified through cation exchange chromatography, and AEX – represents rMBP-Can f 6 purified through anion exchange chromatography. CCD is the glycosylation IgE marker. [1-19] correspond to individual plasma from *C. familiaris*-sensitized patients positive for Can f 6; PP – positive-pool (*C. familiaris* sIgE-positive, but Can f 6 sIgE-negative, n = 7); PN – negative-pool (n = 7). A dashed line represents the cut off.

Post hoc comparisons were conducted using Tukey's HSD test to identify specific protein pairs with significant differences in reactivity. This analysis was done on non-averaged triplicate data, which allowed only substantial differences to reach statistical significance. This approach represented a compromise between Bonferroni-adjusted paired analyses, which detected significant differences across all pairs, and Tukey's HSD applied to averaged, unpaired data, which yielded no significance due to conservative post-ANOVA assumptions. Although applying Tukey's HSD to raw data inflated statistical power, it effectively minimized noise and detected only pronounced differences. The

analysis showed that rMBP-Can f 6 purified through anion exchange chromatography exhibited significantly different reactivity than *E. coli*-produced proteins (Fig. 3.28, Table 3.8). No statistically significant differences were observed between the remaining pairs, suggesting comparable reactivity levels across those groups.

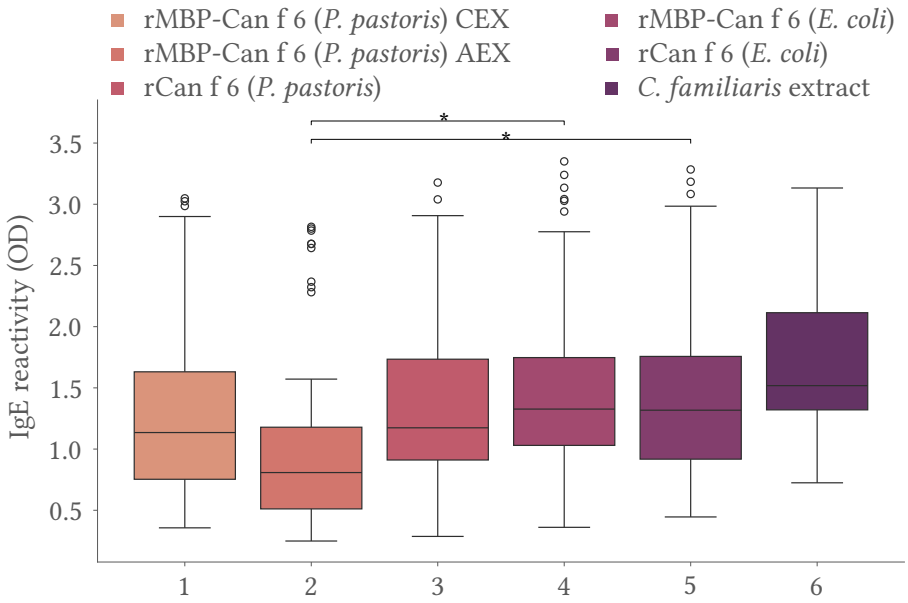


Figure 3.28: IgE reactivity of patient plasma with recombinant Can f 6 proteins visualized in a boxplot. Lanes: *P. pastoris*-produced rMBP-Can f 6 purified through [1] CEX and [2] AEX; [3], *P. pastoris*-produced rCan f 6; [4], *E. coli*-produced rMBP-Can f 6; [5] *E. coli*-produced rCan f 6; [6] *C. familiaris* allergen extract. The data represents normalized IgE reactivity across 19 individual plasma from *C. familiaris*-sensitized patients positive for Can f 6. Reactivity is represented as normalized OD values, and significant differences between proteins are annotated with asterisks ($** = p < 0.01$, $*** = p < 0.001$).

The aim of this research was to examine the production of the glycoprotein Can f 6 in a eukaryotic expression system, focusing on its ability to specifically recognize sIgE antibodies, its non-reactivity toward non-specific IgE, and the feasibility of high-yield production for large-scale applications. The co-expression of rCan f 6 protein with MBP in both systems led to a heightened yield, affirming it as the most effective production method for rCan f 6 across both expression platforms. Specifically, MBP enhanced production by 1.8-fold in *E. coli* (42 mg/L vs. 78 mg/L) and 1.6-fold in *P. pastoris* (36 mg/L vs. 57.5 mg/L in AEX). However, while MBP substantially increased the molecular mass

Table 3.8: Pairwise comparisons after repeated measures ANOVA and Tukey's HSD post hoc analysis.

A	System A	B	System B	meandiff	p-adj	Reject H_0
rCan f 6	*E. coli*	rCan f 6	*P. pastoris*	−0.0890	0.9711	False
rCan f 6	*E. coli*	rMBP-Can f 6	*E. coli*	0.0049	1.0000	False
rCan f 6	*E. coli*	rMBP-Can f 6, AEX	*P. pastoris*	−0.4563	0.0130	True
rCan f 6	*E. coli*	rMBP-Can f 6, CEX	*P. pastoris*	−0.1736	0.7403	False
rCan f 6	*P. pastoris*	rMBP-Can f 6	*E. coli*	0.0938	0.9649	False
rCan f 6	*P. pastoris*	rMBP-Can f 6, AEX	*P. pastoris*	−0.3674	0.0770	False
rCan f 6	*P. pastoris*	rMBP-Can f 6, CEX	*P. pastoris*	−0.0847	0.9758	False
rMBP-Can f 6	*E. coli*	rMBP-Can f 6, AEX	*P. pastoris*	−0.4612	0.0117	True
rMBP-Can f 6	*E. coli*	rMBP-Can f 6, CEX	*P. pastoris*	−0.1785	0.7198	False
rMBP-Can f 6, AEX	*P. pastoris*	rMBP-Can f 6, CEX	*P. pastoris*	0.2827	0.2762	False

meandiff – The difference in means: mean (A) - mean (B); p-adj – adjusted p-value using the Tukey HSD correction for multiple comparison, it controls the family-wise error rate to reduce false positives; Reject H_0 – reject the null hypothesis, this indicates whether the null hypothesis (no difference between the two groups) is rejected after correction. True if p-adj $<$ 0.05 – significant difference between the pair, False if p-adj $\geqslant$ 0.05 – no significant difference.

of the recombinant product, the mass yield did not correspond to a higher calculated molar yield of Can f 6-containing molecules in either expression system (Table 3.9).

Table 3.9: Comparison of mass and molar yields of non-fused and MBP-fused recombinant proteins.

System	**Protein**	**MW (kDa)**	**Yield (mg/L)**	**Mass ratio****	**Molar yield* (µmol/L)**	**Molar ratio****
E. coli	rCan f 6	23.7	42	1.86	1.77	0.68
E. coli	rMBP-Can f 6	64.4	78		1.21	
P. pastoris	rCan f 6	21	36	1.59	1.71	0.51
P. pastoris	rMBP-Can f 6	65	57.5		0.88	

* Molar yield was calculated as yield (mg/L) divided by calculated MW (kDa).
** Ratios indicate the effect of MBP fusion within each expression system.

Importantly, the immunological assays were performed using equal protein mass for coating; therefore, MBP-fusion proteins contained fewer Can f 6-containing molecules than their non-fused counterparts. Nevertheless, no statistically significant differences in IgE reactivity were observed among the recombinant Can f 6 variants. Thus, MBP fusion preserved Can f 6 antigenic performance under mass-normalized assay conditions.

Comparative analysis of recombinant Can f 6 variants demonstrated that all prokaryotic and eukaryotic proteins specifically bound Can f 6 sIgE and showed no non-specific binding, confirming that the recombinant proteins retain essential IgE-binding epitopes and supporting their use in immuno-

logical research and potential diagnostic applications. Glycosylation did not significantly influence IgE reactivity, suggesting that a eukaryotic expression system is not required to produce immunoreactive Can f 6 proteins. However, as basophil degranulation assays were not conducted, further studies are needed to assess the functional allergenic activity of these recombinant proteins.

Taken together, *E. coli* and *P. pastoris* were both suitable for producing immunoreactive rCan f 6. MBP fusion increased recovered protein mass and preserved IgE reactivity despite reducing the calculated molar yield of Can f 6-containing molecules. Considering mass yield, immunoreactivity, and production simplicity, *E. coli*-produced rMBP-Can f 6 is a strong candidate for larger-scale production, among the *P. pastoris*-produced variants, AEX purification provided substantially higher recovery, whereas CEX-purified rMBP-Can f 6 showed stronger IgE reactivity but also greater fragmentation. Therefore, the optimal *P. pastoris* purification strategy depends on the intended application.

3.4. Streamlining PichiaPink™ recombinant allergen production

3.4.1 Construction of optimized vectors for allergen expression and purification

Following our observation of N-terminal proteolytic cleavage in 6xHis-MBP-allergen constructs, we developed two standardized pPink-HC derivatives (see 2.2) to both streamline the cloning workflow and systematically investigate this processing phenomenon across additional allergen families.

The pPink-αF-6xHis vector incorporates a C-terminal 6xHis tag directly into the pPink-αF-HC backbone through the insertion of a synthetic oligonucleotide cassette (generated from pPink_oligo_FR and pPink_oligo_RV oligos, see 2.1, 2.2.2) between the SmaI and KpnI restriction sites (see 2.2.1, 2.2.5). This design enables direct cloning of allergen sequences while ensuring consistent C-terminal tagging across all constructs.

The pPink-αF-6xHis-MBP-TEV vector features an N-terminal 6xHis-MBP-TEV cassette, amplified from pFX7-MBP-TEV-6xHis vector using MBP-Pink-Fr and MBP-Pink-Rv primers (see 2.1), which introduced the RE sites of StuI and KpnI into the fragment. The PCR-generated cassette was inserted at the StuI and KpnI junction (see 2.2.6, 2.2.1, 2.2.5). This construct was designed to

determine whether the observed N-terminal processing represents a general system limitation or an allergen-specific phenomenon. These ready-to-use vectors reduce cloning steps while maintaining the flexibility to test whether fusion to MBP increases recombinant allergen yield without affecting its antigenicity.

3.4.2 Validation of the expression platform using yellow jacket venom allergen Ves v 5

To validate our streamlined PichiaPink™ vectors and investigate whether N-terminal proteolytic processing affects allergens beyond our initial test set, Ves v 5 (antigen 5) from *Vespula vulgaris* yellow jacket venom was selected. This allergen represents an ideal test tool for several reasons: it is the dominant venom component and primary diagnostic target for Hymenoptera allergy; its complex disulfide-bonded structure necessitates eukaryotic expression for proper folding; and previous *E. coli* expression attempts yielded mainly insoluble aggregates requiring *in vitro* refolding, whereas expression in eukaryotic systems obtained properly folded variants resembling the native protein (Borodina et al. 2010; Kischnick et al. 2006; Schiener et al. 2017). Ves v 5 production was explored in *P. pastoris* successfully (Monsalve et al. 1999); however, its production combined with MBP has yet to be investigated. The structural requirements of Ves v 5, combined with its clinical importance in molecular diagnostics and allergen immunotherapy, make it particularly suitable for assessing both the expression efficiency of our vectors and the extent of N-terminal processing in the PichiaPink™ system.

To benchmark against the eukaryotic system, rVes v 5 variants were also produced in *E. coli*, despite literature indicating limited suitability. Cloning followed the rPen m 4 workflow (see 3.2.2), while using the pMA-RQ/Ves v 5 template and pET28F_V5 and pET28R_V5 primers to amplify the codon-optimized 612 bp Ves v 5 coding sequence (see 2.1.3, 2.1). Inserts were cloned into the pET28a(+) and pET28a(+)-MBP-TEV vectors. Expression in BL21(DE3) yielded soluble protein only as an MBP fusion; non-fused rVes v 5 accumulated in inclusion bodies. Switching to the Rosetta strain to improve protein expression had no positive effect on solubility or yield. Purification was performed with Ni-NTA (see 2.2.18) or ZYMO Reasearch columns (see 2.2.19) under denaturing conditions for rVes v 5 and under native conditions for rMBP-Ves v 5. Following purification, dialysis was performed to refold the denatured rVes v 5, and the following fractions were concentrated with protein concentrators (see 2.2.20).

Recombinant protein presence was confirmed by WB using mAb against the 6xHis and MBP tags (Fig. 3.31 lanes 3 and 4). In this study, *E. coli*-produced rVes v 5 yielded 14 mg/L and fusion with MBP increased that yield 3.9-fold to 55 mg/L. However, as observed from Fig. 3.31, rVes v 5 (lane 4), despite optimizations, purification yields two lines, which both interact with mAb against the 6xHis tag. MS analysis showed that the main mass is 121.3 Da smaller than the calculated theoretical mass of rVes v 5 ($\sim$26.9 kDa), and the other protein mass, which is seen in the SDS-PAGE gel, was shown in MS to be approximately 20.3 kDa. It is worth also noting, that while rMBP-Ves v 5 soluble expression was achieved, the instability of the synthesized protein could not be improved by optimization of growth or purification conditions.

Continuing with the PichiaPink™ system assessment, the 204-aa long Ves v 5 protein coding sequence was excised from the pMA-RQ/Ves v 5 plasmid with StuI and KpnI RE and cloned into the SmaI and KpnI sites of pPink-αF-6xHis and pPink-αF-6xHis-MBP-TEV vectors following standard protocols (see 2.2.4, 2.2.1, 2.2.5). This cloning strategy yielded two constructs with only MBP distinguishing the two tagging configurations: the pPink-αF-6xHis-Ves v 5 construct carries an N-terminal 6xHis tag, while the pPink-αF-6xHis-MBP-TEV-Ves v 5 construct features an N-terminal 6xHis-MBP-TEV fusion sequence. Constructs were verified by sequencing at the VU LSC IBT sequencing center, and analyzed using "SnapGene Viewer" software (see 2.4).

Following transformation into four PichiaPink™ *ade2* strains, pilot experiments identified Strain 2 as the optimal producer for both constructs after 48 h methanol induction. SDS-PAGE analysis revealed successful expression of both the 24.3 kDa rVes v 5 and 67.8 kDa rMBP-Ves v 5 proteins, with double banding patterns indicating post-translational glycosylation characteristic of *P. pastoris* expression (Fig. 3.29).

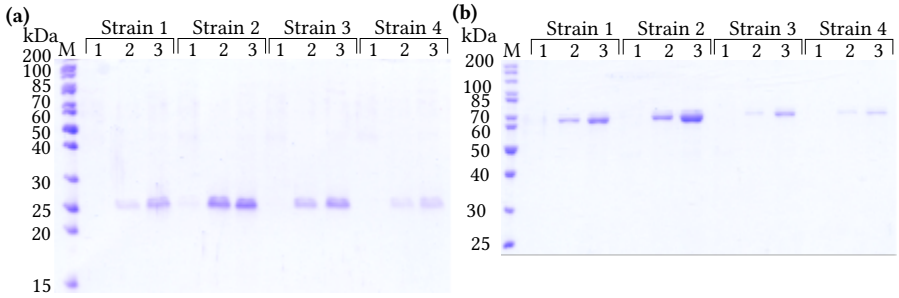


Figure 3.29: Analysis of media collected after heterologous protein induction in PichiaPink™ yeast strains transformed with **(a)** pPink-αF-Ves v 5 and **(b)** pPink-αF-MBP-TEV-Ves v 5 expression vectors in 14% SDS-PAGE gels. Lanes: [1], media before induction with MeOH; [2], media after 24 h induction with 2% MeOH; [3] media after 48 h induction with 2% MeOH. The names of the strains are written above the corresponding lane numbers. M – PageRuler™ Unstained Protein Ladder #26614.

Large-scale expression culture media were collected, dialyzed, and purified using complementary strategies based on tag accessibility: the 6xHis-tagged rVes v 5 was purified via affinity FPLC using a Ni-NTA Superflow column (see 2.2.23), while the 6xHis-MBP fusion protein required ion exchange chromatography, CEX (see 2.2.22), because purification through the N-terminal 6xHis tag was not effective. Purified protein yields, determined by Bradford assay and densitometric analysis using ImageJ 1.50b software (see 2.2.24), demonstrated a noteworthy difference: 36 mg/L for rVes v 5 vs. 134.5 mg/L for rMBP-Ves v 5, representing a 3.7-fold enhancement with MBP fusion (Fig. 3.30). To our knowledge, this 134.5 mg/L yield represents the highest reported production level for recombinant Ves v 5 across all expression systems, substantially exceeding previous yields from *E. coli* (requiring refolding), baculovirus-insect cells, and other yeast platforms (Suck et al. 2000; Monsalve et al. 1999; Henriksen et al. 2001).

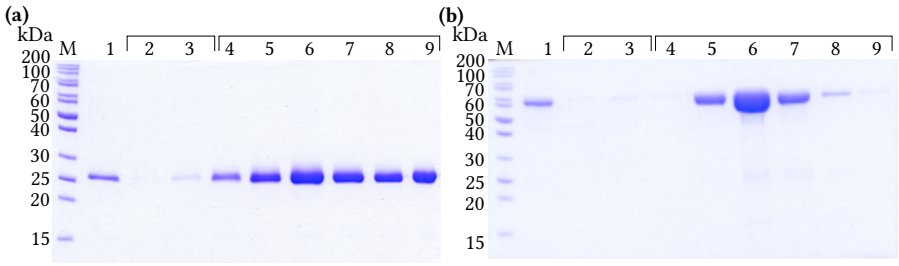


Figure 3.30: Analysis of *P. pastoris* purification of rVes v 5 **(a)** and rMBP-Ves v 5 **(b)** in 14% SDS-PAGE gels. Lanes: [1], media sample after dialysis; [2-3], wash fractions; [4-9], elution fractions. M – PageRuler™ Unstained Protein Ladder #26614.

To investigate the apparent loss of histidine tag detection, WB was performed using antibodies against 6xHis and MBP tags (Fig. 3.31). WB showed that neither rMBP-Ves v 5 (Fig. 3.31 lane 1) nor rMBP-Pen m 4 (Fig. 3.31 lane 5) was detectable with anti-6xHis antibodies. In contrast, downstream fusion partners remained detectable with the corresponding antibodies. These results support that N-terminal processing is a recurrent feature of 6xHis-MBP-allergen fusion constructs expressed in the PichiaPink™ system, independent of the allergen family.

Glycosylation analysis with Con A (not shown) confirmed that both *P. pastoris*-produced Ves v 5 variants are glycosylated. MS analysis of the deglycosylated proteins (not shown) revealed MW deviating from the theoretical values, similar to the rMBP-Pen m 4 MS analysis (see 3.2.6). The observed mass of rVes v 5 was 7 Da higher than expected, whereas the observed mass of rMBP-Ves v 5 was approximately 2 kDa lower than expected. Part of this discrepancy may be related to incomplete deglycosylation and disulfide bond formation. However, the reduced mass of rMBP-Ves v 5, together with the loss of anti-6xHis reactivity, supports proteolytic removal of the N-terminal 6xHis-containing region.

Both 6xHis-containing Ves v 5 constructs begin with a Pro-Met-Gly-Ser-Ser sequence linked to the 6xHis tag, followed by a Ser-Ser-Gly linker leading to the thrombin recognition site. Based on monoisotopic calculations, this region accounts for approximately 1281 Da with the 6xHis tag alone, 1512 Da when extended to the thrombin site, and 2121 Da when the complete thrombin-region segment is included. Therefore, the observed mass shift is consistent with cleavage occurring in or near the 6xHis-linker-thrombin region.

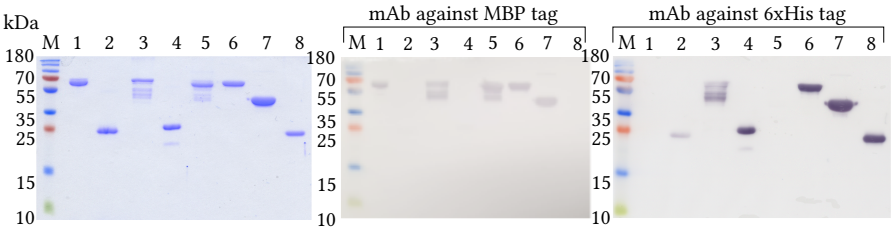


Figure 3.31: Analysis of rVes v 5 variants by SDS-PAGE and WB (annotated by the used mAb designation). Lanes: [1], *P. pastoris*-produced rMBP-Ves v 5; [2], *P. pastoris*-produced rVes v 5; [3], *E. coli*-produced rMBP-Ves v 5; [4], *E. coli*-produced rVes v 5; [5], *P. pastoris*-produced rMBP-Pen m 4; [6], *E. coli*-produced rMBP-Pen m 4; [7], *E. coli*-produced MBP with a 6xHis tag; [8] *E. coli*-produced rPen m 4. M – PageRuler™ Plus Prestained Protein Ladder #26619. Proteins were separated in 14% SDS-PAGE gels.

These findings can be contextualized within the broader literature on N-terminal tag processing in *P. pastoris* systems. The N-terminally tagged αF-6xHis-*protein* constructs in other *P. pastoris* strains, including PichiaPink™, have generally been reported as successful (Smith et al. 2023; Chronopoulou et al. 2023). However, N-terminal proteolytic cleavage has also been reported for N-terminally-tagged fusion constructs, such as αF-6xHis-GFP-*protein*, where WB showed loss of the His-tag signal but retention of GFP with intact proteins detected in some cases but not others (Zheng et al. 2022). Although Zheng et al. 2022 attributed this discrepancy to limited antibody accessibility in folded proteins, the denaturing WB conditions (SDS, reducing agents, heating to 100 °C) should eliminate such structural constraints.

This variability suggests that the design of the fusion tag and the structural context of the expressed protein influence susceptibility to proteolysis, rather than a uniform defect in the expression system. To our knowledge, such N-terminal proteolysis within MBP-fusion constructs expressed in *P. pastoris* has not been previously described. This finding highlights a potential vulnerability that should be considered in future fusion protein design. Future work should investigate whether the thrombin recognition motif or adjacent linker sequence contributes to the observed N-terminal processing and whether sequence modifications can preserve the integrity of the 6xHis tag.

To assess functional integrity, immunoreactivity of Ves v 5 variants was evaluated with three groups of patient plasma (Table 3.10): (i) fourteen individual human plasma samples tested positive for Ves v 5 sIgE; (ii) Pool-positive, pool composed of five human plasma samples that tested positive for Pen m 4

sIgE but negative for Ves v 5 sIgE; (iii) Pool-negative, pool composed of five human plasma samples that tested negative for IgE.

Table 3.10: Characteristics of human plasma and serum samples used in the assays.

	No. of samples	Pen m 4 sIgE	Ves v 5 sIgE	Other allergens	Serum pool
i.	14	Negative[1,2]	Positive[1,2]	Positive[1,2]	Individual samples
ii.	5	Positive[1,2]	Negative[1,2]	Positive[1,2]	Positive control pool
iii.	5	Negative[1]	Negative[1]	Negative[1]	Negative-pool

[1] IgE ALEX[2]® Allergy test — sIgE ≥ 0.30 kU/L = positive; < 0.30 kU/L = negative.

[2] ImmunoCAP™ or Phadia UniCap™ assays — sIgE ≥ 0.35 kU/L = positive; < 0.35 kU/L = negative.

Indirect ELISA analysis revealed clear differences in binding specificity between tested expression systems (Fig. 3.32). *P. pastoris*-produced Ves v 5 variants showed specific binding to sIgE with no detectable cross-reactivity to control pools, confirming preservation of native-like structure. In contrast, *E. coli*-produced variants – both the purified denatured form and the refolded protein – exhibited non-specific binding to control pools. Statistical analysis confirmed these observations. One-way repeated measures ANOVA revealed significant differences among protein variants $F(4,\ 52) = 11.17,\ p < 0.001$, while technical replicates showed no differences $F(2,\ 375) = 0.011,\ p < 0.988$, indicating that all measurements are consistent.

Individual serum analysis highlighted the diagnostic risks of using misfolded allergens. For instance, serum 10 (0.37 kU/L by ALEX) reacted with *E. coli*-produced variants but fell below the ELISA cut off for properly folded *P. pastoris*-produced variants, likely reflecting non-specific binding already observed with control pools rather than genuine allergen recognition. Similarly, serum 14 (1.37 kU/L by ALEX) reacted exclusively with denatured *E. coli* rVes v 5, suggesting non-specific interaction rather than IgE-mediated binding. Importantly, this differential reactivity cannot be attributed simply to low sIgE levels, as sera with comparable ALEX values – such as sera 13 and 7 (1.36 and 0.58 kU/L, respectively) – displayed markedly different binding profiles, confirming that the observed patterns reflect expression system-dependent effects rather than antibody concentration levels.

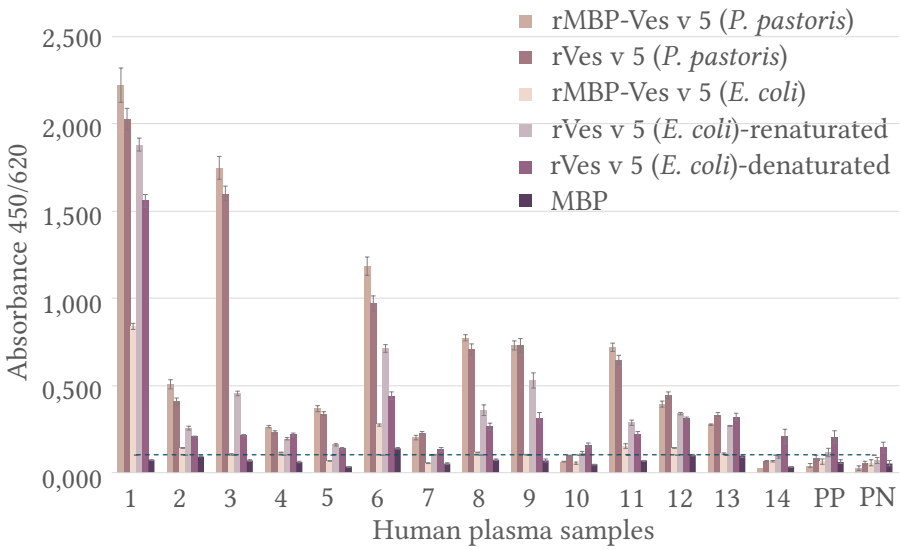


Figure 3.32: IgE reactivity of patient plasma with recombinant Ves v 5 proteins in an indirect ELISA platform. [1-14] correspond to individual plasma from *V. vulgaris*-sensitized patients positive for Ves v 5; PP – positive-pool (*V. vulgaris* and Ves v 5 sIgE-positive, n = 5); PN – negative-pool (n = 5). MBP acts as a negative control. A dashed line represents the cut off.

The lack of reactivity with *P. pastoris* variants in some confirmed sIgE-positive sera may also reflect serum degradation from shipping conditions or freeze-thaw cycles, emphasizing the importance of proper sample handling in diagnostic validation studies. These findings underscore how misfolded recombinant allergens can generate false-positive results, while properly folded variants provide more reliable diagnostic specificity.

Post hoc pairwise comparisons were conducted using the Bonferroni correction (Fig. 3.33). While *E. coli*-produced rVes v 5 variants showed significantly higher reactivity than *E. coli*-produced rMBP-Ves v 5, this apparent binding must be interpreted cautiously given the demonstrated non-specific interactions with control sera. In contrast, *P. pastoris*-produced rMBP-Ves v 5 and rVes v 5 proteins demonstrated the highest binding sensitivity. rMBP-Ves v 5 antigenicity was also confirmed by an inhibition assay conducted by Laimis Silimavičius. This assay showed that just 0.05 mg of *P. pastoris*-produced rMBP-Ves v 5 achieved 92% inhibition of Ves v 5 sIgE binding.

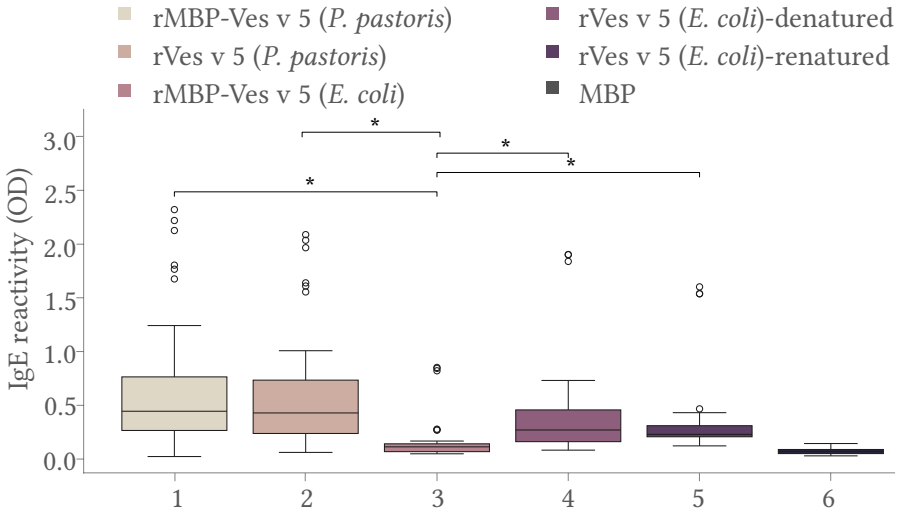


Figure 3.33: IgE reactivity of patient plasma with recombinant Ves v 5 proteins visualized in a boxplot. Lane: [1], *P. pastoris*-produced rMBP-Ves v 5; [2], *P. pastoris*-produced rVes v 5 protein; [3], *E. coli*-produced rMBP-Ves v 5; [4], *E. coli*-produced rVes v 5, denatured form; [4], *E. coli*-produced rVes v 5, renatured form;[6], MBP. The data represents normalized IgE reactivity across 14 individual plasma from *V. vulgaris*-sensitized patients positive for Ves v 5. Reactivity is represented as normalized OD values, and significant differences between proteins are annotated with asterisks ($* = p < 0.05$).

Analyzing further, in contrast to Pen m 4 and Can f 6, MBP fusion increased both recovered protein mass and the calculated molar yield of Ves v 5-containing molecules (Table 3.11). Thus, for Ves v 5, the improved yield was not only a mass-based effect of the larger fusion protein but also reflected increased recovery of allergen-containing molecules.

Table 3.11: Comparison of mass and molar yields of non-fused and MBP-fused recombinant proteins.

System	Protein	MW (kDa)	Yield (mg/L)	Mass ratio**	Molar yield* (µmol/L)	Molar ratio**
E. coli	rVes v 5	26.9	14	3.9	0.52	1.56
E. coli	rMBP-Ves v 5	67.7	55		0.81	
P. pastoris	rVes v 5	24.3	36	3.7	1.48	1.33
P. pastoris	rMBP-Ves v 5	67.8	134.5		1.98	

* Molar yield was calculated as yield (mg/L) divided by calculated MW (kDa).

** Ratios indicate the effect of MBP fusion within each expression system.

Taken together with the stronger IgE reactivity of *P. pastoris*-produced variants, these results support *P. pastoris*-produced rMBP-Ves v 5 as the most suitable construct for high-yield recombinant Ves v 5 production for diagnostic applications.

Interestingly, workflow testing with different *V. vulgaris* allergens, such as the phospholipase A1 allergen Ves v 1, the hyaluronidase allergen Ves v 2, and the dipeptidyl peptidase IV-like serine protease allergen Ves v 3, was not successful with or without MBP despite extensive optimization. Implying that this workflow is more suitable for small, and minimally glycosylated proteins; and is not suitable for more catalytically active, heavily glycosylated enzymes that could cause the ER stress, misfolding, or host toxicity. These findings highlight the strong dependence of recombinant allergen yield in *P. pastoris* on the intrinsic compatibility between protein structure, post-translational requirements, and the yeast secretory machinery.

SUMMARY

In this study, recombinant forms of three clinically important allergens were produced and characterized: the black tiger shrimp *P. monodon* allergen Pen m 4, the dog *C. familiaris* allergen Can f 6, and the yellow jacket *V. vulgaris* allergen Ves v 5. For Pen m 4 and Can f 6, comparison between *E. coli* and *P. pastoris* demonstrated that both platforms could produce properly folded allergens with retained IgE-binding activity, although purification strategy could markedly influence it. N-glycosylation was confirmed as non-essential for IgE recognition. For Ves v 5, *P. pastoris* variants displayed native-like antigenicity, whereas *E. coli*-produced proteins exhibited non-specific binding. Expression strategies included the first application of MBP fusion in *P. pastoris* for recombinant allergen production.

Several novel findings also emerged. First, MBP fusion proteins in *P. pastoris* were unexpectedly susceptible to proteolysis, with two patterns observed: cleavage downstream of the MBP domain, marked by lower MW fragments reactive with anti-MBP antibodies, and proteolysis of the N-terminal 6xHis-containing region, inferred from anti-6xHis WB and MS analyses. The first pattern supports previous reports of MBP-associated proteolytic activity affecting linker/spacer and cargo regions (Z Li et al. 2010; Moua et al. 2016), whereas the second appears to represent a novel vulnerability in multiple *P. pastoris*-produced MBP-allergen fusion proteins. Together, these findings highlight additional constraints for MBP-fusion protein design in *P. pastoris*.

Second, *P. pastoris*-produced Pen m 4 formed stable glycosylation-modulated dimers resistant to standard denaturing conditions, while EDTA anticoagulant reduced Pen m 4 sIgE binding through calcium chelation, revealing an overlooked preanalytical variable. Finally, both *E. coli*-produced Pen m 4 and Can f 6 degraded during lyophilization, explaining the instability of these allergens in commercial extracts.

Together, these findings establish evidence-based protocols for recombinant allergen production and highlight critical variables, including fusion partner stability and sample handling, that influence diagnostic reliability. This work confirms Pen m 4, Can f 6, and Ves v 5 as clinically important allergens suitable for component-resolved diagnostics, and validates *P. pastoris* as a viable production system.

CONCLUSIONS

1. *P. pastoris* and *E. coli* expression systems can produce functionally equivalent recombinant Pen m 4 and Can f 6 allergens, that retain proper folding structure required for IgE recognition. For Ves v 5, only production in *P. pastoris* retained native-like antigenicity.

2. Comparative analysis demonstrated that N-glycosylation is not essential for IgE recognition of Pen m 4 and Can f 6 allergens but posttranslational modifications are essential for allergen Ves v 5.

3. Fusion to the maltose-binding protein (MBP) enhanced protein yield in both expression systems without compromising IgE-binding functionality, validating MBP as a useful tool for recombinant allergen production.

4. Cleavage downstream of the MBP domain observed in *P. pastoris*-produced MBP fusion proteins supports previous reports of MBP-associated proteolytic activity affecting regions C-terminal to the MBP domain.

5. Novel proteolytic processing patterns of MBP fusion proteins in *P. pastoris*, including loss of the N-terminal 6xHis-containing region, were identified and highlight new constraints for fusion protein design in yeast expression systems.

6. *P. pastoris*-produced Pen m 4 formed unusually stable dimers resistant to denaturing conditions, revealing a glycosylation-modulated structural feature with potential diagnostic applications.

7. EDTA anticoagulant reduced IgE recognition of Pen m 4 by 61-73% through calcium chelation, underscoring the need for standardized sample handling.

8. *E. coli*-produced Pen m 4 and Can f 6 were found to degrade during lyophilization, providing a molecular explanation for their instability in commercial extracts and emphasizing the importance of optimized storage strategies.

BIBLIOGRAPHY

1. Ahmad M, Hirz M, Pichler H, and Schwab H. Protein Expression in *Pichia pastoris*: Recent Achievements and Perspectives for Heterologous Protein Production. Appl Microbiol Biotechnol. 2014; 98:5301–17. DOI: 10.1007/s00253-014-5732-5. PMID: 24743983
2. Alaskhar Alhamwe B et al. Histone Modifications and Their Role in Epigenetics of Atopy and Allergic Diseases. Allergy Asthma Clin Immunol. 2018; 14:1–16. DOI: 10.1186/s13223-018-0259-4
3. Altmann F. Coping with Cross-Reactive Carbohydrate Determinants in Allergy Diagnosis. Allergo J Int. 2016; 25:98–105. DOI: 10.1007/s40629-016-0112-6
4. AMVA. American Veterinary Medical Association (AVMA) Pet Ownership and Demographic Sourcebook. 2024
5. Anto JM et al. Mechanisms of the Development of Allergy (MeDALL): Introducing Novel Concepts in Allergy Phenotypes. J Allergy Clin Immunol. 2017; 139:388–99. DOI: 10.1016/j.jaci.2016.12.940. PMID: 28183433
6. Anvari S, Miller J, Yeh CY, and Davis CM. IgE-Mediated Food Allergy. Clinic Rev Allerg Immunol. 2019; 57:244–60. DOI: 10.1007/s12016-018-8710-3. PMID: 30370459
7. Arnau-Soler A et al. Food Allergy Genetics and Epigenetics: A Review of Genome-Wide Association Studies. Allergy. 2025; 80:106–31. DOI: 10.1111/all.16429
8. Asero R, Pravettoni V, Scala E, and Villalta D. House Dust Mite-Shrimp Allergen Interrelationships. Curr Allergy Asthma Rep. 2020; 20:3–7. DOI: 10.1007/s11882-020-0902-2. PMID: 32144500
9. Asero R, Scala E, et al. Shrimp Allergy: Analysis of Commercially Available Extracts for *in vivo* Diagnosis. J Investig Allergol Clin Immunol. 2017; 27:175–82. DOI: 10.18176/jiaci.0127. PMID: 27959286
10. Atiakshin D et al. Carboxypeptidase A3 - A Key Component of the Protease Phenotype of Mast Cells. Cells. 2022; 11. DOI: 10.3390/cells11030570. PMID: 35159379
11. Ayuso R et al. Sarcoplasmic Calcium-Binding Protein Is an EF-hand-type Protein Identified as a New Shrimp Allergen. J Allergy Clin Immunol. 2009; 124:114–20. DOI: 10.1016/j.jaci.2009.04.016. PMID: 19523674
12. Barber D et al. Molecular Allergology and Its Impact in Specific Allergy Diagnosis and Therapy. Allergy. 2021; 76:3642–58. DOI: 10.1111/all.14969
13. Barnes PJ. Pathophysiology of Allergic Inflammation. Immunol Rev. 2011; 242:31–50. DOI: 10.1111/j.1600-065X.2011.01020.x.
14. Baseggio Conrado A et al. Food Anaphylaxis in the United Kingdom: Analysis of National Data, 1998-2018. BMJ. 2021; 372. DOI: 10.1136/bmj.n251. PMID: 33597169
15. Bazon ML, Silveira LH, Simioni PU, and Brochetto-Braga MR. Current Advances in Immunological Studies on the Vespidae Venom Antigen 5: Therapeutic and Prophylaxis to Hypersensitivity Responses. Toxins. 2018; 10. DOI: 10.3390/toxins10080305. PMID: 30042313
16. Bell MR, Engleka MJ, Malik A, and Strickler JE. To Fuse or Not to Fuse: What Is Your Purpose? Protein Sci. 2013; 22:1466–77. DOI: 10.1002/pro.2356. PMID: 24038604
17. Bensnes SS. You Sneeze, You Lose. The Impact of Pollen Exposure on Cognitive Performance during High-Stakes High School Exams. J Health Econ. 2016; 49:1–13. DOI: 10.1016/j.jhealeco.2016.05.005. PMID: 27315202

18. Berin MC. Immunopathophysiology of Food Protein-Induced Enterocolitis Syndrome. J Allergy Clin Immunol. 2015; 135:1108–13. DOI: 10.1016/j.jaci.2014.12.1948. PMID: 25746969
19. Bilò MB et al. Large Local Reactions to Hymenoptera Stings: Outcome of Re-stings in Real Life. Allergy. 2019; 74:1969–76. DOI: 10.1111/all.13863
20. Blank S, Haemmerle S, et al. Prevalence of Hymenoptera Venom Allergy and Sensitization in the Population-Representative German KORA Cohort. Allergo J Int. 2019; 28:183–91. DOI: 10.1007/s40629-018-0089-4
21. Blank S, Michel Y, et al. Evaluation of Different Glycoforms of Honeybee Venom Major Allergen Phospholipase A2 (Api m 1) Produced in Insect Cells. Protein Pept Lett. 2011; 18:415–22. DOI: 10.2174/092986611794653923. PMID: 21171948
22. Bønnelykke K, Sparks R, Waage J, and Milner JD. Genetics of Allergy and Allergic Sensitization: Common Variants, Rare Mutations. Curr Opin Immunol. 2015; 36(Figure 1):115–26. DOI: 10.1016/j.coi.2015.08.002. PMID: 26386198
23. Bonnet B et al. An Update on Molecular Cat Allergens: Fel d 1 and What Else? Chapter 1: Fel d 1, the Major Cat Allergen. Allergy Asthma Clin Immunol. 2018; 14:14. DOI: 10.1186/s13223-018-0239-8
24. Borodina I, Jensen BM, Søndergaard I, and Poulsen LK. Display of Wasp Venom Allergens on the Cell Surface of *Saccharomyces cerevisiae*. Microb Cell Fact. 2010; 9:74. DOI: 10.1186/1475-2859-9-74
25. Boyce JA. Mast Cells and Eicosanoid Mediators: A System of Reciprocal Paracrine and Autocrine Regulation. Immunol Rev. 2007; 217:168–85
26. Branco ACCC, Yoshikawa FSY, Pietrobon AJ, and Sato MN. Role of Histamine in Modulating the Immune Response and Inflammation. Mediators Inflamm. 2018; 2018. DOI: 10.1155/2018/9524075. PMID: 30224900
27. Brondyk WH. Chapter 11 Selecting an Appropriate Method for Expressing a Recombinant Protein. Methods Enzymol. 2009; 463:131–47. DOI: 10.1016/S0076-6879(09)63011-1. PMID: 19892171
28. Brunetto S, Buta F, Gangemi S, and Ricciardi L. Case Report: An Unusual Case of Cardiac Anaphylaxis in the Maintenance Phase of Vespula Venom Immunotherapy. Front Allergy. 2025; 6:1583909. DOI: 10.3389/falgy.2025.1583909
29. Calderone V, Berni R, and Zanotti G. High-Resolution Structures of Retinol-binding Protein in Complex with Retinol: pH-induced Protein Structural Changes in the Crystal State. J Mol Biol. 2003; 329:841–50. DOI: 10.1016/s0022-2836(03)00468-6
30. Carli T, Locatelli I, Košnik M, and Kukec A. Epidemiology and Risk Factors of Self-Reported Systemic Allergic Reactions to a Hymenoptera Venom in Beekeepers Worldwide: A Protocol for a Systematic Review of Observational Studies. BMJ Open. 2022; 12:e058130. DOI: 10.1136/bmjopen-2021-058130
31. Chan SK and Leung DY. Dog and Cat Allergies: Current State of Diagnostic Approaches and Challenges. Allergy Asthma Clin Immunol. 2018; 10:97–105. DOI: 10.4168/aair.2018.10.2.97
32. Chen LL et al. Effect of Ozone Exposure on Airway Responses to Inhaled Allergen in Asthmatic Subjects. Chest. 2004; 125:2328–35. DOI: 10.1378/chest.125.6.2328. PMID: 15189958
33. Chronopoulou S et al. Expression and Purification of Human Interferon Alpha 2a (IFNα2a) in the Methylotrophic Yeast *Pichia pastoris*. Protein Expr Purif. 2023; 211:106339. DOI: 10.1016/j.pep.2023.106339
34. Cianferoni A and Spergel JM. Food Allergy: Review, Classification and Diagnosis. Allergol Int. 2009; 58:457–66. DOI: DOI:10.2332!allergolint.09-RAI-0138

35. Connors L, O'Keefe A, Rosenfield L, and Kim H. Non-IgE-mediated Food Hypersensitivity. Allergy Asthma Clin Immunol. 2018; 14:1–9. DOI: 10.1186/s13223-018-0285-2
36. Costa J et al. Are Physicochemical Properties Shaping the Allergenic Potency of Animal Allergens? Clinic Rev Allerg Immunol. 2022; 62:1–36. DOI: 10.1007/s12016-020-08826-1
37. Crowther JR. The ELISA Guidebook. Vol. 149. 2000. DOI: 10.1385/1592590497. PMID: 11028258
38. Curin M, Garib V, and Valenta R. Single Recombinant and Purified Major Allergens and Peptides: How They Are Made and How They Change Allergy Diagnosis and Treatment. Ann Allergy Asthma Immunol. 2017; 119:201–9. DOI: 10.1016/j.anai.2016.11.022. PMID: 28890016
39. Currie AJ, Stewart GA, and McWilliam AS. Alveolar Macrophages Bind and Phagocytose Allergen-Containing Pollen Starch Granules Via C-Type Lectin and Integrin Receptors: Implications for Airway Inflammatory Disease. J Immunol. 2000; 164:3878–86. DOI: 10.4049/jimmunol.164.7.3878. PMID: 10725750
40. Custovic A et al. Evolution Pathways of IgE Responses to Grass and Mite Allergens throughout Childhood. J Allergy Clin Immunol. 2015; 136:1645–1652.e8. DOI: 10.1016/j.jaci.2015.03.041
41. D'amato G, Cecchi L, Amato D', and Liccardi G. Urban Air Pollution and Climate Change as Environmental Risk Factors of Respiratory Allergy: An Update. 2. 2010 :95–102
42. D'Amato G, Liccardi G, and Frenguelli G. Thunderstorm-Asthma and Pollen Allergy. Allergy. 2007; 62:11–6. DOI: 10.1111/j.1398-9995.2006.01271.x. PMID: 17156336
43. De Oliveira FC et al. Food Protein-polysaccharide Conjugates Obtained via the Maillard Reaction: A Review. Crit Rev Food Sci Nutr. 2016; 56:1108–25. DOI: 10.1080/10408398.2012.755669. PMID: 24824044
44. Dhaliwal B et al. IgE Binds Asymmetrically to Its B Cell Receptor CD23. Sci Rep. 2017; 7:45533. DOI: 10.1038/srep45533
45. Durham SR and Shamji MH. Allergen Immunotherapy: Past, Present and Future. Nat Rev Immunol. 2023; 23:317–28. DOI: 10.1038/s41577-022-00786-1
46. Dvorak AM. Ultrastructural Studies of Human Basophils and Mast Cells. J Histochem Cytochem. 2005; 53:1043–70. DOI: 10.1369/jhc.5R6647.2005. PMID: 15923353
47. Ebihara N et al. Mast Cell Chymase Decreases the Barrier Function and Inhibits the Migration of Corneal Epithelial Cells. Curr Eye Res. 2005; 30:1061–9. DOI: 10.1080/02713680500346625. PMID: 16354619
48. FEDIAF. European Pet Food Industry (FEDIAF) Annual Report. 2024
49. Fekete S, Beck A, Veuthey JL, and Guillarme D. Ion-Exchange Chromatography for the Characterization of Biopharmaceuticals. J Pharm Biomed Anal. 2015; 113:43–55. DOI: 10.1016/j.jpba.2015.02.037
50. Fernández-Caldas E, Cases B, El-Qutob D, and Cantillo JF. Mammalian Raw Materials Used to Produce Allergen Extracts. Ann Allergy Asthma Immunol. 2017; 119:1–8. DOI: 10.1016/j.anai.2016.08.024. PMID: 28668236
51. Ferreira MA et al. Shared Genetic Origin of Asthma, Hay Fever and Eczema Elucidates Allergic Disease Biology. Nat Gen. 2017; 49:1752–7. DOI: 10.1038/ng.3985. PMID: 29083406
52. Flower DR, North AC, and Sansom CE. The Lipocalin Protein Family: Structural and Sequence Overview. Biochim Biophys Acta Biomembr. 2000; 1482:9–24. DOI: 10.1016/s0167-4838(00)00148-5

53. Floyd ML, Adams KE, and Golden DBK. Updates and Recent Advances on Venom Immunotherapy. Curr Treat Options Allergy. 2023; 10:196–214. DOI: 10.1007/s40521-023-00336-7
54. Fogolari F et al. MM/PBSA Analysis of Molecular Dynamics Simulations of Bovine β-lactoglobulin: Free Energy Gradients in Conformational Transitions? Proteins Struct Funct Genet. 2005; 59:91–103. DOI: 10.1002/prot.20384
55. Galli SJ, Grimbaldeston M, and Tsai M. Immunomodulatory Mast Cells: Negative, as Well as Positive, Regulators of Immunity. Nat Rev Immunol. 2008; 8:478–86
56. Galli SJ, Kalesnikoff J, et al. Mast Cells as "Tunable" Effector and Immunoregulatory Cells: Recent Advances. Annu Rev Immunol. 2005; 23:749–86. DOI: 10.1146/annurev.immunol.21.120601.141025. PMID: 15771585
57. Galli SJ and Tsai M. IgE and Mast Cells in Allergic Disease. Nat Med. 2012; 18:693–704. DOI: 10.1038/nm.2755. PMID: 22561833
58. Galli SJ, Tsai M, and Piliponsky AM. The Development of Allergic Inflammation. Nature. 2008; 454:445–54. DOI: 10.1038/nature07204. PMID: 18650915
59. Gámez C et al. Tropomyosin IgE-positive Results Are a Good Predictor of Shrimp Allergy. Allergy. 2011; 66:1375–83. DOI: 10.1111/j.1398-9995.2011.02663.x. PMID: 21651567
60. Gao J et al. Biosynthesis of Catharanthine in Engineered *Pichia pastoris*. Nat Synth. 2023; 2:231–42. DOI: 10.1038/s44160-022-00205-2
61. Gao Y, Gillen CM, and Wheatly MG. Molecular Characterization of the Sarcoplasmic Calcium-Binding Protein (SCP) from Crayfish *Procambarus clarkii*. Comp Biochem Physiol B Biochem Mol Biol. 2006; 144:478–87. DOI: 10.1016/j.cbpb.2006.04.007. PMID: 16807031
62. Gast V et al. Engineering *Saccharomyces cerevisiae* for the Production and Secretion of Affibody Molecules. Microb Cell Fact. 2022; 21. DOI: 10.1186/s12934-022-01761-0
63. Gasymov OK, Abduragimov AR, and Glasgow BJ. pH-Dependent Conformational Changes in Tear Lipocalin by Site-Directed Tryptophan Fluorescence. Biochemistry. 2010; 49:582–90. DOI: 10.1021/bi901435q
64. Al-Ghouleh A et al. The Glycosylation Pattern of Common Allergens: The Recognition and Uptake of Der p 1 by Epithelial and Dendritic Cells Is Carbohydrate Dependent. PLoS ONE. 2012; 7. DOI: 10.1371/journal.pone.0033929. PMID: 22479478
65. Golden DB et al. AAAAI/ACAAI Joint Venom Extract Shortage Task Force Report. Ann Allergy Asthma Immunol. 2017; 118:283–5. DOI: 10.1016/j.anai.2017.02.001
66. Gould HJ and Sutton BJ. IgE in Allergy and Asthma Today. Nat Rev Immunol. 2008; 8:205–17. DOI: 10.1038/nri2273. PMID: 18301424
67. Hale RC, Morais D, Chou J, and Stowell SR. The Role of Glycosylation in Clinical Allergy and Immunology. J Allergy Clin Immunol. 2024; 153:55–66. DOI: 10.1016/j.jaci.2023.09.003
68. Hamby SE and Hirst JD. Prediction of Glycosylation Sites Using Random Forests. BMC Bioinform. 2008; 9:1–13. DOI: 10.1186/1471-2105-9-500. PMID: 19038042
69. Hammad H and Lambrecht BN. Dendritic Cells and Epithelial Cells: Linking Innate and Adaptive Immunity in Asthma. Nat Rev Immunol. 2008; 8:193–204. DOI: 10.1038/nri2275. PMID: 18301423
70. Heinzerling LM et al. GA ^{2} LEN Skin Test Study I: GA^{2}LEN Harmonization of Skin Prick Testing: Novel Sensitization Patterns for Inhalant Allergens in Europe. Allergy. 2009; 64:1498–506. DOI: 10.1111/j.1398-9995.2009.02093.x
71. Heinzerling L et al. The Skin Prick Test – European Standards. Clin Transl Allergy. 2013; 3:3. DOI: 10.1186/2045-7022-3-3

72. Hemmer W et al. ImmunoCAP Cellulose Displays Cross-Reactive Carbohydrate Determinant (CCD) Epitopes and Can Cause False-Positive Test Results in Patients with High Anti-CCD IgE Antibody Levels. J Allergy Clin Immunol. 2018; 141:372–381.e3. DOI: 10.1016/j.jaci.2017.04.028. PMID: 28506851
73. Henriksen A et al. Major Venom Allergen of Yellow Jackets, Ves v 5: Structural Characterization of a Pathogenesis-Related Protein Superfamily. Proteins Struct Funct Genet. 2001; 45:438–48. DOI: 10.1002/prot.1160. PMID: 11746691
74. Hilger C, Kuehn A, and Hentges F. Animal Lipocalin Allergens. Curr Allergy Asthma Rep. 2012; 12:438–47. DOI: 10.1007/s11882-012-0283-2. PMID: 22791068
75. Homann A, Schramm G, and Jappe U. Glycans and Glycan-Specific IgE in Clinical and Molecular Allergology: Sensitization, Diagnostics, and Clinical Symptoms. J Allergy Clin Immunol. 2017; 140:356–68. DOI: 10.1016/j.jaci.2017.04.019. PMID: 28479330
76. Homas T et al. Siblings, Day-Care Attendance, and the Risk of Asthma and Wheezing during Childhood. N Engl J Med. 2000; 343:538–43
77. Hong J et al. Ambient Air Pollution, Weather Changes and Outpatient Visits for Allergic Conjunctivitis: A Retrospective Registry Study. Sci Rep. 2016; 6:23858. DOI: 10.1038/srep23858
78. Hong X et al. Genome-Wide Association Study Identifies Peanut Allergy-Specific Loci and Evidence of Epigenetic Mediation in US Children. Nat Commun. 2015; 6. DOI: 10.1038/ncomms7304. PMID: 25710614
79. Hopkins D et al. Elimination of Diaminopeptidase Activity in *Pichia pastoris* for Therapeutic Protein Production. Appl Microbiol Biotechnol. 2014; 98:2573–83. DOI: 10.1007/s00253-013-5468-7. PMID: 24526360
80. Hulo N et al. The PROSITE Database. Nucl Acids Res. 2006; 34:227–30. DOI: 10.1093/nar/gkj063. PMID: 16381852
81. Hwang PM et al. Solution Structure and Dynamics of the Outer Membrane Enzyme PagP by NMR. Proc Natl Acad Sci USA. 2002; 99:13560–5. DOI: 10.1073/pnas.212344499
82. Jafri L, Khan AH, and Azeem S. Ionized Calcium Measurement in Serum and Plasma by Ion Selective Electrodes: Comparison of Measured and Calculated Parameters. Indian J Clin Biochem. 2014; 29:327–32. DOI: 10.1007/s12291-013-0360-x
83. Jappe U and Schwager C. Relevance of Lipophilic Allergens in Food Allergy Diagnosis. Curr Allergy Asthma Rep. 2017; 17:1–9. DOI: 10.1007/s11882-017-0731-0. PMID: 28795292
84. Jayakrishnan A et al. Evolving Paradigms of Recombinant Protein Production in Pharmaceutical Industry: A Rigorous Review. Sci. 2024; 6:9. DOI: 10.3390/sci6010009
85. Jiang T, Kang C, Yu XW, and Xu Y. High-Level Expression of Prolyl Endopeptidase in *Pichia pastoris* Using PLA_2 as a Fusion Partner. J Mol Catal B Enzym. 2016; 125:81–7. DOI: 10.1016/j.molcatb.2016.01.005
86. Jimenez-Lopez JC et al. Specific Conformational Epitope Features of Pathogenesis-Related Proteins Mediating Cross-Reactivity between Pollen and Food Allergens. Mol Biol Rep. 2012; 39:123–30. DOI: 10.1007/s11033-011-0717-2. PMID: 21598115
87. Johnston EB et al. Defining Specific Allergens for Improved Component-Resolved Diagnosis of Shrimp Allergy in Adults. Mol Immunol. 2019; 112(April):330–7. DOI: 10.1016/j.molimm.2019.05.006. PMID: 31247376
88. Kamalakannan M et al. Identification and Characterization of DC-SIGN-binding Glycoproteins in Allergenic Foods. Allergy. 2016; 71:1145–55. DOI: 10.1111/all.12873. PMID: 26948687

89. Kamata Y et al. Characterization of Dog Allergens Can f 1 and Can f 2. 1. Preparation of Their Recombinant Proteins and Antibodies. Int Arch Allergy Immunol. 2007; 142:291–300. DOI: 10.1159/000097498
90. Khurana T, Newman-Lindsay S, Young PR, and Slater JE. The NPC2 Protein: A Novel Dog Allergen. Ann Allergy Asthma Immunol. 2016; 116:440–446.e2. DOI: 10.1016/j.anai.2016.02.006. PMID: 26995145
91. Kirjavainen PV et al. Farm-like Indoor Microbiota in Non-Farm Homes Protects Children from Asthma Development. Nat Med. 2019; 25:1089–95. DOI: 10.1038/s41591-019-0469-4
92. Kischnick S et al. Bacterial Fermentation of Recombinant Major Wasp Allergen Antigen 5 Using Oxygen Limiting Growth Conditions Improves Yield and Quality of Inclusion Bodies. Protein Expr Purif. 2006; 47:621–8. DOI: 10.1016/j.pep.2006.01.009. PMID: 16495080
93. Kivistö JE et al. Genetic and Environmental Susceptibility to Food Allergy in a Registry of Twins. J Allergy Clin Immunol Pract. 2019; 7:2916–8. DOI: 10.1016/j.jaip.2019.05.016
94. Konieczny A et al. The Major Dog Allergens, Can f 1 and Can f 2, Are Salivary Lipocalin Proteins: Cloning and Immunological Characterization of the Recombinant Forms. Immunology. 1997; 92:577–86. DOI: 10.1046/j.1365-2567.1997.00386.x
95. Konradsen JR et al. Allergy to Furry Animals: New Insights, Diagnostic Approaches, and Challenges. J Allergy Clin Immunol. 2015; 135:616–25. DOI: 10.1016/j.jaci.2014.08.026. PMID: 25282018
96. Kopp EB et al. Modes of Type 2 Immune Response Initiation. Immunity. 2023; 56:687–94. DOI: 10.1016/j.immuni.2023.03.015
97. Kostner L et al. Allergic Contact Dermatitis. Immunol Allergy Clin North Am. 2017; 37:141–52. DOI: 10.1016/j.iac.2016.08.014. PMID: 27886903
98. Kowalski ML et al. Risk and Safety Requirements for Diagnostic and Therapeutic Procedures in Allergology: World Allergy Organization Statement. World Allergy Organ J. 2016; 9:1–42. DOI: 10.1186/s40413-016-0122-3
99. Kozlov E et al. Recombinant IgE-reactive Functional Can f 5 Devoid of Cross-reactive Carbohydrate Determinants. Allergy. 2025 :1–5. DOI: 10.1111/all.16465
100. Kraft S and Kinet JP. New Developments in FcεRI Regulation, Function and Inhibition. Nat Rev Immunol. 2007; 7:365–78. DOI: 10.1038/nri2072. PMID: 17438574
101. Kucinskaite-Kodze I et al. Characterization of Monoclonal Antibodies against Hantavirus Nucleocapsid Protein and Their Use for Immunohistochemistry on Rodent and Human Samples. Arch Virol. 2011; 156:443–56. DOI: 10.1007/s00705-010-0879-6. PMID: 21161552
102. Lambrecht BN. Allergen Uptake and Presentation by Dendritic Cells. Curr Opin Allergy Clin Immunol. 2001; 1:51–9
103. Lambrecht BN. Allergy-Triggering Proteins Share a Common Characteristic. Nature. 2025; 645:323–5. DOI: 10.1038/d41586-025-02136-2. PMID: 40739007
104. Lambrecht BN and Hammad H. The Immunology of the Allergy Epidemic and the Hygiene Hypothesis. Nat Immunol. 2017; 18:1076–83. DOI: 10.1038/ni.3829. PMID: 28926539
105. Lee KH et al. The Gut Microbiota, Environmental Factors, and Links to the Development of Food Allergy. Clin Mol Allergy. 2020; 18. DOI: 10.1186/s12948-020-00120-x. PMID: 32265597
106. Lehmann K et al. High-Yield Expression in *Escherichia coli*, Purification, and Characterization of Properly Folded Major Peanut Allergen Ara h 2. Protein Expr Purif. 2003; 31:250–9. DOI: 10.1016/S1046-5928(03)00190-6. PMID: 14550644
107. Li F et al. Sneezing Reflex Is Mediated by a Peptidergic Pathway from Nose to Brainstem. Cell. 2021; 184:3762–3773.e10. DOI: 10.1016/j.cell.2021.05.017. PMID: 34133943

108. Li Z et al. Secretion and Proteolysis of Heterologous Proteins Fused to the *Escherichia coli* Maltose Binding Protein in *Pichia pastoris*. Protein Expr Purif. 2010; 72:113–24. DOI: 10.1016/j.pep.2010.03.004. PMID: 20230898
109. Licari A et al. Epidemiology of Allergic Rhinitis in Children: A Systematic Review and Meta-Analysis. J Allergy Clin Immunol Pract. 2023; 11:2547–56. DOI: 10.1016/j.jaip.2023.05.016
110. Lima M and Baynes JW. Glycation. Encycl Biol Chem Second Ed. 2013; 2:405–11. DOI: 10.1016/B978-0-12-378630-2.00120-1. PMID: 20446573
111. Lin HP et al. Identification of Novel Macropinocytosis Inhibitors Using a Rational Screen of Food and Drug Administration-approved Drugs. Br J Pharmacol. 2018; 175:3640–55. DOI: 10.1111/bph.14429. PMID: 29953580
112. Liu M et al. A Comprehensive Review of Recombinant Technology in the Food Industry: Exploring Expression Systems, Application, and Future Challenges. Comp Rev Food Sci Food Safe. 2025; 24. DOI: 10.1111/1541-4337.70078
113. Liu X et al. Genetic and Environmental Contributions to Allergen Sensitization in a Chinese Twin Study. Clin Exp Allergy. 2009; 39:991–8. DOI: 10.1111/j.1365-2222.2009.03228.x
114. London SJ et al. Family History and the Risk of Early-Onset Persistent, Early-Onset Transient, and Late-Onset Asthma. Epidemiology. 2001; 12:577–83. DOI: 10.1097/00001648-200109000-00019
115. Lopata AL, Kleine-Tebbe J, and Kamath SD. Allergens and Molecular Diagnostics of Shellfish Allergy. Allergo J Int. 2017; 26:118. DOI: 10.1007/s40629-016-0124-2
116. Lu CW et al. Air Pollution and Meteorological Conditions Significantly Contribute to the Worsening of Allergic Conjunctivitis: A Regional 20-City, 5-Year Study in Northeast China. Light Sci Appl. 2021; 10:190. DOI: 10.1038/s41377-021-00630-6
117. Lukacs NW and Hogan SP. Food Allergy: Begin at the Skin, End at the Mast Cell? Nat Rev Immunol. 2025. DOI: 10.1038/s41577-025-01185-y
118. Magee CN, Boenisch O, and Najafian N. The Role of Costimulatory Molecules in Directing the Functional Differentiation of Alloreactive T Helper Cells. Am J Transplant. 2012; 12:2588–600. DOI: 10.1111/j.1600-6143.2012.04180.x. PMID: 22759274
119. Marcotte DE. Allergy Test: Seasonal Allergens and Performance in School. J Health Econ. 2015; 40:132–40. DOI: 10.1016/j.jhealeco.2015.01.002. PMID: 25680109
120. Matricardi PM et al. EAACI Molecular Allergology User's Guide. Pediatr Allergy Immunol. 2016; 27:1–250. DOI: 10.1111/pai.12563. PMID: 27288833
121. Mattsson L, Lundgren T, Olsson P, et al. Molecular and Immunological Characterization of Can f 4: A Dog Dander Allergen Cross-Reactive with a 23 kDa Odorant-Binding Protein in Cow Dander. Clin Exp Allergy. 2010; 40:1276–87. DOI: 10.1111/j.1365-2222.2010.03533.x. PMID: 20545700
122. Mattsson L, Lundgren T, Everberg H, et al. Prostatic Kallikrein: A New Major Dog Allergen. J Allergy Clin Immunol. 2009; 123. DOI: 10.1016/j.jaci.2008.11.021. PMID: 19135239
123. Maupin KA, Liden D, and Haab BB. The Fine Specificity of Mannose-Binding and Galactose-Binding Lectins Revealed Using Outlier Motif Analysis of Glycan Array Data. Glycobiol. 2012; 22:160–9. DOI: 10.1093/glycob/cwr128. PMID: 21875884
124. Maurer M and Zuberbier T. Undertreatment of Rhinitis Symptoms in Europe: Findings from a Cross-Sectional Questionnaire Survey. Allergy. 2007; 62:1057–63. DOI: 10.1111/j.1398-9995.2007.01367.x. PMID: 17581263

125. Mazur M, Czarnobilska M, Dyga W, and Czarnobilska E. Trends in the Epidemiology of Allergic Diseases of the Airways in Children Growing Up in an Urban Agglomeration. J Clin Med. 2022; 11. DOI: 10.3390/jcm11082188
126. Mckeever TM et al. Siblings, Multiple Births, and the Incidence of Allergic Disease: A Birth Cohort Study Using the West Midlands General Practice Research Database. Thorax. 2001; 56:758–62
127. McKibbin LR et al. Fatal Anaphylaxis Due to Peanut Exposure from Oral Intercourse. Allergy Asthma Clin Immunol. 2021; 17. DOI: 10.1186/s13223-021-00611-9
128. McLeod JJ, Baker B, and Ryan JJ. Mast Cell Production and Response to IL-4 and IL-13. Cytokine. 2015; 75:57–61. DOI: 10.1016/j.cyto.2015.05.019. PMID: 26088754
129. Midodzi WK, Rowe BH, Majaesic CM, and Senthilselvan A. Reduced Risk of Physician-Diagnosed Asthma among Children Dwelling in a Farming Environment. Respirology. 2007; 12:692–9. DOI: 10.1111/j.1440-1843.2007.01134.x. PMID: 17875057
130. Mita H, Koketsu A, Ishizaki S, and Shiomi K. Molecular Cloning and Functional Expression of Allergenic Sarcoplasmic Calcium-Binding Proteins from Penaeus Shrimps. J Sci Food Agric. 2013; 93:1737–42. DOI: 10.1002/jsfa.5961. PMID: 23180551
131. Monsalve RI, Lu G, and King TP. Expressions of Recombinant Venom Allergen, Antigen 5 of Yellowjacket (*Vespula vulgaris*) and Paper Wasp (*Polistes annularis*), in Bacteria or Yeast. Protein Expr Purif. 1999; 16:410–6. DOI: 10.1006/prep.1999.1082
132. Morii A, Mita H, Ishizaki S, and Shiomi K. Importance of Conformation for the IgE Reactivity of Sarcoplasmic Calcium-Binding Protein from the Black Tiger Shrimp *Penaeus monodon*. Eur Food Res Technol. 2013; 236:165–70. DOI: 10.1007/s00217-012-1867-8
133. Motosue MS, Li JT, and Campbell RL. Anaphylaxis. Immunol Allergy Clin North Am. 2022; 42:13–25. DOI: 10.1016/j.iac.2021.09.010
134. Moua PS et al. Differential Secretion Pathways of Proteins Fused to the *Escherichia coli* Maltose Binding Protein (MBP) in *Pichia pastoris*. Protein Expr Purif. 2016; 124:1–9. DOI: 10.1016/j.pep.2016.04.005. PMID: 27079175
135. Müller U, Schmid-Grendelmeier P, Hausmann O, and Helbling A. IgE to Recombinant Allergens A Pi m 1, V Es v 1, and V Es v 5 Distinguish Double Sensitization from Crossreaction in Venom Allergy. Allergy. 2012; 67:1069–73. DOI: 10.1111/j.1398-9995.2012.02847.x
136. Müller UR. Recombinant Hymenoptera Venom Allergens. Allergy. 2002; 57:570–6. DOI: 10.1034/j.1398-9995.2002.02157.x
137. Murray PR, Rosenthal KS, and Pfaller MA. Medical Microbiology. 7th ed. 2013. 874 pp.
138. Nallamsetty S, Austin BP, Penrose KJ, and Waugh DS. Gateway Vectors for the Production of Combinatorially-Tagged His_6-MBP Fusion Proteins in the Cytoplasm and Periplasm of *Escherichia coli*. Protein Sci. 2005; 14:2964–71. DOI: 10.1110/ps.051718605. PMID: 16322578
139. Nallamsetty S and Waugh DS. Solubility-Enhancing Proteins MBP and NusA Play a Passive Role in the Folding of Their Fusion Partners. Protein Expr Purif. 2006; 45:175–82. DOI: 10.1016/j.pep.2005.06.012. PMID: 16168669
140. Nelson M and Ogden J. An Exploration of Food Intolerance in the Primary Care Setting: The General Practitioner's Experience. Soc Sci Med. 2008; 67:1038–45. DOI: 10.1016/j.socscimed.2008.05.025. PMID: 18584930
141. Newcomer ME and Ong DE. Plasma Retinol Binding Protein: Structure and Function of the Prototypic Lipocalin. Biochim Biophys Acta Biomembr. 2000; 1482:57–64. DOI: 10.1016/s0167-4838(00)00150-3

142. Nilsson OB et al. Characterization of the Dog Lipocalin Allergen Can f 6: The Role in Cross-Reactivity with Cat and Horse. Allergy. 2012; 67:751–7. DOI: 10.1111/j.1398-9995.2012.02826.x. PMID: 22515174
143. Noirey N et al. Langerhans-like Dendritic Cells Generated from Cord Blood Progenitors Internalize Pollen Allergens by Macropinocytosis, and Part of the Molecules Are Processed and Can Activate Autologous Naive T Lymphocytes. J Allergy Clin Immunol. 2000; 105(6 II):1194–201. DOI: 10.1067/mai.2000.106545. PMID: 10856155
144. Nowak-Węgrzyn A, Katz Y, Mehr SS, and Koletzko S. Non-IgE-mediated Gastrointestinal Food Allergy. J Allergy Clin Immunol. 2015; 135:1114–24. DOI: 10.1016/j.jaci.2015.03.025. PMID: 25956013
145. Nugraha R et al. Effects of Extraction Buffer on the Solubility and Immunoreactivity of the Pacific Oyster Allergens. Foods. 2021; 10(2 1):16. DOI: 10.3390/foods10020409
146. Obst U, Lu TK, and Sieber V. A Modular Toolkit for Generating *Pichia pastoris* Secretion Libraries. ACS Synth Biol. 2017; 6:1016–25. DOI: 10.1021/acssynbio.6b00337
147. Ogulur I et al. Type 2 Immunity in Allergic Diseases. Cell Mol Immunol. 2025; 22:211–42. DOI: 10.1038/s41423-025-01261-2
148. Okada H, Kuhn C, Feillet H, and Bach JF. The 'hygiene Hypothesis' for Autoimmune and Allergic Diseases: An Update. Clin Exp Immunol. 2010; 160:1–9. DOI: 10.1111/j.1365-2249.2010.04139.x. PMID: 20415844
149. Okayama Y et al. Roles of Omalizumab in Various Allergic Diseases. Allergol Int. 2020; 69:167–77. DOI: 10.1016/j.alit.2020.01.004. PMID: 32067933
150. Onyimba F, Crowe SE, Johnson S, and Leung J. Food Allergies and Intolerances: A Clinical Approach to the Diagnosis and Management of Adverse Reactions to Food. Clin Gastroenterol Hepatol. 2021; 19:2230–2240.e1. DOI: 10.1016/j.cgh.2021.01.025. PMID: 33493695
151. Osguthorpe JD. *In vitro* Allergy Testing. Int Forum Allergy Rhinol. 2014; 4. DOI: 10.1002/alr.21384
152. Owen-Woods C et al. Local Microvascular Leakage Promotes Trafficking of Activated Neutrophils to Remote Organs. J Clin Invest. 2020; 130:2301–18. DOI: 10.1172/JCI133661. PMID: 31971917
153. Pandjaitan B et al. *Escherichia coli* Expression and Purification of Recombinant Dog Albumin, a Cross-Reactive Animal Allergen. J Allergy Clin Immunol. 2000; 105(2 II):279–85. DOI: 10.1016/s0091-6749(00)90077-0. PMID: 10669848
154. Papia F, Bellia C, and Uasuf CG. Tropomyosin: A Panallergen That Causes a Worldwide Allergic Problem. Allergy Asthma Proc. 2021; 42:E145–E151. DOI: 10.2500/aap.2021.42.210057. PMID: 34474717
155. Paul WE and Zhu J. How Are TH2-type Immune Responses Initiated and Amplified? Nat Rev Immunol. 2010; 10:225–35. DOI: 10.1038/nri2735. PMID: 20336151
156. Pavón-Romero GF et al. Allergen Immunotherapy: Current and Future Trends. Cells. 2022; 11:212. DOI: 10.3390/cells11020212
157. Pawankar R et al., eds. World Allergy Organization (WAO) White Book on Allergy. Milwaukee, Wisconsin: World Allergy Organization (WAO), 2013. 1-248
158. Pechsrichuang P and Jacquet A. Molecular Approaches to Allergen-Specific Immunotherapy: Are We so Far from Clinical Implementation? Clin Exp Allergy. 2020; 50:543–57. DOI: 10.1111/cea.13588. PMID: 32078207

159. Pejler G, Åbrink M, Ringvall M, and Wernersson S. Mast Cell Proteases. Adv Immunol. 2007; 95:167–255. DOI: 10.1016/S0065-2776(07)95006-3. PMID: 17869614
160. Perez-Riverol A, Justo-Jacomini DL, de Lima Zollner R, and Brochetto-Braga MR. Facing Hymenoptera Venom Allergy: From Natural to Recombinant Allergens. Toxins. 2015; 7:2551–70. DOI: 10.3390/toxins7072551. PMID: 26184309
161. Peters K and Peters M. The Role of Lectin Receptors and Their Ligands in Controlling Allergic Inflammation. Front Immunol. 2021; 12. DOI: 10.3389/fimmu.2021.635411. PMID: 33995354
162. Platts-Mills TA et al. Carbohydrate Epitopes Currently Recognized as Targets for IgE Antibodies. Allergy. 2021; 76:2383–94. DOI: 10.1111/all.14802
163. Polovic N et al. Dog Saliva - an Important Source of Dog Allergens. Allergy. 2013; 68:585–92. DOI: 10.1111/all.12130. PMID: 23464525
164. Poole JA et al. Organic Dust Exposure Alters Monocyte-Derived Dendritic Cell Differentiation and Maturation. Am J Physiol Lung Cell Mol Physiol. 2009; 297:767–76. DOI: 10.1152/ajplung.00107.2009.-Organic
165. Prabha L et al. Identification of the Dipeptidyl Aminopeptidase Responsible for N-terminal Clipping of Recombinant Exendin-4 Precursor Expressed in *Pichia pastoris*. Protein Expr Purif. 2009; 64:155–61. DOI: 10.1016/j.pep.2008.10.021
166. Pyrhönen K, Näyhä S, and Läärä E. Dog and Cat Exposure and Respective Pet Allergy in Early Childhood. Pediatr Allergy Immunol. 2015; 26:247–55. DOI: 10.1111/pai.12369. PMID: 25735463
167. Qin L, Tang LF, Cheng L, and Wang HY. The Clinical Significance of Allergen-Specific IgG4 in Allergic Diseases. Front Immunol. 2022; 13:1032909. DOI: 10.3389/fimmu.2022.1032909
168. Rapp E and Reichl U, eds. Advances in Glycobiotechnology. Advances in Biochemical Engineering/Biotechnology. Cham: Springer International Publishing, 2021. DOI: 10.1007/978-3-030-69590-3
169. Reily C, Stewart TJ, Renfrow MB, and Novak J. Glycosylation in Health and Disease. Nat Rev Nephrol. 2019; 15:346–66. DOI: 10.1038/s41581-019-0129-4. PMID: 30858582
170. Renz H et al. Food Allergy. Nat Rev Dis Primers. 2018; 4. DOI: 10.1038/nrdp.2017.98. PMID: 29300005
171. Roberts SA et al. Ligand-Induced Heme Ruffling and Bent NO Geometry in Ultra-High-Resolution Structures of Nitrophorin 4. Biochemistry. 2001; 40:11327–37. DOI: 10.1021/bi0109257
172. Roesner LM et al. Patients with Atopic Dermatitis Sensitized to Pet Dander Mount IgE and T-cell Responses to Mammalian Cystatins, Including the Human Self-Protein. J Investig Allergol Clin Immunol. 2022; 32:383–92. DOI: 10.18176/jiaci.0737. PMID: 34489229
173. Rönnberg E, Melo FR, and Pejler G. Mast Cell Proteoglycans. J Histochem Cytochem. 2012; 60:950–62. DOI: 10.1369/0022155412458927. PMID: 22899859
174. Rothenberg ME. Eosinophilic Gastrointestinal Disorders (EGID). J Allergy Clin Immunol. 2004; 113:11–28. DOI: 10.1016/j.jaci.2003.10.047. PMID: 14713902
175. Rouvinen J et al. Transient Dimers of Allergens. PLoS ONE. 2010; 5. DOI: 10.1371/journal.pone.0009037. PMID: 20140203
176. Ruethers T et al. Seafood Allergy: A Comprehensive Review of Fish and Shellfish Allergens. Mol Immunol. 2018; 100:28–57. DOI: 10.1016/j.molimm.2018.04.008

177. Salamon P et al. Human Mast Cells Release Interleukin-8 and Induce Neutrophil Chemotaxis on Contact with Activated T Cells. Allergy. 2005; 60:1316–9. DOI: 10.1111/j.1398-9995.2005.00886.x. PMID: 16135000
178. Salo PM et al. Prevalence of Allergic Sensitization in the United States: Results from the National Health and Nutrition Examination Survey (NHANES) 2005-2006. J Allergy Clin Immunol. 2014; 134:350–9. DOI: 10.1016/j.jaci.2013.12.1071. PMID: 24522093
179. Sánchez-Borges M et al. The Importance of Allergic Disease in Public Health: An iCAALL Statement. World Allergy Organ J. 2018; 11:10–12. DOI: 10.1186/s40413-018-0187-2
180. Sato T. Lectin-Probed Western Blot Analysis. *Lectins: Methods and Protocols, Methods in Molecular Biology*. Ed. by Hirabayashi J. Vol. 1200. 1992. New York: Springer Science+Business Media, 2014 :93–100. DOI: 10.1007/978-1-4939-1292-6
181. Schiener M et al. Application of Recombinant Antigen 5 Allergens from Seven Allergy-Relevant Hymenoptera Species in Diagnostics. Allergy. 2017; 72:98–108. DOI: 10.1111/all.13000. PMID: 27496543
182. Schmidt M and Hoffman DR. Expression Systems for Production of Recombinant Allergens. Int Arch Allergy Immunol. 2002; 128:264–70. DOI: 10.1159/000063865. PMID: 12218364
183. Schneider CA, Rasband WS, and Eliceiri KW. NIH Image to ImageJ: 25 Years of Image Analysis. Nat Met. 2012; 9:671–5. DOI: 10.1038/nmeth.2089. PMID: 22930834
184. Schoos AMM, Nwaru BI, and Borres MP. Component-Resolved Diagnostics in Pet Allergy: Current Perspectives and Future Directions. J Allergy Clin Immunol. 2021; 147:1164–73. DOI: 10.1016/j.jaci.2020.12.640. PMID: 33444632
185. Shi K et al. Epithelial Cell Membrane Perforation Induces Allergic Airway Inflammation. Nature. 2025. DOI: 10.1038/s41586-025-09331-1
186. Shin YH et al. Global, Regional, and National Burden of Allergic Disorders and Their Risk Factors in 204 Countries and Territories, from 1990 to 2019: A Systematic Analysis for the Global Burden of Disease Study 2019. Allergy. 2023; 78:2232–54. DOI: 10.1111/all.15807
187. Shiomi K et al. Sarcoplasmic Calcium-Binding Protein: Identification as a New Allergen of the Black Tiger Shrimp *Penaeus monodon*. Int Arch Allergy Immunol. 2008; 146:91–8. DOI: 10.1159/000113512. PMID: 18204275
188. Sicherer SH et al. Genetics of Peanut Allergy: A Twin Study. J Allergy Clin Immunol. 2000; 106(1 I):53–6. DOI: 10.1067/mai.2000.108105. PMID: 10887305
189. Singh MB and Bhalla PL. Recombinant Expression Systems for Allergen Vaccines. Inflamm Allergy Drug Targets. 2006; 5:53–9. DOI: 10.2174/187152806775269312. PMID: 16613564
190. Sližiene A et al. Cross-Reactive Monoclonal Antibodies against Fish Parvalbumins as a Tool for Studying Antigenic Similarity of Different Parvalbumins and Analysis of Fish Extracts. Mol Immunol. 2023; 154:80–95. DOI: 10.1016/j.molimm.2023.01.001
191. Sližienė A et al. Monoclonal Antibodies against the Newly Identified Allergen β-Enolase from Common Carp (*Cyprinus carpio*). Food Agric Immunol. 2022; 33:129–49. DOI: 10.1080/09540105.2022.2028741
192. Smith ET et al. High Yield Expression in *Pichia pastoris* of Human Neutrophil Elastase Fused to Cytochrome B5. Protein Expr Purif. 2023; 206:106255. DOI: 10.1016/j.pep.2023.106255
193. Sogo A, Cruz MJ, Amengual MJ, and Muñoz X. Identification of Pen m 4 as a Potential Cause of Occupational Asthma to Gammarus Shrimp. Clin Transl Allergy. 2018; 8:18–20. DOI: 10.1186/s13601-018-0232-9

194. Sokol CL, Barton GM, Farr AG, and Medzhitov R. A Mechanism for the Initiation of Allergen-Induced T Helper Type 2 Responses. Nat Immunol. 2008; 9:310–8. DOI: 10.1038/ni1558. PMID: 18300366
195. Soon JM and Abdul Wahab IR. Global Food Recalls and Alerts Associated with Labelling Errors and Its Contributory Factors. Trends Food Sci Technol. 2021; 118:791–8. DOI: 10.1016/j.tifs.2021.11.001
196. Spergel JM. Nonimmunoglobulin E-mediated Immune Reactions to Foods. Allergy Asthma Clin Immunol. 2006; 2:78–85. DOI: 10.2310/7480.2006.00009
197. Spolidoro GCI et al. Prevalence Estimates of Eight Big Food Allergies in Europe: Updated Systematic Review and Meta-analysis. Allergy. 2023; 78:2361–417. DOI: 10.1111/all.15801
198. Stanworth DR. The Discovery of IgE. Allergy. 1993; 48:67–71. DOI: 10.1111/j.1398-9995.1993.tb00687.x. PMID: 8457034
199. Steensma DP. The Kiss of Death: A Severe Allergic Reaction to a Shellfish Induced by a Good-Night Kiss. Mayo Clin Proc. 2003; 78:221–2. DOI: 10.4065/78.2.221
200. Stefanovic N and Irvine AD. Filaggrin and Beyond. Ann Allergy Asthma Immunol. 2024; 132:187–95. DOI: 10.1016/j.anai.2023.09.009
201. Steinke JW and Borish L. Genetics of Allergic Disease. Med Clin N Am. 2006; 90:1–15. DOI: 10.1016/j.mcna.2005.08.005. PMID: 16310521
202. Stevanovic K, Sinkkonen A, Pawankar R, and Zuberbier T. Urban Greening and Pollen Allergy: Balancing Health and Environmental Sustainability. J Allergy Clin Immunol Pract. 2025; 13:275–9. DOI: 10.1016/j.jaip.2024.12.017
203. Strachan DP. Hay Fever, Hygiene, and Household Size. BMJ. 1989; 299:1259–60. DOI: 10.1136/bmj.299.6710.1259
204. Suck R et al. Purification and Immunobiochemical Characterization of Folding Variants of the Recombinant Major Wasp Allergen Ves v 5 (Antigen 5). Int Arch Allergy Immunol. 2000; 121:284–91. DOI: 10.1159/000024341. PMID: 10828718
205. Swoboda I et al. A Recombinant Hypoallergenic Parvalbumin Mutant for Immunotherapy of IgE-Mediated Fish Allergy. J Immunol. 2007; 178:6290–6. DOI: 10.4049/jimmunol.178.10.6290. PMID: 17475857
206. Takizawa H. Diesel Exhaust Particles and Their Effect on Induced Cytokine Expression in Human Bronchial Epithelial Cells. Curr Opin Allergy Clin Immunol. 2004; 4:355–9. DOI: 10.1097/00130832-200410000-00005
207. Tamai K et al. Association between Cesarean Delivery and Childhood Allergic Diseases in a Longitudinal Population-Based Birth Cohort from Japan. Sci Rep. 2025; 15:19206. DOI: 10.1038/s41598-025-03703-3
208. Tanno LK, Chalmers R, et al. Global Implementation of the World Health Organization's International Classification of Diseases (ICD)-11: The Allergic and Hypersensitivity Conditions Model. Allergy. 2020; 75:2206–18. DOI: 10.1111/all.14468
209. Tanno LK and Demoly P. Allergy in the World Health Organization's International Classification of Diseases (ICD)-11. Pediatr Allergy Immunol. 2022; 33:5–7. DOI: 10.1111/pai.13616. PMID: 35080297
210. Taylor PE and Jonsson H. Thunderstorm Asthma. Curr Allergy Asthma Rep. 2004; 4:409–13. DOI: 10.1007/s11882-004-0092-3
211. Tegoni M et al. Mammalian Odorant Binding Proteins. Biochim Biophys Acta Biomembr. 2000; 1482:229–40. DOI: 10.1016/s0167-4838(00)00167-9

212. Thery C and Amigorena S. The Cell Biology of Antigen Presentation in Dendritic Cells. Curr Opin Immunol. 2001; 13:45–51. DOI: 10.1016/s0952-7915(00)00180-1
213. Tomura S, Ishizaki S, Nagashima Y, and Shiomi K. Reduction in the IgE Reactivity of Pacific Mackerel Parvalbumin by Mutations at Ca^{2+}-Binding Sites. Fish Sci. 2008; 74:411–7. DOI: 10.1111/j.1444-2906.2008.01538.x
214. Tsakok T et al. Does Atopic Dermatitis Cause Food Allergy? A Systematic Review. J Allergy Clin Immunol. 2016; 137:1071–8. DOI: 10.1016/j.jaci.2015.10.049
215. Uriarte SA and Sastre J. Clinical Relevance of Molecular Diagnosis in Pet Allergy. Allergy. 2016; 71:1066–8. DOI: 10.1111/all.12917. PMID: 27108666
216. Valenta R, Hochwallner H, Linhart B, and Pahr S. Food Allergies: The Basics. Gastroenterology. 2015; 148:1120–1131.e4. DOI: 10.1053/j.gastro.2015.02.006. PMID: 25680669
217. Valenta R, Karaulov A, et al. Allergen Extracts for in vivo Diagnosis and Treatment of Allergy: Is There a Future? J Allergy Clin Immunol. 2018; 6:1845–1855.e2. DOI: 10.1016/j.jaip.2018.08.032. PMID: 30297269
218. Von Ehrenstein OS et al. Reduced Risk of Hay Fever and Asthma among Children of Farmers. Clin Exp Allergy. 2000; 30:187–93. DOI: 10.1046/j.1365-2222.2000.00801.x. PMID: 10651770
219. Vredegoor DW et al. Can f 1 Levels in Hair and Homes of Different Dog Breeds: Lack of Evidence to Describe Any Dog Breed as Hypoallergenic. J Allergy Clin Immunol. 2012; 130:904–909.e7. DOI: 10.1016/j.jaci.2012.05.013
220. Walker AK et al. N-Linked Glycosylation Modulates Dimerization of Protein Disulfide Isomerase Family A Member 2 (PDIA2). FEBS J. 2013; 280:233–43. DOI: 10.1111/febs.12063. PMID: 23167757
221. Walker S et al. Seasonal Allergic Rhinitis Is Associated with a Detrimental Effect on Examination Performance in United Kingdom Teenagers: Case-control Study. J Allergy Clin Immunol. 2007; 120:381–7. DOI: 10.1016/j.jaci.2007.03.034. PMID: 17560637
222. Walsh J et al. Differentiating Milk Allergy (IgE and Non-IgE Mediated) from Lactose Intolerance: Understanding the Underlying Mechanisms and Presentations. Br J Gen Pract. 2016; 66:e609–e611. DOI: 10.3399/bjgp16X686521. PMID: 27481986
223. Wang YJ et al. *Canis familiaris* Allergen Can f 6: Expression, Purification and Analysis of B-cell Epitopes in Chinese Dog Allergic Children. Oncotarget. 2017; 8:90796–807. DOI: 10.18632/oncotarget.21822
224. Wang W. Lyophilization and Development of Solid Protein Pharmaceuticals. Vol. 203. 1–2. 2000. 1-60. DOI: 10.1016/S0378-5173(00)00423-3. PMID: 10967427
225. White AJ et al. Characterization of Sarcoplasmic Calcium Binding Protein (SCP) Variants from Freshwater Crayfish *Procambarus clarkii*. Comp Biochem Physiol B Biochem Mol Biol. 2011; 160:8–14. DOI: 10.1016/j.cbpb.2011.04.003. PMID: 21530674
226. Wintersand A et al. Allergens in Dog Extracts: Implication for Diagnosis and Treatment. Allergy. 2019; 74:1472–9. DOI: 10.1111/all.13785. PMID: 30888707
227. Wnuk W and Jauregui-Adell J. Polymorphism in High-Affinity Calcium-Binding Proteins from Crustacean Sarcoplasm. Eur J Biochem. 1983; 131:177–82. DOI: 10.1111/j.1432-1033.1983.tb07246.x. PMID: 6832140
228. Worm M et al. First European Data from the Network of Severe Allergic Reactions (NORA). Allergy. 2014; 69:1397–404. DOI: 10.1111/all.12475

229. Xu CJ et al. Shared DNA Methylation Signatures in Childhood Allergy: The MeDALL Study. J Allergy Clin Immunol. 2021; 147:1031–40. DOI: 10.1016/j.jaci.2020.11.044. PMID: 33338541
230. Yamamoto K et al. Crystal Structure of the Dog Allergen Can f 6 and Structure-Based Implications of Its Cross-Reactivity with the Cat Allergen Fel d 4. Sci Rep. 2019; 9. DOI: 10.1038/s41598-018-38134-w. PMID: 30728436
231. Yemaneberhan H et al. Prevalence of Wheeze and Asthma and Relation to Atopy in Urban and Rural Ethiopia. Lancet. 1997; 350:85–90. DOI: 10.1016/S0140-6736(97)01151-3. PMID: 9228959
232. Yu W, Freeland DM, and Nadeau KC. Food Allergy: Immune Mechanisms, Diagnosis and Immunotherapy. Nat Rev Immunol. 2016; 16:751–65. DOI: 10.1038/nri.2016.111. PMID: 27795547
233. Zemelka-Wiacek M et al. Hot Topics in Allergen Immunotherapy, 2023: Current Status and Future Perspective. Allergy. 2024; 79:823–42. DOI: 10.1111/all.15945
234. Zeng J et al. Recent Insights in Cow's Milk Protein Allergy: Clinical Relevance, Allergen Features, and Influences of Food Processing. Trends Food Sci Technol. 2025; 156:104830. DOI: 10.1016/j.tifs.2024.104830
235. Zhang Q et al. Dietary Advanced Glycation End-Products Promote Food Allergy by Disrupting Intestinal Barrier and Enhancing Th2 Immunity. Nat Commun. 2025; 16:4960. DOI: 10.1038/s41467-025-60235-0
236. Zhang Y, Matsuo H, and Morita E. Cross-Reactivity among Shrimp, Crab and Scallops in a Patient with a Seafood Allergy. J Dermatol. 2006; 33:174–7. DOI: 10.1111/j.1346-8138.2006.00040.x. PMID: 16620221
237. Zhang Y, Liu R, and Wu X. The Proteolytic Systems and Heterologous Proteins Degradation in the Methylotrophic Yeast *Pichia pastoris*. Ann Microbiol. 2007; 57:553–60. DOI: 10.1007/BF03175354
238. Zhang Z et al. Seafood Allergy: Allergen, Epitope Mapping and Immunotherapy Strategy. Crit Rev Food Sci Nutr. 2023; 63:1314–38. DOI: 10.1080/10408398.2023.2181755
239. Zheng L et al. A Fluorescent Screening Method for Optimization of Conotoxin Expression in *Pichia pastoris*. Biotech and App Biochem. 2022; 69:1611–21. DOI: 10.1002/bab.2231
240. Zimmer J et al. The History, Present and Future of Allergen Standardization in the United States and Europe. Front Immunol. 2021; 12. DOI: 10.3389/fimmu.2021.725831. PMID: 34594335
241. Ziogiene D et al. Mutations of Kluyveromyces lactis Dolichol Kinase Enhances Secretion of Recombinant Proteins. FEMS Yeast Res. 2019; 19. DOI: 10.1093/femsyr/foz024. PMID: 30865773
242. Zuberbier T et al. Economic Burden of Inadequate Management of Allergic Diseases in the European Union: A GA2LEN Review. Allergy. 2014; 69:1275–9. DOI: 10.1111/all.12470. PMID: 24965386

SANTRAUKA

SANTRUMPOS

6xHis	Šešių histidinų inkarinė seka
ade2	Genas, koduojantis fosforibozil-aminoimidazolo karboksilazę; trūkstant aplinkoje adenino yra kaupiamas raudonas pigmentas
AEX	Anijonų mainų chromatografija (angl. *anion exchange chromatography*)
Ag	Antigenas
AIT	Alergenams specifinė imunoterapija
ANOVA	Vienfaktorinė blokuotų duomenų dispersinė analizė (angl. *analysis of variance*)
BGMY	Mielių augimo buferinė terpė, praturtinta gliceroliu (angl. *buffered glycerol-complex medium*)
BMMY	Mielių augimo buferinė terpė, praturtinta metanoliu (angl. *buffered methanol-complex medium*)
CCD	Kryžmiškai reaguojančios angliavandenių determinantės (angl. *cross-reactive carbohydrate determinants*)
CEX	Katijonų mainų chromatografija (angl. *cation exchange chromatography*)
EDTA	Etilendiamintetraacto rūgštis (angl. *ethylenediaminetetraacetic acid*)
HRP	Krienų peroksidazė (angl. *horse radish peroxidase*)
IFA	Imunofermentinė analizė
IgE	Imunoglobulinas E
IPTG	Izopropil-β-D-tiogalaktopiranozidas
JSA	Jaučio serumo albuminas
K	Kamienas
Kon A	Konkanavalinas A
KT	Kambario temperatūra
LB	*Luria-Bertani* mitybinė terpė
LMM	Linijinis mišrių efektų modelis
mAk	Monokloniniai antikūnai
MBP	Maltozę surišantis baltymas (angl. *maltose-binding protein*)
MFI	Vidutinis fluorescencijos intensyvumas (angl. *mean fluorescence intensity*)

MW	Molekulinė masė (angl. *molecular weight*)
Ni-NTA	Nikelio nitrilo-triacetinė rūgštis (angl. *nickel-nitrilotriacetic acid*)
NDS-PAA	Natrio dodecilsulfato poliakrilamidinis gelis
OT	Optinis tankis
PAD	*P. pastoris* mielių augimo terpė, kurioje trūksta adenino (angl. *pichia adenine dropout*)
PAD-Ag	Agarizuota PAD terpė
PBS	Fosfatinis buferinis tirpalas (angl. *phosphate buffered saline*)
pep4	Genas, koduojantis vakuolinę aspartil proteazę (proteinazė A)
PGR	Polimerazės grandininė reakcija
prb1	Genas, koduojantis vakuolinę serino proteazę (proteinazė B)
PTM	Potransliacinės modifikacijos
PVDF	Polivinildifluoridas
sIgE	Žmonių kraujo serume esantys alergenui specifiniai IgE
SD	Standartinis nuokrypis (angl. *standard deviation*)
SKB	Sarkoplazminis kalcį surišantis baltymas
SNR	Signalo ir triukšmo santykis (angl. *signal-to-noise ratio*)
SPT	Odos dūrio testas (angl. *skin prick test*)
T_H	T pagalbinės ląstelės
VU GMC BTI	Vilniaus universiteto Gyvybės mokslų centro Biotechnologijos institutas
YPDS	Mielių augimo terpė su mielių ekstraktu, peptonu, dekstroze ir sorbitoliu (angl. *yeast extract peptone dextrose - sorbitol*)

ĮVADAS

Tarptautinė ligų klasifikacija nurodo, kad alergija yra ketvirta dažniausiai pasitaikanti lėtinė liga pasaulyje. Tai grindžiama per pastaruosius dešimtmečius stebimu nuosekliu alerginių ligų skaičiaus augimu (Shin et al. 2023; Tanno, Chalmers, et al. 2020). Šis augimas siejamas su sparčia urbanizacija, industrializacija ir gyvenimo būdo pokyčiais. Dėl padidėjusio oro užterštumo, pakitusių mitybos įpročių ir mažėjančios mikrobiologinės įvairovės, keičiasi organizmo atsakas į alergenus (Shin et al. 2023; Q Zhang et al. 2025; Kirjavainen et al. 2019).

Alerginė reakcija, tarpininkaujant IgE antikūnams, yra greita ir stipri imuninės sistemos reakcija į alergeną, būdinga 2 tipo imuniniam atsakui. Šiame procese dalyvauja alergenui specifiniai IgE (sIgE) antikūnai ir T pagalbinės ląstelės (T_H2), atpažįstančios iš alergeno kilusius antigenus (Ag). Alerginės reakcijos metu išsiskiriančios biologiškai aktyvios medžiagos, sukelia įvairius simptomus: nuo rinito ir dilgėlinės iki anafilaksinio šoko (Renz et al. 2018; Anvari et al. 2019).

Molekuliniu lygmeniu tik nedidelė dalis proteomos yra priskiriama alergenams. Alerginis potencialas nėra nulemtas vienos baltymų savybės. Jį lemia tarpusavyje sąveikaujantys veiksniai: baltymų gausa, struktūrinis stabilumas, biocheminės modifikacijos ir patekimo į organizmą kelias (Costa et al. 2022). Net ir gimininguų baltymų šeimose stebimas didelis kintamumas, rodantis, kad papildomi veiksniai – tokie kaip trimatė struktūra, epitopų prieinamumas ir šeimininko specifiniai imuniniai atsakai, gali turėti įtakos įsijautrinimui (Lukacs and Hogan 2025). IgE prisijungimą gali paveikti potransliacinės modifikacijos (PTM), ypač glikozilinimas, taip pat maisto apdorojimo metu (pvz.: kaitinimo, pasterizacijos) susidarančios modifikacijos (Zeng et al. 2025). Alergiją sukeliančių veiksnių ar savybių apibrėžimas išlieka vienu pagrindinių molekulinės alergologijos iššūkių.

Istoriškai alerginių ligų diagnozavimui ir gydymui buvo naudojami iš natūralių šaltinių išskirti alergenų ekstraktai. Tačiau jų sudėties kintamumas, nestabilumas, standartizacijos trūkumas bei užterštumas nealerginiais komponentais riboja jų patikimumą, pritaikymą ir saugumą (Zimmer et al. 2021). Kaip alternatyva alergenų ekstraktams, rekombinantinių alergenų naudojimas alergologijos srityje leidžia taikyti nuoseklesnius ir tikslesnius alergijos diagnostikos metodus. Individualių alergenų naudojimas diagnostikoje padeda nustatyti specifinius alergenus, sukeliančius alergines reakcijas. Šis *in vitro* metodas, vadinamas komponentų atskyrimo diagnostika, leidžia atskirti tikrą

įsijautrinimą nuo kryžminio reaktyvumo, tiksliau įvertinti riziką ir sudaryti individualizuotas gydymo strategijas (Sánchez-Borges et al. 2018).

Nors alergijų diagnostikoje ir gydyme pasiekta reikšminga pažanga, išlieka esminių iššūkių. Alergijas galima valdyti vengiant alergenų, alergijos simptomus slopinančiais vaistais arba taikant alergenams specifinę imunoterapiją (AIT). AIT yra terapinės vakcinacijos forma, kuria siekiama indukuoti ilgalaikę imuninę toleranciją alergenams (Durham and Shamji 2023).

Tačiau AIT gydymas yra ilgas, susijęs su tam tikra rizika pacientui ir ne visada veiksmingas; iš dalies dėl to, kad naudojami natūralios kilmės alergenų ekstraktai. Naujos kartos technologijos, įskaitant rekombinantinius alergenus, hipoalergenus, peptidų arba nukleorūgščių pagrindu sukurtas vakcinas, yra perspektyvus kelias, kuriant saugesnius ir veiksmingesnius gydymo būdus (Zemelka-Wiacek et al. 2024). Tačiau jų taikymas priklauso nuo tikslaus molekulinio apibūdinimo ir patikimų biotechnologinių sintezės sistemų, galinčių užtikrinti kliniškai reikšmingų baltymų gavimą.

Renkantis rekombinantinių alergenų raiškos sistemą, būtina įvertinti kelis kriterijus: potransliacinių modifikacijų poreikį, struktūrinį sudėtingumą reikalingą konformacinių epitopų išsaugojimui ir baltymų tirpumą. Kartu reikia atsižvelgti į praktinius apribojimus, įskaitant sintezės laiką, išeigą ir produkto saugumą terapiniam taikymui (M Liu et al. 2025). Todėl išsamus raiškos sistemų savybių supratimas yra būtinas tiek diagnostikos metodų tobulinimui, tiek naujos kartos terapinių priemonių kūrimui.

Darbo tikslas

Susintetinti ir apibūdinti rekombinantinius juodosios tigrinės krevetės *Penaeus monodon* Pen m 4, šuns *Canis familiaris* Can f 6 ir paprastosios vapsvos *Vespula vulgaris* Ves v 5 alergenus *Pichia pastoris* sistemoje bei įvertinti šios sistemos tinkamumą diagnostinės kokybės alergenams gaminti.

Darbo uždaviniai

1. Palyginti *Saccharomyces cerevisiae* ir *Pichia pastoris* raiškos sistemas rekombinantinių alergenų baltymų sintezei.

2. Susintetinti Pen m 4 ir Can f 6 alergenus, nesulietus ir sulietus su maltoze surišančiu baltymu (MBP), *E. coli* ir *P. pastoris* raiškos sistemose.

3. Įvertinti rekombinantinių alergenų antigeniškumą, naudojant alergiškų pacientų kraujo serumus.

4. Palyginti IgE reaktyvumą tarp *E. coli* ir *P. pastoris* sintetintų alergenų, siekiant įvertinti potransliacinių modifikacijų svarbą.

5. Validuoti optimizuotą *P. pastoris* rekombinantinio baltymo sintezės ir gryninimo procesą, naudojant Ves v 5 kaip modelinį vabzdžių nuodų alergeną, ir įvertinti susintetintų alergenų diagnostines savybes.

Darbo aktualumas ir mokslinis naujumas

Šiame darbe pirmą kartą sistemingai ištirta Pen m 4 ir Can f 6 alergenų sintezė *P. pastoris* raiškos sistemoje, gauti rezultatai buvo lyginami su tradicine *E. coli* sistema. Norint padidinti rekombinantinių baltymų raišką bei tirpumą, jie buvo sintetinami sujungus su maltozę surišančiu baltymu (MBP). Suliejimo strategija su MBP yra dažnai naudojama *E. coli* raiškos sistemose, tačiau jos pritaikymas *P. pastoris* sistemoje iki šiol buvo mažai tiriamas. MBP panaudojimas padidino rekombinantinių alergenų raišką ir nedarė neigiamos įtakos tolimesniems tyrimams, kadangi MBP nesąveikauja su IgE antikūnais. *P. pastoris* pagrindu sukurta rekombinantinių alergenų sintezės sistema buvo išbandyta su *Vespula vulgaris* alergenais, parodant, jog ši sistema yra tinkama mažiems ir minimaliai glikozilintiems baltymams. Sintezės mechanizmo tyrimai atskleidė proteolitinį skilimą už MBP domeno, atitinkantį literatūroje aprašytus atvejus, bei pirmą kartą nustatytą su MBP sulieto baltymo N-galo proteolitinį skilimą *P. pastoris* raiškos sistemoje. Šie atradimai išryškina kritiškai svarbius veiksnius, kuriuos būtina įvertinti konstruojant su MBP sulietus baltymus *P. pastoris* sistemoje. Kartu, nustatyta, jog *P. pastoris* mielėse susintetintas Pen m 4 alergenas formuoja ypač stabilius dimerus, kurie yra atsparūs cheminiam ar terminiam poveikiui, ir išlaiko IgE atpažinimo funkcijas. Šis glikozilinimo nulemtas konformacinis pokytis iki šiol literatūroje aprašytas tik kelis kartus (Z Li et al. 2010; AK Walker et al. 2013).

Klinikiniu požiūriu šis darbas sprendžia esminius alergijos diagnostikos trūkumus: 1) komerciniai krevečių alergijos testai remiasi tropomiozinu – panalergenu, kurio platus kryžminio reaktyvumo spektras lemia klaidingai teigiamus rezultatus; 2) šunų alergijos diagnostika netiksli dėl alergenų ekstraktų skirtingos kokybės bei neišsamaus alergenų spektro. Sintetinant ir apibūdinant Pen m 4 ir Can f 6 alergenus, kurie gali sukelti monosensibilizaciją ir apibrėžia atskiras pacientų grupes, šis darbas sukuria naujus molekulinius įrankius alergijos komponentų diagnostikai. Sukurta metodika taip pat tinka

kliniškai svarbiems alergenams, tokiems kaip paprastųjų vapsvų nuodų alergenas Ves v 5, nes metodika užtikrina aukštą išeigą ir kokybę. Tyrime nustatyti diagnostikos patikimumą ribojantys veiksniai: EDTA antikoaguliantas, dėl kalcio chelatavimo sumažinantis Pen m 4 IgE reaktyvumą 61-73% ir liofilizacija, sukelianti bakterijose susintetintų Pen m 4 ir Can f 6 baltymų degradaciją bei galimai nulemianti jų nestabilumą komerciniuose ekstraktuose. Apibendrinus, šie atradimai nustato moksliškai pagrįstus rekombinantinių alergenų sintezės protokolus ir atskleidžia kritinius analizės bei stabilumo veiksnius, kuriuos būtina kontroliuoti siekiant užtikrinti diagnostikos tikslumą.

Disertacijos teiginiai

1. *P. pastoris* PichiaPink™ raiškos sistema yra tinkama rekombinantinių alergenų sintezei su natyviai konformacijai artima antigenine struktūra, o gauti alergenai tinkami naudoti diagnostinėse sistemose.

2. Suliejimas su MBP padidina rekombinantinių alergenų išeigą *P. pastoris* sistemoje, nedarant neigiamos įtakos jų pritaikymui *in vitro* diagnostinėse sistemose.

3. *P. pastoris* raiškos sistemoje susintetintų baltymų, sulietų su MBP, N-galo proteolizė rodo anksčiau nežinomą MBP konstruktų jautrumą proteolitiniam skaldymui, tai suteikia naujų įžvalgų, kuriant suliejimo baltymus *P. pastoris* raiškos sistemoje.

4. Įprastinės procedūros, atliekamos ruošiantis diagnostiniams tyrimams, apimančios tiek EDTA antikoaguliantų naudojimą kraujo ėmimo proceso metu, tiek alergenų stabilizavimą liofilizacijos metodu, gali daryti įtaką ir reikšmingai keisti diagnostinių testų jautrumą ir tikslumą.

5. *P. pastoris* glikozilinimas rekombinantiniuose alergenuose gali sukelti struktūrinius pokyčius, kurie padidina baltymo stabilumą, tačiau kartu išlaiko diagnostikai reikalingą imunologinį antigeniškumą.

MEDŽIAGOS IR METODAI

Žmogaus kraujo serumai

Tyrimuose buvo naudojami ImmunoCAP™ arba ALEX^2® testais analizuoti žmogaus kraujo serumai. Mėginiai su sIgE koncentracijomis $\geq$0,35 kU/l ir $\geq$0,30 kU/l buvo laikomi teigiamais. Kraujo mėginiai buvo nuperkami iš PlasmaLab International (Everett, WA, JAV), o Pen m 4 tyrimams gaunami iš Imunodiagnostika Ltd. (Lietuva; etikos komiteto sutikimas Nr. 158200-17-926-430). Serumų mėginiai buvo laikomi $-20\,^{\circ}\mathrm{C}$ iki naudojimo.

Rekombinantinių baltymų sintezė E. coli raiškos sistemoje

E. coli Dh5α kamieno ląstelės buvo naudojamos vektorių dauginimui, o BL21(DE3) ir Rosetta(DE3) kamienai – rekombinantinių baltymų raiškai. Kompetentinės bakterijų ląstelės buvo paruošiamos $CaCl_2$ metodu ir transformuojamos su plazmidine DNR šiluminio šoko metodu. Po transformacijos buvo atliekama teigiamų klonų atranka kolonijų PGR metodu. Plazmidinė DNR iš *E. coli* bakterijų buvo gryninama šarminės lizės metodu arba rinkiniu pagal gamintojo rekomendacijas. Išgrynintos plazmidės buvo tikrinamos Vilniaus universiteto Gyvybės mokslų centro Biotechnologijos instituto (VU GMC BTI) DNR sekoskaitos centre.

Į sterilią skystą LB terpę, praturtintą atitinkamais antibiotikais, buvo užsėjami teigiami klonai ir auginami per naktį $37\,^{\circ}\mathrm{C}$ temperatūroje purtant. Kitą dieną, į naują mėgintuvėlį su skysta LB terpe buvo užsėjama naktinė kultūra (50 kartų skiesta) ir auginama tomis pačiomis sąlygomis kol pasiekiamas reikiamas 0,4-0,6 optinis tankis (OT). Rekombinantinių baltymų raiška indukuota su 1 mM IPTG (0,5 mM IPTG rVes v 5 baltymui) ir tęsiama pernakt (3 val. indukcija rMBP-Ves v 5 baltymui) $24\,^{\circ}\mathrm{C}$ temperatūroje ($30\,^{\circ}\mathrm{C}$ temperatūroje rVes v 5 ir rMBP-Ves v 5 baltymams). Ląstelių biomasė buvo surenkama centrifuguojant ir toliau ardoma natyviomis arba denatūruojančiomis sąlygomis. Atskirtas supernatantas arba nuskaidrintas lizatas buvo saugomi $-20\,^{\circ}\mathrm{C}$ temperatūroje.

Rekombinantinių baltymų sintezė P. pastoris raiškos sistemoje

P. pastoris PichiaPink™ raiškos sistema buvo naudojama rekombinantinių baltymų sekrecijai į mielių augimo terpę. Pilotiniuose eksperimentuose buvo naudojami skirtingų proteazių neturintys kamienai (K): K1 – *ade*2; K2 – *ade*2, *pep*4; K3 – *ade*2, *prb*1; ir K4 – *ade*2, *pep*4, *prb*1. Elektrokompetentinės *P. pastoris* ląstelės buvo paruošiamos su 1 M D-sorbitoliu ir transformuojamos su

linearizuota plazmidine DNR elektroporacijos būdu (1,5 kV, 200 Ω, 25 μF). Transformuotų mielių ląstelės buvo atgaivinamos YPDS terpėje ir teigiami klonai atrenkami PAD-Ag lėkštutėse po 3-10 dienų inkubacijos 30 °C temperatūroje. Atrinktos kolonijos buvo auginamos BMGY terpėje 30 °C temperatūroje purtant kol pasiekiamas reikiamas 5-6 OT. Tada ląstelės buvo surenkamos centrifuguojant ir rekombinantinių baltymų raiška indukuojama su BMMY terpe, praturtina metanoliu. Mielių kultūra buvo auginama 30 °C temperatūroje purtant. Kas 24 val., indukcijos terpė buvo papildoma metanoliu iki galutinės 0,5% koncentracijos. Po 3 dienų auginimo, kultūra centrifuguojama 1500 rpm 10 min; surinktas supernatantas buvo laikomas 4 °C temperatūroje iki dializės. Viduląstelinių baltymų analizei, po centrifugavimo atskirta mielių biomasė buvo ardoma mechaniškai su stiklo rutuliukais atitinkamame buferiniame tirpale. Nuskaidrinti lizatai buvo saugomi 4 °C arba −20 °C temperatūroje.

Baltymų gryninimas Ni-NTA agaroze

Histidino inkarine seka (6xHis) žymėtų rekombinantinių baltymų gryninimas buvo atliekamas su Ni-NTA agaroze natyviomis bei denatūruojančiomis sąlygomis. Lizės buferiniu tirpalu pusiausvyrinta Ni-NTA agarozė buvo inkubuojama su nuskaidrintais lizatais 4 °C temperatūroje (2 val. natyviomis arba 12-16 val. denatūruojančiomis sąlygomis). Po inkubacijos buvo atliekamas agarozės plovimas: 3 kartus fiziologiniu arba 10 kartų mažėjančio pH buferiniais tirpalais su vis didėjančia imidazolo koncentracija. Tikslinis baltymas buvo eliuojamas atitinkamai didinant imidazolo koncentraciją ir/arba mažinant pH.

Baltymų gryninimas jonų mainų chromatografija

Rekombinantiniai su MBP sulieti baltymai buvo gryninami iš surinktos *P. pastoris* augimo terpės jonų mainų chromatografijos būdu. Dr. Mindaugas Zaveckas atliko gryninimą su *Ä*KTApurifier 100 sistema 4 °C temperatūroje. Prieš gryninimą terpės buvo dializuojamos atitinkamuose buferiniuose tirpaluose ir filtruojamos naudojant 0,45 μM porų PVDF membraną su Stericup Quick Release-HV sterilia vakuumo filtravimo sistema. Katijonų (SP Sepharose Fast Flow) ir anijonų (Q Sepharose Fast Flow) sorbentai buvo naudojami rMBP-Pen m 4 ir rMBP-Can f 6 gryninimui, tačiau rMBP-Ves v 5 buvo gryninamas naudojant tik katijonų sorbentą. Baltymai buvo eliuojami linijiniu druskos gradientu (0-100%).

Baltymų gryninimas Ni-NTA chromatografija

Histidino inkarine seka pažymėti rekombinantiniai alergenai buvo gryninami iš surinktos *P. pastoris* augimo terpės afininės chromatografijos būdu. Analogiškai jonų mainų chromatografijai – gryninimas buvo atliekamas Dr. Mindaugo Zavecko, o terpės prieš gryninimą buvo dializuojamos ir filtruojamos. Gryninimas buvo atliekamas su Ni-NTA Superflow kolonėle, pusiausvyrinta su dializės buferiniu tirpalu. Baltymai buvo eliuojami linijiniu imidazolo gradientu (0-100%).

Imunoblotas

Po baltymų elektroforezės, baltymai buvo perkeliami ant aktyvintos PVDF membranos pusiau sausu pernešimo metodu (20 V, 700 mA, 40 min). Po pernašos, membrana 1 val. buvo blokuojama su Roti®Block blokavimo tirpalu. Po blokavimo ir plovimo etapų su fosfatiniu buferiniu tirpalu, papildytu 0,1% Tween-20 tirpalo (PBS-T), membrana buvo inkubuojama su pirminiais monokloniniais antikūnais (mAk) prieš 6xHis arba mAk 19C19 prieš MBP. Po inkubacijos, membrana buvo praplaunama ir inkubuojama su antriniais ožkos mAk prieš pelės IgG (H + L) konjuguotais su krienų peroksidaze (HRP). Imunodetekcijai su žmogaus kraujo serumais buvo naudojami antriniai pelės mAk prieš žmogaus IgE Fc konjuguoti su HRP. Chromogeninė reakcija buvo stebima pridėjus 4-chloro-1-napthol ir H_2O_2 arba 1-Step™ Ultra TMB-Blotting tirpalo.

Glikozilinimo ir deglikozilinimo analizės

Baltymų glikozilinimo analizė buvo atliekama naudojant manozės liekanoms specifinį Konkanavalino A (Kon A) lektiną. Tyrimas buvo atliekamas pagal pusiau sauso imunobloto metodiką, naudojant Kon A peroksidazės konjugatą, praskiestą iki 0,1 μg/ml koncentracijos PBS-T buferiniame tirpale. Chromogeninė reakcija buvo stebima pridėjus 1-Step™ Ultra TMB-Blotting tirpalo.

Deglikozilinimas buvo atliekamas naudojant Protein Deglycosylation Mix II rinkinį pagal gamintojo rekomendacijas. Išgryninti tiriami baltymai (100 μg) buvo denatūruojami 75 °C temperatūroje; tada, pridėjus deglikozilazių, buvo inkubuojami 30 min 25 °C temperatūroje ir 1 val. 37 °C temperatūroje. Gauti mėginiai buvo analizuojami baltymų elektroforezės ir imunobloto metodais.

Netiesioginė imunofermentinė analizė (IFA)

Diagnostinė 96 šulinėlių IFA plokštelė (Nunc MaxiSorp™) buvo padengiama 5 µg/ml koncentracijos antigenu, praskiestu karbonatiniame imobilizacijos buferiniame tirpale. Po 12 val. inkubacijos 4 °C temperatūroje, plokštelė buvo blokuojama laikant 2 val. KT su Roti®Block blokavimo tirpalu. Po blokavimo, plokštelė buvo praplaunama du kartus su dejonizuotu H_2O. Nusausinus plokšteles, į šulinėlius buvo įpilama po 100 µl 1:10 santykiu PBS-T buferiniame tirpale praskiestų žmogaus kraujo serumų. Po 2 val. inkubacijos KT purtant, plokštelė buvo plaunama keturis kartus su PBS-T buferiniu tirpalu, kuris buvo 1:1 santykiu praskiestas su dejonizuotu H_2O. Tada buvo įpilama po 50 µl pelės mAk prieš žmogaus IgE Fc, konjuguotų su HRP, 1:1000 santykiu praskiestų PBS-T buferiniame tirpale. Po analogiškos inkubacijos ir plovimo, plokštelės dar papildomai buvo praplaunamos du kartus su dejonizuotu H_2O ir detekcija buvo atliekama su chromogeniniu TMB tirpalu; reakcija buvo stabdoma su 3,6% H_2SO_4 rūgštimi. OT buvo matuojamas 450/620 nm spektrofotometru.

Kalcio chelatavimo poveikiui tirti serumų skiedimo buferinis tirpalas buvo papildomas 5 mM EDTA.

Teigiamo atsako ribinės vertės buvo apskaičiuojamos kaip neigiamų kontrolių vidurkis prie kurio pridėti trys standartiniai nuokrypiai (SD). Kontroliniai mėginiai: rekombinantinis MBP (*E. coli* kilmės), *P. monodon* bei *C. familiaris* alergenų ekstraktai ir HRP fermentas.

Atvirkštinės fazės baltymų mikrogardelių tyrimas

Atvirkštinės fazės baltymų mikrogardelių eksperimentai buvo atliekami Laimio Silimavičiaus įmonėje „Imunodiagnostika“ (Vilnius, Lietuva); duomenų analizė buvo atliekama Jutos Dvareckienės. Rekombinantinių baltymų koncentracija buvo standartizuojama iki 1 mg/ml, mėginiai buvo atspausdinami serijiniais skiedimais (nuo 2 iki 32 kartų) ant 2D-epoksidinių stiklų, naudojant sciFLEXARRAYER SX spausdintuvą. Alergenai buvo spausdinami 420 pl tūrio lašeliais su trimis pakartojimais, išlaikant 320 µm intervalus tarp taškų. Plokštelės kontrolei buvo naudojami Streptavidin-Cy5 žymėti taškai. Kontroliniai komponentai: natūralūs alergenų ekstraktai (0,2 µg/ml), rekombinantiniai alergenai (0,2 µg/ml), peroksidazė (0,05 µg/ml), žmogaus serumo albuminas (0,1 µg/ml) bei MBP (0,1 µg/ml).

Paruoštos mikrogardelės buvo laikomos 4 °C temperatūroje per naktį. Kitą dieną, mikrogardelės buvo džiovinamos 15 min 37 °C temperatūroje, tada blokuojamos 30 min su 2% jaučio serumo albuminu (JSA) PBS-T buferiniame tirpale. Po inkubacijos, mikrogardelė buvo išpurtoma ir inkubuojama su

žmogaus serumais (1:5 skiedimo) 2 val. 37 °C temperatūroje 24-šulinėlių hibridizacijos kasetėse. Po plovimo su dejonizuotu H_2O (30 sek inkubacija) ir nusausinimo, buvo naudojami antriniai antikūnai – pelės mAk prieš žmogaus IgE, konjuguoti su Alexa Fluor 647 žymeniu. Inkubacija buvo vykdoma 1 val. KT. Fluorescencinė reakcija buvo stebima InnoScan 710 AL skeneriu. Vaizdo analizė buvo vykdoma MAPIX programine įranga, vidutinis fluorescencijos intensyvumas (MFI) buvo apskaičiuojamas iš trijų pakartojimų duomenų.

Kryžmiškai reaguojančių angliavandenių determinančių (CCD) įtakai įvertinti, serumai, prieš tyrimą su mikrogardelėmis, buvo inkubuojami (10 min, KT) su 100 μg/mL HRP, kuris sumažina CCD reaktyvumą ∼94%. Mikrogardelėse esanti HRP buvo naudojama kaip teigiama kontrolė neapdorotų serumų analizėje ir kaip neigiama kontrolė su HRP inkubuotiems serumams.

Statistinė IFA duomenų analizė

IFA duomenys buvo normalizuojami, siekiant sulyginti skirtingų plokštelių rezultatus. Iš kiekvieno mėginio OT reikšmės buvo atimamas tuščių šulinėlių vidurkis, tada gauta vertė buvo padalinama iš toje pačioje plokštelėje esančio atitinkamo kontrolinio mėginio OT vidurkio vertės (pvz., *C. familiaris* ekstrakto reakcijos su teigiamu kraujo serumų mišiniu). Normalizavus plokštelių vertes pagal plokštelės kontrolę, kiekviena reikšmė buvo padauginama iš visų plokštelių teigiamų kontrolių vidurkio, sulyginant visų plokštelių duomenis. Rezultatai buvo pateikiami kaip normalizuotas vidutinis OT $\pm$ SD, o ribinė reikšmė (cut off) buvo nustatoma kaip neigiamų kontrolių vidurkis pridėjus tris SD.

Statistinė analizė buvo atliekama Python aplinkoje (PyCharm IDE), naudojant pandas ir numpy duomenų apdorojimui, scipy.stats ir statsmodels statistiniams testams, bei matplotlib ir seaborn rezultatų vizualizacijai. Baltymų reaktyvumo skirtumai tarp rekombinantinių variantų buvo vertinami vienfaktorinės blokuotų duomenų dispersinės analizės testu, o poriniai palyginimai (post hoc) buvo atliekami taikant Tukey HSD arba Bonferroni korekciją. Papildomai, suporuoti t-testai (dviejų imčių, suporuotų vidurkių palyginimas) buvo atliekami naudojant Microsoft Excel v2504 (Build 16.0.18730.20122). Statistiniai skirtumai buvo laikomi reikšmingais, kai $p < 0{,}05$.

Statistinė mikrogardelių duomenų analizė

MFI duomenys buvo apdorojami ir normalizuojami, siekiant pašalinti technines klaidas ir užtikrinti palyginamumą tarp mikrogardelių. Zondai, pažymėti „−100FLAG“ arba turintys signalo ir triukšmo santykį (SNR) $\leqslant 2$,

buvo pašalinami iš tolimesnės analizės. Likusios MFI reikšmės buvo transformuojamos $\log_2$ masteliu, siekiant stabilizuoti dispersiją ir sumažinti dešininį pasiskirstymo šališkumą. Iš kiekvieno taško MFI vertės buvo atimamas medianinis fono signalas, gaunant normalizuotas $\log_2$ reikšmes. Kadangi eksperimento metu buvo naudojamos kelios mikrogardelės, buvo atliekamas verčių normalizavimas tarp mikrogardelių. Palyginus kontrolinių baltymų skaitines vertes, nustatyta, jog matavimai tarp mikrogardelių skiriasi 0,16 vieneto $\log_2$ skalėje, t.y., apie 12%. Todėl atitinkamai buvo koreguojami duomenys ir plokštelių vertės buvo suvienodinamos. Po šios korekcijos, kontrolinių baltymų (MBP, HRP, alergenų ekstraktai, bendras IgE, HSA ir kontroliniai baltymai) reikšmės buvo sugrupuotos ir buvo apskaičiuojamas jų vidurkis, suformuojant vieningą duomenų rinkinį.

Statistiniam vertinimui buvo taikomas linijinis mišrių efektų modelis (LMM), pagrįstas normalizuotomis $\log_2$ MFI reikšmėmis. Baltymo variantas buvo laikomas fiksuotu efektu, o mikrogardelė ir serumo donoras – atsitiktiniais efektais, siekiant įvertinti likutinį tarpindividinį ir tarpplokštelinį kintamumą. Modelis buvo įgyvendintas Python aplinkoje (PyCharm IDE), naudojant pandas ir numpy duomenų tvarkymui, scipy.stats ir statsmodels statistiniams testams bei matplotlib ir seaborn vizualizacijai. Poriniai palyginimai (post hoc) buvo koreguojami taikant Bonferroni korekciją, siekiant kontroliuoti bendrą I rūšies klaidos lygį.

CCD įtakos analizei buvo lyginamos HRP-inhibuotų ir neinhibuotų serumų suporuotos $\log_2$ MFI reikšmės kiekvienam baltymui. Shapiro-Wilk testas buvo taikomas porinių skirtumų normalumui įvertinti: jei $p \geqslant 0{,}05$, buvo taikomas suporuoto Student t-testo analizė, o nenormaliems duomenims – Wilcoxon Signed Rank testas. Gautos p-reikšmės buvo koreguotos taikant Benjamini-Hochberg metodą (klaidingų atradimų dažnis = 5%), siekiant išlaikyti klaidingų atradimų lygį. CCD inhibicijos poveikiui įvertinti buvo taikomas dviejų faktorių blokuotų duomenų dispersinės analizės modelis, kur „baltymas“ ir „HRP inhibicija“ buvo laikomi vidinių kintamųjų faktoriais, o „serumo donoras“ – subjekto kintamasis. Visuose testuose reikšmingumo riba buvo nustatoma, kai $\alpha = 0{,}05$.

REZULTATAI IR JŲ APTARIMAS

1. Rekombinantinių alergenų sintezės sistemų pasirinkimas

Pirmiausia buvo atlikta rekombinantinių alergenų sintezės sistemų analizė, išrenkant optimalią prokariotinę ir eukariotinę sistemas. Raiškai bakterijose buvo pasirinkta tradicinė *E. coli* BL21(DE3) sistema, paremta pET28a(+) ir pET28a(+)-MBP-TEV vektorių naudojimu. Sintezei bei tirpumui padidinti – alergenai sulieti su MBP baltymu, o užtikrinti efektyvų gryninimą – su 6xHis inkarine seka, kadangi MBP dažnai nesąveikauja su amilozės sorbentu (Nallamsetty and Waugh 2006; Nallamsetty, Austin, et al. 2005; Bell et al. 2013).

Raiškai mielėse pirmiausia pasirinkta *Saccharomyces cerevisiae* mielių sistema, paremta viduląsteline (pFX7 tipo vektoriai) ir sekrecine (pFG-PGK1-α tipo vektoriai) sinteze. Analogiškai prokariotiniam modeliui, MBP buvo naudotas kaip sulietinis baltymas gerinantis sintezę ir tirpumą. Tačiau, nors MBP yra dažnai naudojamas *E. coli* raiškos sistemose, jo pritaikymas mielėse iki šiol buvo mažai tiriamas (Z Li et al. 2010; Moua et al. 2016; Jiang et al. 2016; J Gao et al. 2023). Šis darbas pirmą kartą aprašo MBP strategijos taikymą rekombinantinių alergenų sintezei mielėse, siekiant ištirti, ar ši strategija galėtų padidinti išeigas ir pagerinti tinkamai susilanksčiusių baltymų sintezę.

Pilotiniai eksperimentai buvo atlikti su aštuoniais alergenais iš įvairių baltymų šeimų. Viduląstelinė ir sekrecinė heterologinių baltymų raiška *S. cerevisiae* mielių kamienuose buvo nedidelė arba neaptinkama, imunobloto analizėse buvo matoma, jog baltymai sudegradavo. *S. cerevisiae* AH22-214Δmn10 mutantinio kamieno išbandymas hiperglikozilinimo sumažinimui irgi nepasiteisino. Kadangi glikozilinimo analizės parodė, jog PTM lygis priklausė nuo pasirinktos mielių kolonijos, indukcijos temperatūros ir indukcijos trukmės. Apžvelgiant susintetintų baltymų išeigas, PTM kintamumą ir *S. cerevisiae* sekretuojamų foninių baltymų gausą kultūros terpėse, buvo nuspręsta tolimesnius eksperimentus vykdyti *Pichia pastoris* raiškos sistemoje.

Literatūroje aprašyta didelė įvairovė rekombinantinių alergenų susintetintų *P. pastoris* mielėse. Raiškai eukariotinėje sistemoje pasirinkta PichiaPink™ sistema, naudojanti pPink-HC vektorių su AOX1 promotoriaus kontroliuojamu α-faktoriaus signaliniu peptidu, skirtu sekretuojamai raiškai. Ši sistema pasižymi efektyvia sekrecija, mažu foninių baltymų kiekiu, palengvinančiu gryninimą, bei mažesniu hiperglikozilinimu.

Pasirinkus sistemas bei apibendrinus literatūros duomenis, nuspręsta toliau dirbti su baltymais, kurių sintezė mielėse iki šiol netirta – Pen m 4 ir Can f 6 – bei su vienu iš svarbiausių vabzdžių nuodų alergenu Ves v 5.

2. Rekombinantinio alergeno Pen m 4 sintezė ir imunologinis apibūdinimas

2.1 E. coli raiškos sistema: rPen m 4 ir rMBP-Pen m 4 sintezė ir apibūdinimas

Rekombinantinio Pen m 4 sintezei *E. coli* bakterijose buvo sukonstruoti raiškos vektoriai, naudojant pET28a(+) ir pET28a(+)-MBP-TEV plazmides, į kurias buvo įterpta optimizuota Pen m 4 geno DNR seka. Sukonstruotų vektorių pET28a(+)-Pen m 4 ir pET28a(+)-MBP-TEV-Pen m 4 sekoskaita buvo atlikta VU GMC BTI DNR sekoskaitos centre.

Sukurtomis plazmidėmis transformavus *E. coli* BL21(DE3) kamieną, rPen m 4 ir rMBP-Pen m 4 baltymų sintezė buvo indukuota į terpę pridėjus IPTG. Surinkta bakterijų biomasė buvo suardyta ultragarsu ir išanalizuota baltymų elektroforezės metodu. Vektoriumi pET28a(+)-Pen m 4 transformuotų bakterijų supernatanto mėginyje buvo aptikta rPen m 4 baltymą atitinkanti $\sim$25 kDa dydžio baltymo juostelė, o pET28a(+)-MBP-TEV-Pen m 4 vektoriumi – $\sim$66 kDa juostelė. Tirpūs tiksliniai baltymai buvo išgryninti su Ni-NTA agaroze ir gautos, atitinkamai, 52 mg/l ir 66 mg/l išeigos. Suliejimas su 6xHis-MBP seka padidino Pen m 4 baltymo išeigą 1,3 karto.

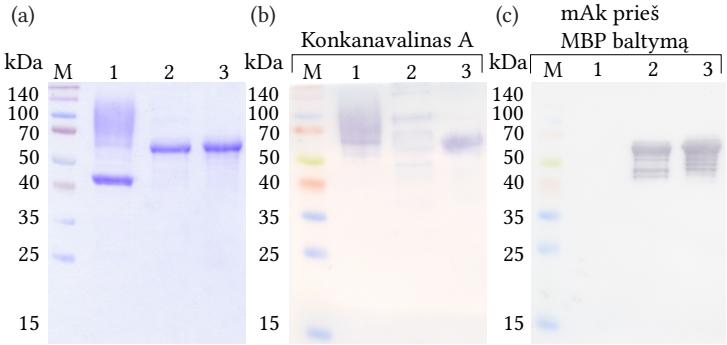
E. coli sistemoje susintetintų Pen m 4 baltymų apibūdinimas masių spektrometrijos ir imunobloto metodais parodė, jog tikslinių baltymų dydžiai, 25 ir 66 kDa, atitinka teorines molekulines mases (MW). Rekombinantinių Pen m 4 baltymų laikymo sąlygų analizė atskleidė, jog liofilizacija neigiamai veikia tiek rPen m 4, tiek rMBP-Pen m 4 stabilumą. Kitų sąlygų – $4\,^{\circ}\mathrm{C}$, $20\,^{\circ}\mathrm{C}$ ir $-20\,^{\circ}\mathrm{C}$ su 40% gliceroliu – poveikis baltymo stabilumui po mėnesio laikymo buvo panašus. Tačiau patikrinus mėginius po 24 mėn buvo galima stebėti, jog laikymas be glicerolio tiek $4\,^{\circ}\mathrm{C}$, tiek $-20\,^{\circ}\mathrm{C}$ temperatūroje sukėlė stiprią degradaciją, o laikymas $-20\,^{\circ}\mathrm{C}$ temperatūroje su 40% gliceroliu užtikrino geriausią stabilumą. Kadangi rPen m 4 numatytas naudoti kaip farmacinis produktas, šis jautrumas šaldymo ir džiovinimo stresui turi būti vertinamas ruošiant alergenų ekstraktus (W Wang 2000). Šie rezultatai gali paaiškinti, kodėl Pen m 4 molekulė neaptinkama 2 iš 5 komercinių krevečių ekstraktų, dažniausiai gaminamų iš užšaldytų krevečių (Asero, Scala, et al. 2017).

2.2 P. pastoris raiškos sistema: rPen m 4 ir rMBP-Pen m 4 sintezė ir apibūdinimas

Rekombinantinio Pen m 4 sintezei *P. pastoris* mielėse buvo sukonstruoti raiškos vektoriai, naudojant pPink-αF-HC vektorių, į kurį įterpta arba optimizuota Pen m 4 geno DNR seka arba sulietinį MBP-TEV-Pen m 4 baltymą koduojanti DNR seka. Sukonstruotų vektorių pPink-αF-Pen m 4 ir pPink-αF-MBP-TEV-Pen m 4 sekoskaita buvo atlikta VU GMC BTI DNR sekoskaitos centre.

Sukurtomis plazmidėmis transformavus PichiaPink™ sistemos mielių kamienus, pasižyminčius skirtingais proteazių aktyvumais, rPen m 4 ir rMBP-Pen m 4 baltymų sintezė indukuota į terpę pridėjus metanolio. Surinkta augimo terpė buvo sukoncentruota ir išanalizuota baltymų elektroforezės metodu. Vektoriumi pPink-αF-MBP-TEV-Pen m 4 transformuotų mielių mėginiuose buvo aptikta rMBP-Pen m 4 baltymą atitinkanti ~67 kDa juostelė. Baltymo rMBP-Pen m 4 sintezės efektyvumas buvo įvertintas ir imunobloto analizėje su mAk prieš MBP baltymą (Pav. 1 takeliai 2 ir 3). Didelės apimties sintezei pasirinktas *pep*4 geno neturintis kamienas, kuris sintetino didžiausią kiekį rMBP-TEV-Pen m 4 baltymo.

Tuo tarpu vektoriumi pPink-αF-Pen m 4 transformuotų mielių mėginiuose buvo aptiktos ~45-65 kDa baltymų juostos, neatitinkančios teorinės 23 kDa rPen m 4 molekulinės masės (Pav. 1 takelis 1). Dr. Mindaugui Zaveckui išgryninus rPen m 4 baltymą, naudojant Ni-NTA chromatografiją, gauta 17 mg/l išeiga. Mėginių su išgrynintu rPen m 4 analizė imunobloto metodu, parodė, jog su mAk prieš 6xHis inkarinę seką buvo stebimos tik 45-60 kDa baltymų juostelės ir nebuvo aptinkama juostelė, atitinkanti teorinę rPen m 4 MW. Plazmidės pPink-αF-Pen m 4 sekos patikrinimas, pakartotinės transformacijos ar skirtingos auginimo sąlygos šio rezultato nepakeitė.



Pav. 1: Išgrynintų *P. pastoris* susintetintų rPen m 4 ir rMBP-Pen m 4 baltymų analizė **(a)** NDS-PAA gelyje ir **(b, c)** imunoblote (pažymėta naudojamu mAk arba lektino pavadinimu). Takeliai: [1], rPen m 4; [2], rMBP-Pen m 4, išgrynintas naudojant AEX; [3], rMBP-Pen m 4, išgrynintas naudojant CEX. M – Spectra™ Multicolor Broad Range Protein Ladder. **(b)** Glikozilinimo analizė atlikta naudojant Kon A lektiną; **(c)** Imunobloto analizė atlikta naudojant mAk prieš MBP baltymą. Baltymai atskirti 14% SDS-PAGE geliuose.

Įsitikinti, jog tai *P. pastoris* mielėms specifinė modifikacija, rPen m 4 alergenas buvo susintetintas *Kluveromyces lactis* mielių WGI, WSS ir WSSΔpep4

kamienuose. Išanalizavus iš transformuotų *K. lactis* mielių išgryninto rPen m 4 mėginius, visų trijų kamienų gryninimo frakcijose buvo stebima rPen m 4 baltymą atitinkanti ∼23 kDa juostelė. Šie rezultatai parodė, jog modifikacija specifinė *P. pastoris* mielėms. *K. lactis* vektorių konstravimas ir mielių transformacija atlikta Dr. Danguolės Žiogienės ir Andriaus Budrulio, gryninimas atliktas Jutos Dvareckienės.

Siekiant įvertinti, ar padidėjusi MW gali būti susijusi su glikozilinimu, buvo atlikta bioinformatinė analizė, deglikozilinimas su PNGaze F ir analizė su Kon A lektinu. Pen m 4 aminorūgščių sekoje nustatytos dvi O-glikozilinimo ir viena N-glikozilinimo vieta (Hamby and Hirst 2008). Kon A analizė ir pokyčiai po PNGazės F veikimo parodė, jog rPen m 4 yra N-glikozilintas. Po deglikozilinimo likusi sąveika su manozę surišančiu Kon A lektinu rodo, jog ne visos angliavandenių modifikacijos buvo pašalintos. Šis signalas gali būti susijęs su nepilnu deglikozilinimu, O-glikozilinimu arba glikacija (de Oliveira et al. 2016). Imunoblote su antikūnais prieš 6xHis žymenį deglikozilintame mėginyje buvo stebimos keturios juostos – 55, 47, 35 ir 23 kDa. Mažiausia juosta atitiko apskaičiuotą monomerinio rPen m 4 MW.

Šie rezultatai rodo, jog aptinkamas platus 45-65 kDa rPen m 4 signalas bent iš dalies paaiškinamas heterogenišku glikozilinimu, atliekamu *P. pastoris* raiškos sistemoje. Tačiau glikozilinimas nepaaiškina po deglikozilinimo išlikusių didesnės MW baltymų, todėl buvo daroma prielaida, jog vyrauja papildomas veiksnys. *P. monodon* Pen m 4 alergenas priklauso sarkoplazminių kalcį surišančių baltymų (SKB) šeimai, kurios nariai natūraliai aptinkami kaip dimerai (Ayuso et al. 2009). rPen m 4 masių spektrometrijos analizėje pagrindinis signalas buvo nustatytas ties ∼47 kDa. Ši masė yra artima apskaičiuotai rPen m 4 homodimero masei ir yra apie 1,3 kDa didesnė už teorinę vertę. Šis neatitikimas galėjo būti stebimas dėl deglikozilinimo metu pilnai nepašalintų angliavandenių masės.

Literatūros duomenimis, SKB dimerai turi disocijuoti urėjos ar SDS turinčioje aplinkoje (Wnuk and Jauregui-Adell 1983; Ayuso et al. 2009). Tačiau, šiame tyrime atlikti eksperimentai redukuojančiose sąlygose (4-6 M urea, 4-6 M GuaHCl, 50-250 mM DTT, 100 mM EDTA ir kaiti-nimas 30 min 100 °C temperatūroje) nesuardė *P. pastoris* susintetinto rPen m 4. Todėl padaryta išvada, jog *P. pastoris* atliekamos PTM formuoja itin stabilų dimerą; tokie glikozilinimo moduliuojami stabilūs dimerai aprašyti Z Li et al. 2010 ir AK Walker et al. 2013 darbuose.

Glikacija taip pat galėjo prisidėti prie stebėto NDS-PAA gelyje baltymo

profilio, nes Pen m 4 sekoje yra santykinai daug lizino liekanų (7,2%), o baltymų šleifai ir likutinė sąveika su Kon A po PNGazės F apdorojimo gali rodyti kovalentiškai prijungtus angliavandenius (Lima and Baynes 2013; de Oliveira et al. 2016). Vis dėlto tiksli PTM, atsakinga už didesnės molekulinės masės rPen m 4 formas, nebuvo nustatyta.

Nepriklausomai nuo stebimo heterogeniškumo, imunoblotuose rPen m 4 baltymas parodė specifinį IgE prisijungimą su *P. monodon* alergiškų žmonių kraujo serumais ir nerodė kryžminio reaktyvumo su kitais antikūnais. Apibendrinant, šie rezultatai rodo, kad *P. pastoris* susintetintas rPen m 4 yra stabilus, IgE reaktyvus rekombinantinis alergenas, tačiau naudojant šį baltymą diagnostiniuose tyrimuose reikėtų įtraukti glikozilinimo kontrolę.

Tuo tarpu, rekombinantinis baltymas rMBP-Pen m 4 buvo išgrynintas Dr. Mindaugo Zavecko naudojant katijonų mainų (CEX) ir anijonų mainų (AEX) chromatografiją. Šie metodai pasirinkti norint ištirti galimą gryninimo strategijos įtaką IgE sąveikai. AEX ir CEX metodais išgryninto rMBP-Pen m 4 išeigos buvo, atitinkamai, 39,5 mg/l ir 1,5 mg/l (Pav. 1 takeliai 2 ir 3). Skirtingi gryninimo metodai lėmė ne tik išeigos skirtumus, bet ir rMBP-Pen m 4 glikozilinimo profilių variaciją. Masių spektrometrijos analizė parodė, jog išgrynintų rMBP-Pen m 4 baltymų dydis buvo 2,28-2,29 kDa mažesnis nei teorinė MW. Imunoblotuose su mAk prieš 6xHis inkarinę seką signalas nebuvo aptiktas, tačiau naudojant mAk prieš MBP baltymą buvo stebimos ryškios tikslinių baltymų juostelės. Šie rezultatai rodo, jog *P. pastoris* susintetintuose rMBP-Pen m 4 baltymuose buvo prarasta N-galinė 6xHis žymenį turinti sritis. Nepaisant šio proteolitinio apdorojimo, rMBP-Pen m 4 specifiškai sąveikavo su *P. monodon* alergiškų žmonių kraujo serumais, o kryžminių reakcijų nebuvo stebėta.

Imunobloto analizėje su mAk prieš MBP taip pat buvo stebimas ∼45 kDa fragmentas, pagal masių spektrometrijos analizę apie 42 kDa, kuris atitinka nuskelto MBP dydį. Tai rodo papildomą skilimą už MBP domeno esančiose sulieto baltymo srityse. Remiantis literatūros analize, *P. pastoris* mielėse yra aprašytas su MBP susijusios proteazės veikimas, kuris, kaip manoma, atpažįsta ne specifinę aminorūgščių seką, o trimatę baltymo struktūrą. Papildomi tyrimai parodė, jog ∼45 kDa fragmento susidarymo nepanaikino nei proteazių inhibitoriai, nei skirtingų proteazių neturintys *P. pastoris* kamienai. Literatūros duomenimis, MBP jungimas prie baltymo N- ar C-galo gali pagerinti sekreciją ir išeigą, tačiau poveikis priklauso nuo tikslinio baltymo ir reikalauja empirinių tyrimų. Šiame tyrime MBP sujungimas padidino rPen m 4 išeigą 2,3 karto,

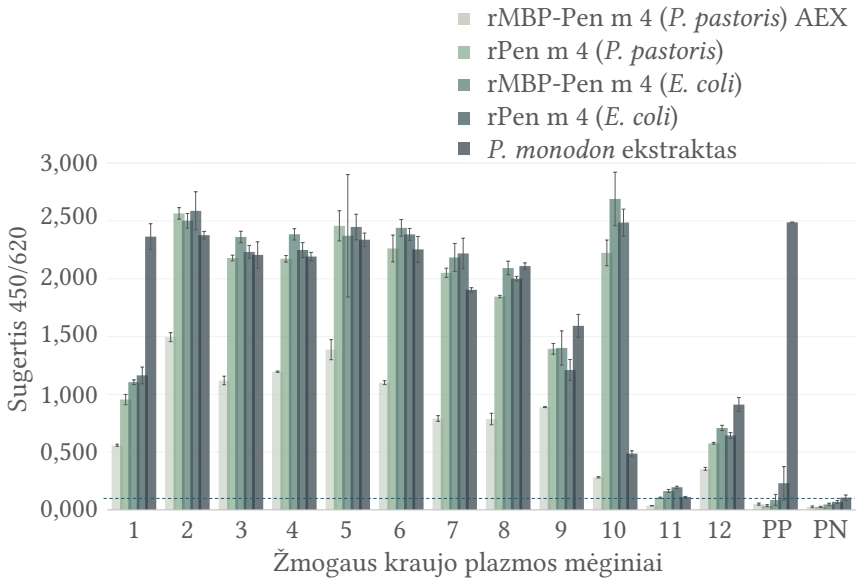
priklausomai nuo gryninimo strategijos.

2.4 rPen m 4 ir rMBP-Pen m 4 baltymų variantų imunologinis apibūdinimas

Susintetintų rekombinantinių baltymų antigeniškumas tirtas su alergiškų žmonių kraujo plazmos mėginiais. Imunologinei analizei mėginiai buvo suskirstyti į keturias grupes: (i) 12 individualių mėginių, teigiamų Pen m 4 sIgE; (ii) teigiama grupė – 8 mėginiai, teigiami *P. monodon* ekstraktui, bet neigiami Pen m 4 sIgE; (iii) neigiama grupė – 8 IgE neigiami mėginiai; (iv)∼100 individualių *P. monodon* alergijai neigiamų mėginių, tačiau teigiamų kitoms alergijoms. Tyrimas atliktas netiesiogine IFA ir atvirkštinės fazės baltymų mikrogardele.

Tiriant rekombinantinių Pen m 4 baltymų antigeniškumą, pirmiausia, buvo nustatomas gryninimo strategijos poveikis baltymo imunogeniškumui. IFA rezultatai parodė, jog *P. pastoris* susintetintas rMBP-Pen m 4, išgrynintas AEX metodu, statistiškai stipriau reagavo su sIgE nei rMBP-Pen m 4, išgrynintas CEX metodu. Šie rezultatai rodo, jog gryninimas fiziologinėse sąlygose pH (7,5) padeda išlaikyti natyvią baltymo konformaciją ir konformacinius epitopus.

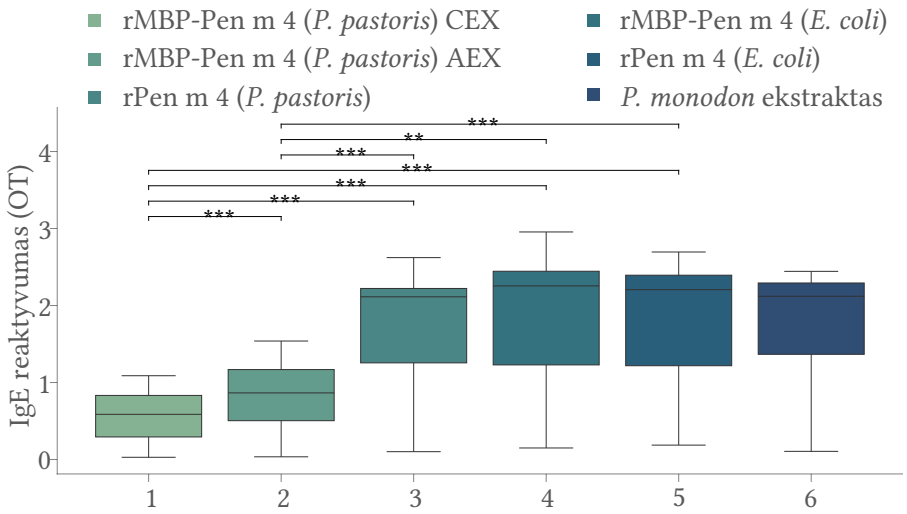
Susintetintų variantų IFA analizė parodė, jog visi tirti baltymai specifiškai sąveikauja su Pen m 4 sIgE ir nerodo kryžminio reaktyvumo (Pav. 2).



Pav. 2: rPen m 4 baltymų variantų imunoreaktyvumo analizė netiesioginėje IFA sistemoje. AEX – rMBP-Pen m 4, išgrynintas anijonų mainų chromatografijos būdu. [1-12] yra individualūs *P. monodon* alergija sergančių žmonių kraujo plazmos mėginiai, kurie yra teigiami Pen m 4 alergenui; PP – teigiama grupė (*P. monodon* sIgE teigiama, bet Pen m 4 sIgE neigiama, n = 8); PN – neigiama grupė (n = 8). Ribinė reikšmė pažymėta punktyrine linija.

Analizuojant, ar susintetintų baltymų variantų reaktyvumas su sIgE yra statistiškai skirtingas, atlikta vienfaktorinė blokuotų duomenų dispersinė analizė. Gauti ANOVA rezultatai $F(4,\ 44) = 41,28,\ p < 0,001$ patvirtina, jog Pen m 4 variantai reaguoja statistiškai skirtingai. Įdomu, jog *E. coli* sistemoje susintetinti Pen m 4 variantai stipriau sąveikavo su 11 serumu, kuris pasižymėjo žemu sIgE kiekiu, rodant šių variantų jautrumą. Tačiau *E. coli* susintetintas rPen m 4 kartu parodė ir didesnę nespecifišką sąveiką su teigiama (PP) grupe.

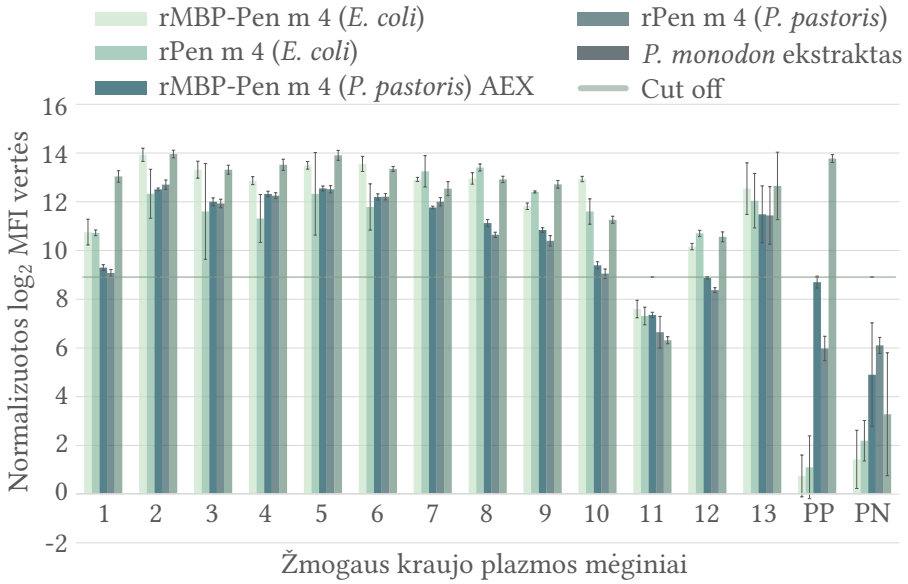
Post hoc analizė su Bonferroni korekcija atskleidė reikšmingus skirtumus tarp *P. pastoris* susinte-tintų rMBP-Pen m 4 baltymų (grynintų AEX ir CEX metodais) ir visų kitų tirtų Pen m 4 variantų (Pav. 3). *P. pastoris* susintetinti rMBP-Pen m 4 baltymai turėjo žemiausią medianą ir statistiškai silpniau reagavo su sIgE lyginant su kitais tirtais baltymais.



Pav. 3: Grafikas, vaizduojantis rPen m 4 baltymų variantų imunoreaktyvumo pasiskirstymą. [1], *P. pastoris* susintetintas rMBP-Pen m 4, išgrynintas naudojant katijonų mainų chromatografiją (CEX) ir [2] – naudojant anijonų mainų chromatografiją (AEX); [3], *P. pastoris* susintetintas rPen m 4 baltymas; [4], *E. coli* susintetintas rMBP-Pen m 4; [5], *E. coli* susintetintas rPen m 4; [6], *P. monodon* alergenų ekstraktas. Duomenys atspindi normalizuotą IgE reaktyvumą su 13 individualių *P. monodon* alergija sergančių žmonių kraujo plazmos mėginių, kurie yra teigiami Pen m 4 alergenui. Reaktyvumas pateikiamas kaip normalizuotos OT vertės, o reikšmingi skirtumai tarp baltymų pažymėti žvaigždutėmis ($** = p < 0,01, \quad *** = p < 0,001$).

Apibendrinus, nors silpna reakcija su sIgE antikūnais nėra tinkama diagnostiniams testams, tokia savybė būtų tinkama imunoterapijai, kadangi sukeliamas silpnesnis imuninis atsakas. Kiti Pen m 4 variantai reagavo panašiu stiprumu ir statistiškai neišsiskyrė.

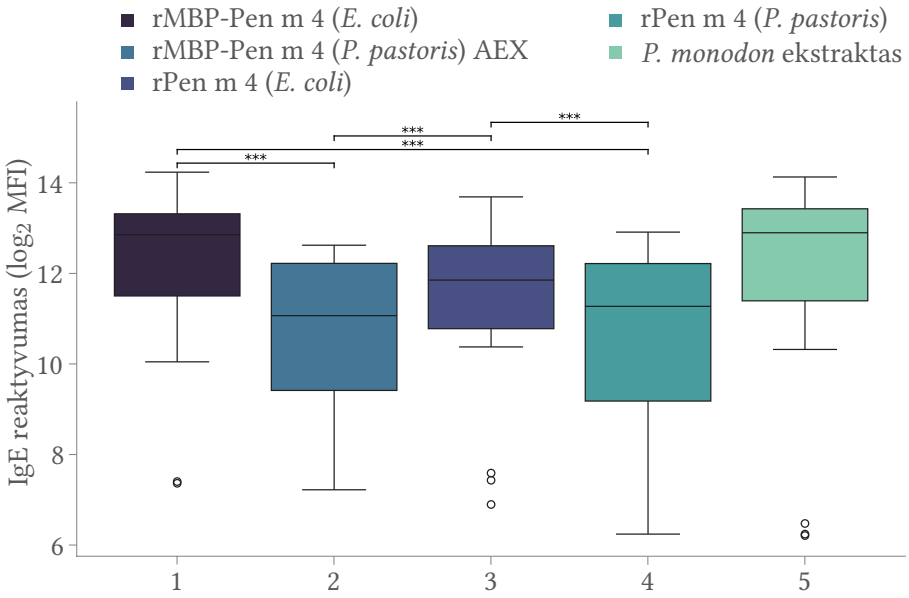
Imunologinei analizei naudojant atvirkštinės fazės baltymų mikrogardelių sistemą, gauti rezultatai patvirtino netiesioginės IFA duomenis, jog *E. coli* susintetinti variantai stipriau sąveikauja su sIgE (Pav. 4). Jautresnis detekcijos metodas parodė, jog *E. coli* susintetintas rMBP-Pen m 4 turėjo didžiausias vidutinio fluorescencijos intensyvumo (MFI) vertes. Pav. 4 matomos normalizuotos MFI vertės, t.y. gauti duomenys buvo $\log_2$ transformuoti, norint sumažinti sklaidą ir lengviau interpretuoti santykinius pokyčius



Pav. 4: rPen m 4 baltymų variantų imunoreaktyvumo analizė atvirkštinės fazės baltymų mikrogardelių sistemoje. Grafike pateikti 0,25 mg/ml koncentracijos baltymų imunoreaktyvumo analizės rezultatai. AEX – rMBP-Pen m 4, išgrynintas anijonų mainų chromatografijos būdu. [1-13] yra individualūs *P. monodon* alergija sergančių žmonių kraujo plazmos mėginiai, kurie yra teigiami Pen m 4 alergenui; PP – teigiama grupė (*P. monodon* sIgE teigiama, bet Pen m 4 sIgE neigiama, n = 8); PN – neigiama grupė (n = 8). Y-ašyje pateikiamos normalizuotos $\log_2$ MFI vertės; kiekvienas vienetas šioje skalėje atitinka dvigubą neapdoroto MFI intensyvumo pokytį. Cut off – ribinė reikšmė.

Mikrogardelės eksperimentai parodė, kad optimali Pen m 4 koncentracija padengimui yra 0,25 mg/ml: mažesnės koncentracijos mažina jautrumą, o didesnės didina klaidingai teigiamų rezultatų tikimybę. Tiriant kryžminį reaktyvumą su (iv) kraujo grupe, užfiksuotas atvejis, kai pacientas buvo jautrus tik Pen m 4, bet nejautrus Pen m 1 ir Pen m 2 alergenams, patvirtinant Pen m 4 sukeliamos monosensibilizacijos galimybę (Sogo et al. 2018). Taip pat nustatyta, jog *P. pastoris* susintetinti alergenai dėl glikozilinimo gali didinti I tipo klaidų tikimybę (I tipo klaidos – klaidingai teigiami, o II tipo klaidos – klaidingai neigiami), todėl diagnostinėse plokštelėse būtina naudoti glikozilinimo kontrolę (Kozlov et al. 2025). Pav. 4 matoma, jog 66% atvejų stipriausiai reagavo *E. coli* rMBP-Pen m 4, o 33% – *E. coli* rPen m 4, tačiau Pen m 4 variantai su 11 serumu neperžengė teigiamos ribos. Tai rodo, kad nors mikrogardelės yra jautresnės, didesnis fonas gali maskuoti silpnus teigiamus signalus ir didinti II tipo klaidų tikimybę.

Linijinis mišrių efektų modelis (LMM) patvirtino, jog matomi statistiškai reikšmingi skirtumai tarp tirtų baltymo variantų (tikimybės santykis $\chi^2(3) = 88{,}78$, $p = 1,77 \times 10^{-18}$) (Pav. 5).



Pav. 5: Grafikas, vaizduojantis rPen m 4 baltymų variantų imunoreaktyvumo pasiskirstymą. [1], *E. coli* susintetintas rMBP-Pen m 4; [2], *P. pastoris* susintetintas rMBP-Pen m 4, išgrynintas naudojant anijonų mainų chromatografiją (AEX); [3], *E. coli* susintetintas rPen m 4; [4], *P. pastoris* susintetintas rPen m baltymas; [5], *P. monodon* alergenų ekstraktas. Y-ašyje pateikiamos normalizuotos $\log_2$ MFI vertės; kiekvienas vienetas šioje skalėje atitinka dvigubą neapdoroto MFI intensyvumo pokyt. Reikšmingi skirtumai tarp baltymų pažymėti žvaigždutėmis ($*** = p < 0,001$).

Post hoc Bonferroni korekcija parodė, jog matomi skirtumai tarp naudotų baltymų raiškos sistemų, tačiau ne tarp variantų tose pačiose sistemoje. Tai rodo, jog net jei *E. coli* rMBP-Pen m 4 turėjo aukščiausius MFI lygius, tas skirtumas nebuvo statistiškai reikšmingas. Tačiau apibendrinus antigeniškumą su gautomis išeigomis, rMBP-Pen m 4 (*E. coli*) išlieka optimaliu kandidatu *P. monodon* alergijos diagnostikai.

2.5 CCD poveikis IgE prisijungimui prie rPen m 4

Šiame darbe mielėse susintetinti rekombinantiniai alergenai yra glikozilinti *P. pastoris* būdingu manozilinimu, kuris skiriasi nuo klasikinių CCD struktūrų, sąveikaujančių su IgE antikūnais. Todėl norint ištirti manozilinimo poveikį rPen m 4 alergenų sąveikoje su sIgE, buvo atliktas eksperimentas su krienų

peroksidaze (HRP). Eksperimente HRP buvo naudota kaip kontrolinis žymuo netipiniam glikozilinimui nustatyti. Normalizuotos $\log_2$ MFI reikšmės iš HRP-inhibuotų ir neinhibuotų serumų buvo lyginamos taikant Shapiro-Wilk testą; jei $p \geqslant 0,05$ – naudotas suporuotas Student t-testas, kitu atveju – Wilcoxon testas. p reikšmės buvo koreguotos Benjamini-Hochberg metodu (FDR = 5%), o bendram inhibicijos poveikiui vertinti taikytas dviejų faktorių pakartotinių matavimų dispersinės analizės modelis (α = 0,05). Duomenų analizėje matoma, jog HRP-inhibuotose mėginiuose statistiškai reikšmingas skirtumas matomas tik HRP kontrolėje (Subsection 3.2.8 Fig.3.17). Tai rodo, jog manozilinimas, bent šiuo atveju, nedaro statistiškai reikšmingo skirtumo gaunamoms vertėms.

Vis dėlto, apibendrinus ankščiau gautus rezultatus, kuriuose su *P. pastoris* susintetintais rPen m 4 baltymais buvo stebimos I tipo klaidos, būtina atsižvelgti į glikozilinimo tipą ir įtraukti CCD kontroles į diagnostinius testus (Kozlov et al. 2025). Daugumoje atvejų manozilinimas reikšmingai nepaveikia sIgE jungimosi, tačiau pacientams su dideliu kiekiu anti-CCD sIgE gali pasitaikyti klaidingai teigiamų rezultatų net ir naudojant komercinius ImmunoCAP testus, kurie sudaryti iš celiuliozės. Tokias sąveikas galima pašalinti naudojant CCD inhibitorius prieš *in vitro* IgE analizę arba panaikinant N-glikozilinimo vietas (Kozlov et al. 2025; Hemmer et al. 2018). Kita vertus, AIT kontekste mielių glikozilinimas yra naudingas, nes manozilinimas atitinka daugelio natūralių alergenų potransliacines modifikacijas ir leidžia susintetinti rekombinantinius baltymus artimus jų natyviai formai (Al-Ghouleh et al. 2012).

2.6 IgE sąveika su rPen m 4 baltymais Ca^{2+}-neturinčioje aplinkoje

Pen m 4 alergeno IgE reaktyvumą lemia konformacinių ir linijinių epitopų derinys, o dalis konformacinių epitopų, kaip ir daugumoje SKB šeimos baltymų, yra palaikomi Ca^{2+} jonų. Ankstesni tyrimai parodė, kad stiprūs Ca^{2+} chelatoriai gali sumažinti Pen m 4 diagnostinį jautrumą, kaip ir kitiems EF-rankos Ca^{2+} surišantiems alergenams (Morii et al. 2013; Tomura et al. 2008). Alergijų diagnostikoje naudojami tiek serumai, tiek kraujo plazma, kuri būna surinkta, naudojant antikoaguliantus (EDTA, natrio citratą, fluoridą ar hepariną), pašalinančius Ca^{2+} jonus. *In vitro* diagnostikos sistemų kūrime ir taikyme nėra atsižvelgiamą į kraujo surinkimo metodikos poveikį testo rezultatui.

Eksperimento metu buvo įvertintas EDTA poveikis *in vitro* Pen m 4 diagnostikai. Pen m 4 teigiamas serumo mišinys buvo praskiestas buferiniu tirpalu, papildytu 5 mM EDTA; koncentracija būdinga kraujo ėmimo mėgintuvėliams. Rezultatai parodė, kad EDTA sumažino IgE reaktyvumą su rPen m 4 net 61-73%, tuo tarpu reakcija su *P. monodon* ekstraktu reikšmingai nepakito. Didinant

EDTA iki 6 mM papildomo poveikio nepastebėta, tačiau didesnis pakartojimų skaičius padidino statistinį jautrumą, atskleidžiant reikšmingus skirtumus ir ekstrakto atveju.

Nors šie rezultatai preliminarūs, jie rodo, kad kraujas, surinktas į EDTA turinčius mėgintuvėlius, gali iškreipti IgE tyrimų rezultatus tiriant Ca^{2+} surišančius alergenus. Šie tyrimai rodo, kad tik įvertinus alergenų struktūrines savybes galima juos įtraukti į diagnostinius testus.

3. Rekombinantinio alergeno Can f 6 sintezė ir imunologinis apibūdinimas

3.1 E. coli raiškos sistema: rCan f 6 ir rMBP-Can f 6 sintezė ir apibūdinimas

Analogiškai rekombinantinių Pen m 4 variantų sintezei *E. coli* raiškos sistemoje, buvo sukonstruoti pET28a(+)-Can f 6 ir pET28a(+)-MBP-TEV-Can f 6 vektoriai. Sukurtomis plazmidėmis transformavus *E. coli* BL21(DE3) kamieną, rCan f 6 ir rMBP-Can f 6 baltymų sintezė indukuota į terpę pridėjus IPTG. Tirpūs tiksliniai baltymai buvo grynininami kaip anksčiau aprašyta su Pen m 4 baltymais, gautos, atitinkamai, 42 mg/l ir 78 mg/l išeigos. Suliejimas su 6xHis-MBP seka, padidino Can f 6 baltymo išeigą 1,8 karto.

Gryninimo frakcijų analizė baltymų elektroforezės metodu bei imunoblotais su mAk prieš 6xHis inkarinę seką bei prieš MBP baltymą parodė, jog tikslinių baltymų dydžiai, 24 kDa ir 64 kDa, atitinka teorines MW. Can f 6 stabilumo tyrimai parodė, kad *E. coli* pagaminti baltymai yra jautrūs šalčiui ir džiovinimo stresui, analogiškai Pen m 4 variantams, o geriausias stabilumas pasiektas laikant $-20\,^{\circ}\mathrm{C}$ su 40% gliceroliu. Literatūroje aprašomas Can f 6 kiekio kintamumas komerciniuose SPT tirpaluose ir jo žemi kiekiai gali riboti diagnostinį patikimumą, ypač monosensibilizuotiems pacientams (Wintersand et al. 2019; Chan and Leung 2018). Kadangi Can f 6 yra aptinkamas požandikaulinėje seilių liaukoje ir pleiskanose, stebimas ekstraktų kintamumas neturėtų būti priskirtas skirtingų šunų veislių naudojimui gamybos procese (Yamamoto et al. 2019; Fernández-Caldas et al. 2017). Tai rodo, jog tikėtina to priežastis yra firmose taikoma metodika alergenų ekstraktų produkcijai, ypač, pačio ekstrakto pateikimas klientams liofilizuota forma. Gauti rezultatai pabrėžia stabilumo svarbą klinikinių produktų kūrimo procese.

3.2 P. pastoris raiškos sistema: rCan f 6 ir rMBP-Can f 6 sintezė ir apibūdinimas

Rekombinantinių Can f 6 sintezei *P. pastoris* mielėse sukonstruoti, analogiškai Pen m 4 konstruktams, pPink-αF-Can f 6 ir pPink-αF-MBP-TEV-Can f 6 vektoriai. Sukurtomis plazmidėmis transformuoti PichiaPink™ sistemos mielių kamienai ir tikslinių baltymų sintezė indukuota į terpę pridėjus metanolio. At-

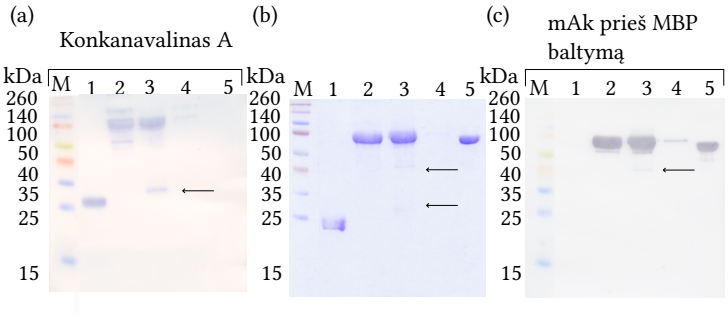
likus baltymų elektroforezę su koncentruotais terpių mėginiais, buvo išrinktas kamienas, sintetinantis didžiausius kiekius tikslinių baltymų – PichiaPink™ K2 (*pep*4 geno nokautas).

Gryninimus su *Ä*KTApurifier 100 sistema atliko Dr. Mindaugas Zaveckas. Baltymas rCan f 6 grynintas, naudojant Ni-NTA chromatografiją; gauta 36 mg/l išeiga. Baltymas rMBP-Can f 6 grynintas, naudojant jonų mainų chromatografiją. CEX būdu išgryninto rMBP-Can f 6 išeiga 18,8 mg/l, o CEX būdu – 57,5 mg/l. MBP baltymo prijungimas, priklausomai nuo gryninimo metodikos, gali padidinti rCan f 6 išeigas 1,6 karto.

Įdomu, jog, analizuojant visų kamienų, transformuotų su pPink-αF-Can f 6 vektoriumi, terpės mėginius baltymų elektroforezėje buvo galima stebėti 21 kDa baltymų juostelę, atitinkančią rCan f 6 MW, bei $\sim$25 kDa baltymų juostelę, kuri galimai atitinka glikozilintą rCan f 6 formą. Siekiant nustatyti, ar tai glikozilintas tikslinio baltymo variantas, atliktos deglikozilinimo ir glikozilinimo analizės su išgrynintais rCan f 6 baltymais. Imunoblotas su mAk prieš 6xHis inkarinę seką parodė, jog deglikozilinus rCan f 6 mėginį, buvo matoma tik viena 21 kDa baltymų juostelė.

Išgrynintų rMBP-Can f 6 baltymų analizė parodė, kad CEX būdu grynintas baltymas degradavo ryškiau nei AEX būdu grynintas mėginys, nors jų glikozilinimo profiliai buvo panašūs (Pav. 6 takeliai 2 ir 3). Baltymų elektroforezėje ir imunoblote su mAk prieš MBP buvo matoma ryški juostelė, atitinkanti MBP turintį fragmentą, o papildoma $\sim$25 kDa juosta, nereaguojanti su mAk prieš MBP, bet sąveikaujanti su Kon A, galimai atitiko Can f 6 turintį fragmentą. Kartu su rMBP-Pen m 4 tyrimų duomenimis ir literatūroje aprašytais sulietų su MBP baltymų skilimo modeliais, šie rezultatai leidžia daryti prielaidą, jog proteolitinis skilimas gali vykti už MBP domeno esančiose sulieto baltymo srityse, t. y. tarpinėje/jungiamojoje srityje ir/arba tikslinio baltymo regione (Z Li et al. 2010). Toks skilimas paaiškintų stebėtus fragmentus, atitinkančius MBP ir rCan f 6 dydžius.

rMBP-Can f 6 degradacija buvo ryškiausia CEX būdu išgrynintuose mėginiuose, todėl galima daryti prielaidą, jog gryninimo sąlygos galėjo prisidėti prie fragmentų susidarymo arba jų išryškėjimo. Tačiau, atsižvelgus į rMBP-Pen m 4 eksperimentų duomenis, kur skilimas kaip tik matomas tik AEX būdu išgrynintuose mėginiuose, galime sakyti, jog proteazės veikimui nedaro įtakos gryninimo pH sąlygos, tikslinio baltymo kiekis, nei tai yra susiję su proteazės A veikla, kadangi baltymai susintetinti *pep*4 geno nokaute.



Pav. 6: Išgrynintų *P. pastoris* susintetintų rCan f 6 ir rMBP-Can f 6 baltymų analizė **(b)** NDS-PAA gelyje ir **(a, c)** imunoblote (pažymėta naudojamu mAk arba lektino pavadinimu). Takeliai: [1], rCan f 6; [2], rMBP-Can f 6, išgrynintas naudojant AEX; [3], rMBP-Can f 6, išgrynintas naudojant CEX; [4], plovimo frakcija iš rMBP-Can f 6 gryninimo naudojant CEX; [5], *E. coli* susintetintas rMBP-Pen m 4 . M – Spectra™ Multicolor Broad Range Protein Ladder. **(a)** Glikozilinimo analizė atlikta naudojant Kon A lektiną; **(c)** Imunobloto analizė atlikta naudojant mAk prieš MBP baltymą. Baltymai atskirti 14% SDS-PAGE geliuose. Rodyklėmis pažymėtos juostelės, galimai turinčios MBP- ir Can f 6-fragmentus.

Kitas svarbus pastebėjimas buvo tai, kad nei rMBP-Can f 6, nei rMBP-Pen m 4 variantai nesąveikavo su mAk prieš 6xHis inkarinę seką. Tai leidžia manyti, kad šiuose *P. pastoris* susintetintuose MBP sulietuose baltymuose buvo prarasta N-galinė 6xHis žymenį turinti sritis. Literatūroje aprašytuose MBP sulietų baltymų modeliuose dažniausiai stebėtas skilimas už MBP domeno esančiose srityse, nepriklausomai nuo MBP padėties konstrukte (Z Li et al. 2010; Moua et al. 2016). Tuo tarpu šiame darbe naudotuose konstruktuose 6xHis žymuo buvo N-gale, t. y. srityje, kuri anksčiau nebuvo aprašyta kaip jautri tokio tipo proteolitiniam skilimui.

Kaip ir kiti aukštesnieji eukariotai, *P. pastoris* sintetina kelių rūšių proteazes, kurios skirstomos į tris pagrindines grupes: citozolinių proteosomų komponentų, vakuolinių proteazių ir sekrecijos kelyje aptinkamų proteazių (Y Zhang, R Liu, et al. 2007). Tyrimuose naudotas PichiaPink™ K2 kamienas yra *pep*4 geno nokautas, nesintetinantis vakuolinės proteazės A, todėl kartu pasižymintis ir sumažėjusiu kitų vakuolinių proteazių aktyvumu (Gast et al. 2022). Taip pat įvertintos sekrecijos kelyje veikiančios dipeptidil aminopeptidazės Ste13p ir Dap2p, tačiau naudotuose konstruktuose nebuvo jų tipinių atpažinimo motyvų, todėl 6xHis žymuo teoriškai turėjo išlikti nepažeistas (Hopkins et al. 2014; Prabha et al. 2009).

Kadangi publikuotų tyrimų su identiškomis ar pakankamai panašiomis konstrukcijomis rasti nepavyko, tiksli N-galinės srities praradimo priežastis lieka neaiški. Kartu, mažai tikėtina, jog 6xHis inkarinė seka buvo erdviškai paslėpta, kadangi su 6xHis-MBP sujungti baltymai sėkmingai aptinkami po produkcijos *E. coli* sistemoje. Todėl tikėtina, kad endogeninė *P. pastoris* proteazė galėjo skelti N-galinę MBP sulietų baltymų sritį.

3.3 rCan f 6 ir rMBP-Can f 6 baltymų variantų imunologinis apibūdinimas

Rekombinantinių Can f 6 alergenų antigeniškumo tyrimams buvo atlikti eksperimentai su alergiškų žmonių kraujo serumais. Imunologinei analizei mėginiai suskirstyti į tris grupes: (i) 19 individualių mėginių, teigiamų Can f 6 sIgE; (ii) teigiama grupė – 7 mėginiai, teigiami *C. familiaris* ekstraktui, bet neigiami Can f 6 alergenui; (iii) neigiama grupė – 7 IgE neigiami mėginiai.

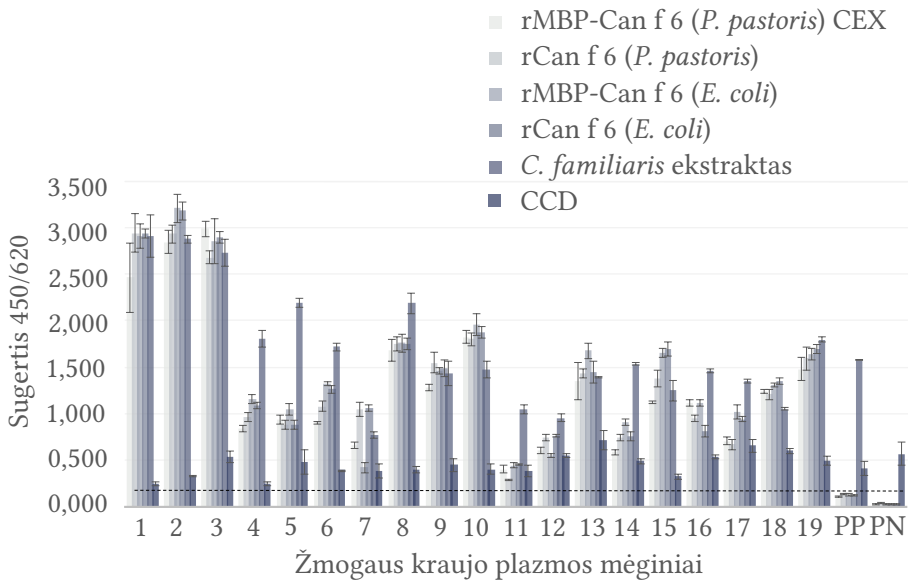
Analizuojant bakterijose ir mielėse susintetintus variantus, pirmiausia, atliktas palyginamasis tyrimas tarp CEX ir AEX metodais išgrynintų rMBP-Can f 6 baltymų. Netiesioginės IFA rezultatai parodė, jog CEX metodu išgrynintas rMBP-Can f 6 reikšmingai stipriau sąveikavo su sIgE nei rMBP-Can f 6, išgrynintas AEX būdu. Šis skirtumas rodo, jog lipokalinas Can f 6 galimai patiria nuo pH-priklausomą struktūrinį pokytį.

Literatūroje nuo pH-priklausomi struktūriniai pokyčiai aprašyti analizuojant ir kitus lipokalinus, tokius kaip nitroforinas (Roberts et al. 2001), ašarų lipokalinas (Gasymov et al. 2010), plazmos retinolį surišantis baltymas (Newcomer and Ong 2000; Calderone et al. 2003), membranos fermentas PagP (Hwang et al. 2002) bei β-laktoglobulinas (Fogolari et al. 2005). Nors lipokalinų aminorūgščių sekos identiškumas dažnai neperžengia 20% ribos, juos sieja keli labai konservatyvūs struktūros motyvai. Lipokalinams būdinga tretinė struktūra yra sudaryta iš aštuonių antiparalelinių β-klosčių, formuojančių β-statinę, ir vienos α-spiralės (Yamamoto et al. 2019; Flower et al. 2000). Ši vandeniliniais ryšiais stabilizuota β-statinė formuoja hidrofobinę ertmę, kuri gali surišti mažas hidrofobines molekules, tokias kaip retinolis, steroidai, kvapiosios medžiagos ir feromonai (Hilger et al. 2012). Galvijų β-laktoglobulino atveju, hidrofobinių ligandų prisijungimas reguliuojamas grįžtamaisiais β-statinės gale esančių kilpų judesiais. Šios kilpos veikia, kaip „dangtelis“, kontroliuojantis prieigą prie statinės ertmės. Manoma, kad šis ligandų prisijungimo ir valdymo mechanizmas yra bendras visiems lipokalinams (Gasymov et al. 2010; Fogolari et al. 2005).

Konformaciniai pokyčiai paaiškintų ir IFA rezultatuose matomus sIgE sąveikos skirtumus, kadangi eksponuojami skirtingi konformaciniai epitopai.

Kita vertus, pH pokyčiai galėjo tiesiog denatūruoti baltymą, eksponuojant alergeno linijinius epitopus (YJ Wang et al. 2017). Kita galimybė yra ta, kad CEX gryninimui reikalingas teigiamas krūvis skirtingai nei AEX naudojamos neigiamos krūvio sąlygos keičia baltymo konformaciją ar stabilumą. Jonų mainų chromatografija yra laikoma auksiniu standartu apibūdinant terapinius preparatus, kadangi skirtingų krūvių atskyrimo procesas gali netiesiogiai parinkti baltymo izoformas su labiau eksponuotais arba nepažeistais IgE prisijungimo epitopais (Fekete et al. 2015).

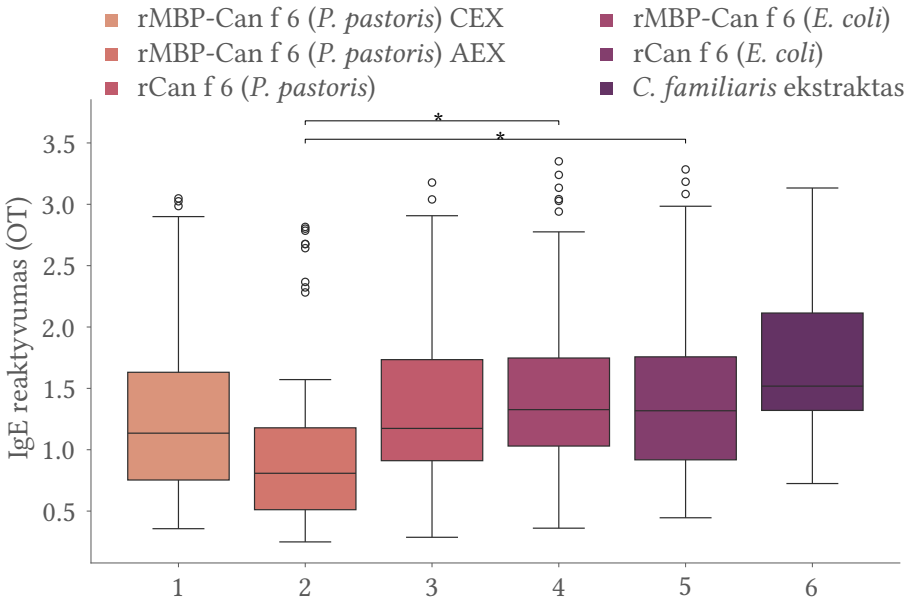
Tęsiant visų rCan f 6 baltymų analizę netiesioginės IFA metodu, galime stebėti, jog visi susintetinti baltymai specifiškai sąveikauja su Can f 6 sIgE ir nerodo kryžminės sąveikos su nespecifiniais IgE (Pav. 7).



Pav. 7: rCan f 6 baltymų variantų imunoreaktyvumo analizė netiesioginėje IFA sistemoje. CEX – rMBP-Can f 6, išgrynintas katijonų mainų chromatografijos būdu, o AEX – anijonų mainų chromatografijos būdu. CCD yra glikozilinimo markeris. [1-19] yra individualūs *C. familiaris* alergija sergančių žmonių kraujo plazmos mėginiai, kurie yra teigiami Can f 6 alergenui; PP – teigiama grupė (*C. familiaris* sIgE teigiama, bet Can f 6 sIgE neigiama, n = 7); PN – neigiama grupė (n = 7). Ribinė reikšmė pažymėta punktyrine linija.

Atliktos vienfaktorinės blokuotų duomenų dispersinės analizės rezultatai parodė, jog tarp tirtų baltymų matomi statistiškai reikšmingi skirtumai ($F(4,\ 72) = 29,8,\ p < 0,001$). Kartu, glikozilinimo įtaką rezultatams buvo įvertinta su CCD kontrole, tačiau nebuvo rasta reikšmingų skirtumų.

IFA rezultatų post hoc analizė atlikta su Tukio HSD testu identifikuojant specifines baltymų poras su reikšmingais reaktyvumo skirtumais (Pav. 8). Ši analizė atlikta su nesugrupuotais duomenimis, išryškinant tik esminius skirtumus tarp porų. Toks statistinis metodas pasirinktas, kadangi Bonferoni analizė parodo, jog visos tirtos baltymų poros statistiškai reikšmingai skiriasi viena nuo kitos, o Tukio HSD analizė su sugrupuotais duomenimis dėl ypač konservatyvių ANOVA prielaidų neranda nei vienos statistiškai reikšmingos poros. Tad nors Tukio HSD taikymas nesugrupuotiems duomenims padidino statistinę galią, jis veiksmingai sumažino atsitiktinės variacijos įtaką ir nustatė tik esminius skirtumus. Post hoc analizė parodė, jog AEX būdų išgrynintas rMBP-Can f 6 statistiškai reikšmingai išsiskyrė nuo *E. coli* susintetintų Can f 6 baltymų. Tarp likusių porų nebuvo pastebėta statistiškai reikšmingų skirtumų, o tai rodo, kad šių grupių reaktyvumo lygis yra panašus.



Pav. 8: Grafikas, vaizduojantis rCan f 6 baltymų variantų imunoreaktyvumo pasiskirstymą. *P. pastoris* susintetintas rMBP-Can f 6, išgrynintas [1] CEX ir [2] AEX būdu; [3], *P. pastoris* susintetintas rCan f 6; [4], *E. coli* susintetintas rMBP-Can f 6; [6], *C. familiaris* alergenų ekstraktas. Duomenys atspindi normalizuotą IgE reaktyvumą su 19 individualių *C. familiaris* alergija sergančių žmonių kraujo plazmos mėginių, kurie yra teigiami Can f 6 alergenui. Reaktyvumas pateikiamas kaip normalizuotos OT vertės, o reikšmingi skirtumai tarp baltymų pažymėti žvaigždutėmis ($** = p < 0.01$, $*** = p < 0.001$).

Šio tyrimo tikslas buvo įvertinti baltymo Can f 6 sintezę eukariotinėje

raiškos sistemoje, analizuojant jo gebėjimą specifiškai atpažinti sIgE antikūnus, nesąveikauti su nespecifiniais IgE bei galimybes užtikrinti didelės išeigos sintezę. rCan f 6 suliejimas su MBP abejose sistemose reikšmingai padidino baltymo išeigą, patvirtinant šį metodą kaip efektyviausią rCan f 6 sintezės strategiją. MBP padidino produkciją 1,8 karto *E. coli* sistemoje (42 mg/l į 78 mg/l) ir 1,6 karto *P. pastoris* sistemoje (36 mg/l į 57,5 mg/l, AEX).

Visų rekombinantinių alergenų, sujungtų ir nesujungtų su MBP, palyginamoji analizė parodė, kad tiek prokariotinė, tiek eukariotinė raiškos sistema yra tinkamos rekombinantinio Can f 6 sintezei. Visi baltymai specifiškai reagavo su Can f 6 sIgE ir nerodė nespecifinio prisijungimo, patvirtinant esminių sIgE prisijungimo epitopų išsaugojimą. Tai rodo jų tinkamumą imunologiniams tyrimams ir galimiems diagnostiniams taikymams. Vis dėlto, neatlikus bazofilų degranuliacijos tyrimų, reikalingi papildomi tyrimai alergeniniam aktyvumui *in vivo* įvertinti.

Išskyrus rMBP-Can f 6, išgrynintą AEX metodu, statistiškai reikšmingų IgE reaktyvumo skirtumų tarp kitų rekombinantinių baltymų nenustatyta. Glikozilinimas, dažnai siejamas su kryžminiu reaktyvumu *in vitro* alergijos diagnostikoje, neturėjo įtakos rekombinantinių Can f 6 variantų IgE reaktyvumui, todėl eukariotinė raiškos sistema nėra būtina. Can f 6 galima efektyviai gaminti *E. coli* sistemose.

Tyrimas parodė, kad skirtingos raiškos sistemos neturi reikšmingos įtakos rCan f 6 sIgE prisijungimui, tad ir *E. coli* ir *P. pastoris* raiškos sistemos gali būti naudojamos imunoreaktyvių Can f 6 baltymų sintezei. N-glikozilinimas nėra būtinas Can f 6 struktūrai ar jo sąveikai su sIgE. Suliejimas su MBP baltymu yra efektyvi rekombinantinių alergenų sintezės strategija ir tinkama tolimesniems tyrimams, kadangi MBP nesąveikauja su IgE. Atsižvelgiant į baltymo išeigą ir reaktyvumą su sIgE, *E. coli* pagamintas rMBP-Can f 6 yra perspektyvus kandidatas didelio masto sintezei. Tačiau įvertinus galimą užteršimą su bakteriniais endotoksinais, optimalia alternatyva laikytinas *P. pastoris* sintetintas rMBP-Can f 6.

4. PichiaPink™ sistemos metodologijos validavimas

4.1 Alergenų raiškos vektorių konstravimas

Nustačius 6xHis-MBP-alergeno konstrukcijų N-galinę proteolizę, buvo sukurti pPink-αF-6xHis ir pPink-αF-6xHis-MBP-TEV standartizuoti pPink-HC vektoriai, siekiant supaprastinti klonavimo eigą ir sistemiškai ištirti proteolizės reiškinį kitose alergenų šeimose.

4.2 Raiškos sistemos validavimas su paprastųjų vapsvų nuodų alergenu Ves v 5

Sistemos validavimui pasirinktas paprastųjų vapsvų nuoduose aptinkamas baltymas Ves v 5, kuris yra pagrindinis Hymenoptera alergijos diagnostinis taikinys. Šio alergeno struktūra yra stabilizuota disulfidiniais tilteliais, todėl rekombinantinei sintezei būtina eukariotinė sistema. Bandymai susintetinti baltymą *E. coli* sistemose buvo nesėkmingi, gaunami inkliuziniai kūneliai, kuriuos būtina *in vitro* perlankstyti. Tuo tarpu eukariotinėse sistemose gaunami tinkamai susilankstę variantai, artimi natyviam baltymui (Borodina et al. 2010; Kischnick et al. 2006; Schiener et al. 2017). Ves v 5 sintezė *P. pastoris* sistemoje jau buvo sėkmingai ištirta (Monsalve et al. 1999), tačiau jo sintezė kartu su MBP dar nebuvo nagrinėta. Dėl struktūrinių reikalavimų ir klinikinės svarbos molekulinei diagnostikai bei alergenų imunoterapijai Ves v 5 yra tinkamas įvertinti PichiaPink™ sistemos efektyvumą, ir galimą N-galo proteolizę.

Siekiant palyginti gautus rekombinantinius Ves v 5 variantus, rVes v 5 buvo susintetinti ir *E. coli* sistemoje, nors, remiantis literatūros duomenimis, ši sistema nėra tinkama šio baltymo sintezei. Išbandyti Bl21(DE3) ir Rosetta kamienai, tačiau rVes v 5 buvo aptinkamas tik inkliuziniuose kūneliuose, o rMBP-Ves v 5 buvo tirpus tik Bl21(DE3) kamiene. Ni-NTA bei ZYMO kolonėlėmis išgryninti baltymai buvo dializuoti bei koncentruoti. Baltymų rVes v 5 ir rMBP-Ves v 5 gautos, atitinkamai, 14 mg/l ir 55 mg/l išeigos. Suliejimas su 6xHis-MBP seka, padidino Ves v 5 baltymo išeigą 3,9 karto.

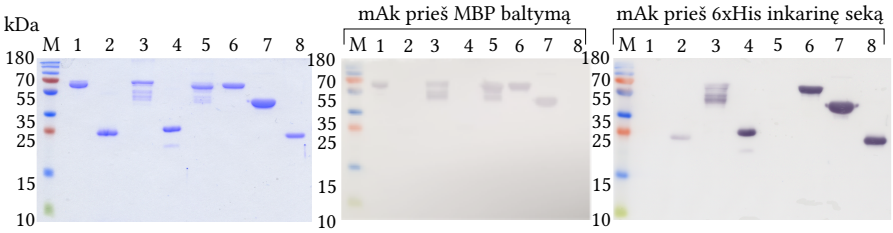
Rekombinantinių Ves v 5 sintezei *P. pastoris*, 204-a.r. ilgio Ves v 5 baltymą koduojanti seka įklonuota į sukonstruotų pPink-αF-6xHis ir pPink-αF-6xHis-MBP-TEV vektorių SmaI ir KpnI restrikcijos vietas. pPink-αF-6xHis-Ves v 5 ir pPink-αF-6xHis-MBP-TEV vektoriai patikrinti VU GMC BTI sekoskaitos centre.

Po transformacijos į PichiaPink™ kamienus, bandomieji eksperimentai parodė, jog po 48 val. indukcijos, didžiausia baltymų sintezė buvo matoma K2 kamiene. Baltymų elektroforezė patvirtino sėkmingą 24,3 kDa rVes v 5 ir 67,8 kDa rMBP-Ves v 5 sintezę; dvigubos juostos, stebimos gelio nuotraukose, parodė glikozilinimą, būdingą *P. pastoris* sistemai. Imunoblotas su mAk prieš 6xHis inkarinę seką parodė, jog *E. coli*-susintetinti rVes v 5 variantai abu sąveikavo, tačiau *P. pastoris*-susintetinti baltymai reagavo skirtingai. Todėl, atitinkamai, rVes v 5 buvo grynintas taikant Ni-NTA chromatografiją, o rMBP-Ves v 5 buvo grynintas naudojant jonų mainų chromatografiją. Gryninimus su $\ddot{A}$KTApurifier 100 sistema atliko Dr. Mindaugas Zaveckas. Išeigos, nustatytos

Bradford metodu ir denzitometrine analize, reikšmingai skyrėsi: rVes v 5 – 36 mg/l, rMBP-Ves v 5 – 134,5 mg/l, t. y. 3,7 karto didesnė išeiga suliejus su MBP. Palyginus rezultatus su randamais literatūroje, suliejimas su MBP sėkmingai pralenkia ankstesnius rezultatus su *E. coli*, bakulovirusų-vabzdžių ląstelių ar kitų mielių sistemomis (Suck et al. 2000; Monsalve et al. 1999; Henriksen et al. 2001).

Norint ištirti 6xHis inkarinės sekos buvimą, buvo atlikti imunoblotai su mAk prieš 6xHis inkarinę seką ir prieš MBP (Pav. 9). rMBP-Ves v 5 (Pav. 9 takelis 1) ir rMBP-Pen m 4 (Pav. 9 takelis 5) nebuvo aptinkami su mAk prieš 6xHis inkarinę seką, tačiau abu baltymai reagavo su mAk prieš MBP. Tai rodo, kad PichiaPink™ sistemoje susintetintuose 6xHis-MBP-*alergenas* konstruktuose pasikartoja N-galinės 6xHis žymenį turinčios srities praradimas, nepriklausomai nuo alergeno šeimos. taip pat parodė neatitikimą tarp nustatytų ir teorinių MW: rVes v 5 masė buvo 7 Da didesnė už teorinę, o rMBP-Ves v 5 – apie 2 kDa mažesnė. Nedidelį MW padidėjimą galima sieti su nepilnu deglikozilinimu ar disulfidinių ryšių susidarymu, o rMBP-Ves v 5 MW sumažėjimas atitinka N-galinės srities proteolitinį pašalinimą. Remiantis monoizotopinių masių skaičiavimais, galima skilimo vieta yra 6xHis jungties ir trombino atpažinimo motyvo regione.

Literatūros duomenimis, αF-6xHis-*baltymas* konstruktai gali būti sėkmingai sintetinami įvairiuose *P. pastoris* kamienuose, įskaitant PichiaPink™ sistemą (Smith et al. 2023; Chronopoulou et al. 2023). Tačiau yra duomenų ir apie N-galinės srities proteolizinį apdorojimą: tiriant αF-6xHis-GFP-*baltymas* konstruktus, 6xHis inkarinė seka nebuvo aptinkama, nors GFP signalas išliko (Zheng et al. 2022).

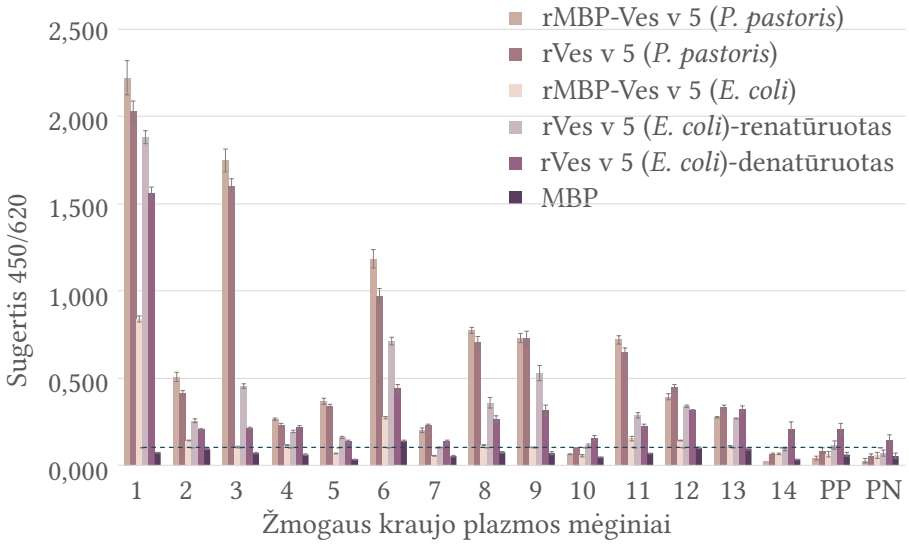


Pav. 9: Išgrynintų *P. pastoris* susintetintų rVes v 5 variantų analizė NDS-PAA gelyje ir imunoblote (pažymėta naudojamu mAk arba lektino pavadinimu). Takeliai: [1], *P. pastoris*-susintetintas rMBP-Ves v 5; [2], *P. pastoris*-susintetintas rVes v 5; [3], *E. coli*-susintetintas rMBP-Ves v 5; [4], *E. coli*-susintetintas rVes v 5; [5], *P. pastoris*-susintetintas rMBP-Pen m 4; [6], *E. coli*-susintetintas rMBP-Pen m 4; [7], *E. coli*-susintetintas MBP su 6xHis inkarine seka; [8] *E. coli*-susintetintas rPen m 4. M – PageRuler™ Plus Prestained Protein Ladder #26619. Baltymai atskirti 14% NDS-PAA geliuose.

Mūsų žiniomis, *P. pastoris* atliekama N-galo proteolizė baltymuose, sulietuose su MBP, anksčiau nebuvo dokumentuota. Tolimesniuose darbuose bus tiriama, ar N-galo proteolizę skatina trombino atpažinimo motyvas ir ar sekos modifikacijos galėtų išsaugoti 6xHis inkarinės sekos integralumą.

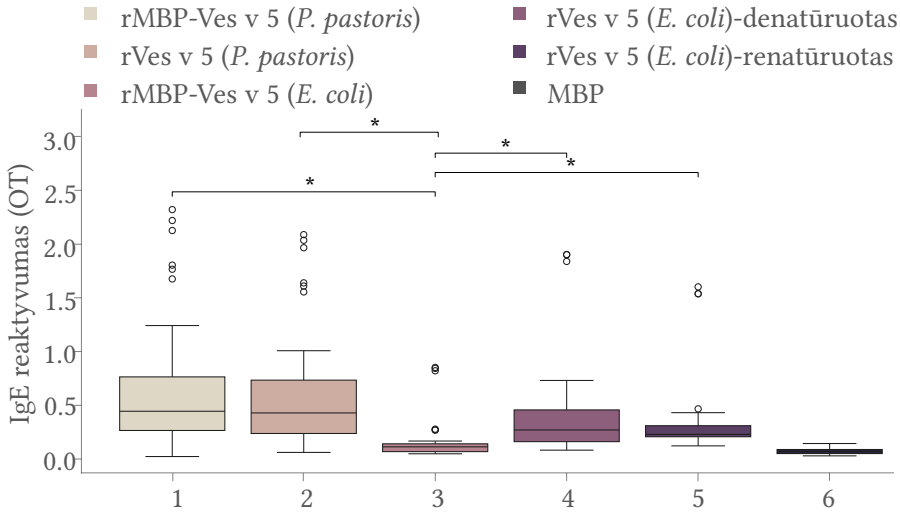
Rekombinantinių Ves v 5 alergenų antigeniškumui ištirti buvo atliekami tyrimai su trimis grupėmis alergiškų žmonių kraujo serumų: (i) 14 individualių mėginių, teigiamų Ves v 5 sIgE; (ii) teigiama grupė – 5 mėginiai, teigiami Pen m 4 sIgE, bet neigiami Ves v 5 sIgE; (iii) neigiama grupė – 5 IgE neigiami mėginiai.

Netiesioginės IFA rezultatai atskleidė reikšmingus skirtumus tarp *P. pastoris* ir *E. coli* susintetintų rekombinantinių Ves v 5 baltymų (Pav. 10). *P. pastoris* susintetinti rVes v 5 parodė specifinę sąveiką su sIgE ir kryžmiškai nereagavo su kontrolinėmis grupėmis. *E. coli* susintetinti variantai – tiek išgryninta denatūruota forma, tiek perlankstyta – nespecifiškai reagavo su kontrolinėmis grupėmis. Vienfaktorinė blokuotų duomenų analizė patvirtino, jog tarp tirtų baltymų variantų matomi statistiškai reikšmingi skirtumai ($F(4,\ 52) = 11,17$, $p < 0,001$). Imunologinė analizė parodė, jog galimai neteisingai susilankstę *E. coli* susintetinti Ves v 5 variantai gali sukelti klaidingai teigiamą signalą, kadangi keli serumai reagavo su *E. coli* variantais, tačiau nereagavo su tinkamai susilanksčiusiais *P. pastoris* variantais. Nors šis skirtumas galėtų būti sąlygotas sIgE koncentracijos, IFA naudotų serumų analizė parodė, jog panašias ALEX vertes turintys serumai demonstravo skirtingus prisijungimo profilius, todėl stebimas reaktyvumas turėtų būti priklausomas nuo sintezės sistemos.



Pav. 10: rVes v 5 baltymų variantų imunoreaktyvumo analizė netiesioginėje IFA sistemoje. [1-14] yra individualūs *V. vulgaris* alergija sergančių žmonių kraujo plazmos mėginiai, kurie yra teigiami Ves v 5 alergenui; PP – teigiama grupė (*V. vulgaris* ir Ves v 5 sIgE teigiama, n = 5); PN – neigiama grupė (n = 7). MBP naudojamas kaip neigiama kontrolė. Ribinė reikšmė pažymėta punktyrine linija.

Poriniai palyginimai su Bonferroni korekcija patvirtino, jog *P. pastoris* susintetinti rVes v 5 ir rMBP-Ves v 5 reikšmingai išsiskyrė tarp visų variantų (Pav. 11). rMBP-Ves v 5 antigeniškumą papildomai patvirtino inhibicinis tyrimas, atliktas Laimio Silimavičiaus: 0,05 mg *P. pastoris* susintetinto rMBP-Ves v 5 slopino Ves v 5 sIgE prisijungimą 92%. Apibendrinus IFA rezultatus bei pasiektas išeigas, *P. pastoris* susintetintas rMBP-Ves v 5 yra optimalus pasirinkimas paprastųjų vapsvų alergijų diagnostikai.



Pav. 11: Grafikas, vaizduojantis rVes v 5 baltymų variantų imunoreaktyvumo pasiskirstymą. Takeliai: [1], *P. pastoris*-susintetintas rMBP-Ves v 5; [2], *P. pastoris*-susintetintas rVes v 5; [3], *E. coli*-susintetintas rMBP-Ves v 5; [4], *E. coli*-susintetintas rVes v 5, denatūruota forma; [4], *E. coli*-susintetintas rVes v 5, renatūruota forma;[6], MBP. Duomenys atspindi normalizuotą IgE reaktyvumą su 14 individualių *V. vulgaris* alergija sergančių žmonių kraujo plazmos mėginių, kurie yra teigiami Ves v 5 alergenui. Reaktyvumas pateikiamas kaip normalizuotos OT vertės, o reikšmingi skirtumai tarp baltymų pažymėti žvaigždutėmis ($* = p < 0.05$).

Įdomu tai, kad bandymai su kitais *V. vulgaris* alergenais (Ves v 1 – fosfolipazė A1, Ves v 2 – hialuronidazė, Ves v 3 – į dipeptidilpeptidazę IV panaši serino proteazė) nebuvo sėkmingi nei su MBP, nei be jo, nepaisant intensyvios optimizacijos. Tai rodo, jog sukurta metodika yra labiau tinkama mažiems, mažiau glikozilintiems baltymams, bet netinka katalitiškai aktyviems, stipriai glikozilintiems fermentams, kurie gali sukelti ER stresą, netaisyklingą lankstymąsi ar toksiškumą šeimininkui. Šie rezultatai parodo, jog rekombinantinių alergenų sintezė *P. pastoris* sistemoje turi būti empiriškai ištirta, nes ji priklauso nuo baltymo struktūros, potransliacinių modifikacijų poreikių ir suderinamumo su mielių sekrecine sistema.

IŠVADOS

1. *P. pastoris* ir *E. coli* raiškos sistemose susintetinti rekombinantiniai Pen m 4 ir Can f 6 alergenai yra antigeniškai panašūs į natūralius alergenus, nes reaguoja su jiems specifiniais IgE klasės antikūnais. Rekombinantinio Ves v 5 sintezei tinkama tik *P. pastoris* raiškos sistema.

2. N-glikozilinimas nėra būtinas Pen m 4 ir Can f 6 alergenų sąveikai su alergenams specifiniais IgE klasės antikūnais, tačiau potransliacinės modifikacijos yra būtinos Ves v 5 alergenui.

3. Suliejimo su maltoze-surišančiu baltymu (MBP) strategija padidino baltymų išeigas tirtose raiškos sistemose, nedarant įtakos sąveikai su IgE klasės antikūnais, patvirtinant MBP kaip naudingą įrankį rekombinantinių alergenų sintezėje.

4. *P. pastoris* susintetintuose MBP-*alergenas* baltymų konstruktuose stebimas proteolitinis skilimas už MBP domeno patvirtina ankstesnius duomenis apie su MBP susijusį proteolitinį aktyvumą, veikiantį C-galinėje MBP domeno pusėje esančias sritis.

5. *P. pastoris* susintetintuose MBP-*alergenas* baltymų konstruktuose nustatytas anksčiau neaprašytas N-galinės srities proteolitinis skilimas, lemiantis 6xHis žymenį turinčios srities praradimą ir ribojantis tokių konstruktų taikymą PichiaPink™ sistemoje.

6. *P. pastoris* susintetintas Pen m 4 formuoja neįprastai stabilius dimerus, atsparius cheminiam ir terminiam poveikui, o tai gali būti susiję su glikozilinimo įtaka baltymo konformacijai ir turėti potencialią diagnostinį reikšmę.

7. EDTA, kalcio chelatorius, sumažino Pen m 4 alergeno reaktyvumą, tai rodo kalcio jonų svarbą IgE atpažinimui ir pabrėžia būtinybę diagnostiniuose tyrimuose standartizuoti mėginių paruošimo ir laikymo sąlygas.

8. Nustatyta, kad *E. coli* susintetinti Pen m 4 ir Can f 6 alergenai liofilizacijos metu degraduoja, tai galimai paaiškina jų nestabilumą komerciniuose ekstraktuose ir pabrėžia laikymo sąlygų svarbą.

PUBLICATION LIST

Thesis relevant publications:

1. **Rainyte J.**, Zvirblis G., Zaveckas M., Kucinskaite-Kodze I., Silimavicius L., and Petraityte-Burneikiene R. Immunological comparison of recombinant shrimp allergen Pen m 4, produced in *Pichia pastoris* and *Escherichia coli*. *Journal of Biotechnology* 2023;369:1-13. doi:10.1016/j.jbiotec.2023.05.002

2. **Dvareckiene J.**, Zvirblis G., Zaveckas M., and Petraityte-Burneikiene R. Enhancing production and assessing IgE reactivity of dog allergen Can f 6 in *Pichia pastoris* and *Escherichia coli*. *Applied Microbiology and Biotechnology* 2025;109(78):2-16. doi:10.1007/s00253-025-13465-7

Thesis irrelevant publications:

1. Statkevičiūtė R., Sadauskas M., **Rainytė J.**, Kavaliauskaitė K., and Meškys R. Comparative Analysis of Mesophilic YqfB-Type Amidohydrolases. *Biomolecules* 2022;12(10):1492. doi:10.3390/biom12101492

CONFERENCES

1. International conference FEBS3+. Poster presentation *Pichia pastoris expression and purification of Canis familliaris allergen Can f 6.* 06 17-19, 2019. Riga, Latvia

2. International conference The COINS. Poster presentation *Canis familiaris allergen Can f 6 and Penaeus monodon allergen Pen m 4: expression, purification and analysis.* 02 26-28, 2020. Vilnius, Lithuania

3. Conference Microbiology 2022. Poster presentation *Yellow jacket major venom allergen Ves v 5 immunological characterization.* 04 28-29, 2022. Birštonas, Lithuania

4. International conference FEBS3+. Oral presentation *Immunological comparison of recombinant shrimp allergen Pen m 4, produced in Pichia pastoris and Escherichia coli.* 06 15-17, 2022. Tallinn, Estonia

CURRICULUM VITAE

Name, Surname	**Juta Dvareckienė (Rainytė)**
Work address	Vilnius University, Life Sciences Center, Institute of Biotechnology, Eukaryote Gene Engineering Department (VU LSC IBT EGED), Saulėtekio al. 7, Vilnius
Email address	juta.rainyte@bti.vu.lt
Education	
2012–2016	VU, B.Sc. in Molecular Biology
2016–2018	VU, M.Sc. in Molecular Biology, *Magna Cum Laude*
2018–2022	Chemical engineering doctoral studies, VU LSC IBT EGED
Experience	
2016–2017	VU LSC IBT EGED – Laboratory Technician
2018–2022	VU LSC IBT EGED – Specialist
2018–2023	VU LSC IBT EGED – Biologist Researcher
2020–2021	Baltymas Ltd. – Bioengineer
2023–2024	VU LSC Institue of Biosciences – Junior Teaching Assistant
2020–present	VU LSC – Junior Researcher
2025–present	VU LSC IBT EGED – Biologist Researcher
Society fellow	
2018–present	Member, Lithuanian Biochemical Society
Outreach	
2015	Protein electrophoresis lab, Kaunas Fortress VII Fort.
2017	Joint protein electrophoresis lab (Kaunas Fortress VII Fort + Lithuanian Biology Teachers Association).
2018–2025	European Biotechnology Week (annual): conducting laboratory work
2021	Student lecture (COINS “Who needs it?”)
2022	Popular-science article on delfi.lt
Service	
2020-2021	Representative of PhD students at IBT

PADĖKA

Dėliojant paskutinius taškus disertacijos lapuose, aš noriu išreikšti didžiausią padėką savo vadovei dr. Rasai Petraitytei-Burneikienei. Pirmus mano žingsnius laboratorijoje lydėjo Jūsų ranka, pirmiausia, nuvedusi pas dr. Justą Lazutką, kuris išmokė visų mokslo subtilybių bakalauro ir magistro laipsnių siekiant. Ir, galiausiai, atvedusi į doktorantūros erą, kurioje išmokau stovėti ant savo kojų kaip mokslininkė. Tikrai, nuoširdžiai Jums dėkoju už Jūsų kantrybę, atvirumą, neišsenkančius patarimus ir galimybę atlikti visus mokslinius darbus EGIS skyriuje.

Mokslinis darbas negali būti atliktas tik vieno žmogaus – patarimai, mokslinės diskusijos, patirties perdavimas, palaikymas ir pagalba darbuose – kiekvienas mokslinis etapas yra lydimas kitų mokslininkų. Todėl noriu, pirmiausia, padėkoti savo esamiems ir buvusiems laboratorijos kolegoms: Justui Lazutkai, Alionai Avižinienei, Pauliui Lukui Tamošiūnui, Arūnei Verbickaitei ir Laimai Čepulytei. Ačiū skiriu ir studentams, ypač Ievai Lingytei už vertingą vadovavimo patirtį. Taip pat dėkoju visam EGIS kolektyvui: Mindaugui Zaveckui, Danguolei Žiogienei, Mildai Norkienei, Eimantui Žitkui, Evaldui Čipliui, Editai Bakūnaitei ir visiems kitiems EGIS kolegoms. Atskirą padėką skiriu buvusiam EGIS vadovui – dr. Gintautui Žvirbliui už prijungimą į alergenų projektą, kuris atvėrė man kelią į šią ilgai svajotą tematiką.

Dėkoju Laimiui Silimavičiui ir Imunodiagnostikai už galimybę mokytis dirbti su mikrogardelių sistemomis. Taip pat dėkoju ILBS kolektyvui už bendradarbiavimą, Jūsų nuostabių antikūnų pasidalinimą.

Ačiū draugams už nuolatinį palaikymą ir galimybę išsilieti dėl visų vargų ir džiaugsmų. Jūsų buvimas šalia yra dovana.

Ir, svarbiausias ačiū yra šeimai. Ačiū, kad palaikėte, kad tikėjote, kad rūpinotės. Ačiū už supratingumą, kad mokslininko gyvenimas yra keistas ir įdomus. Doktorantūros neįsivaizduoju ir be savo antros pusės, Danieliaus, kuris buvo ir yra mano ramstis ir užuovėja. Ačiū, kad mane jau nuo bakalauro vežioji naktimis į laboratoriją, nes reikia užsėti ar indukuoti mieles; ačiū, kad niekad nepykai, kad visus savaitgalius praleisdavau darbe; ačiū, kad ramindavai ir guosdavai mane visais sunkiais etapais. Ačiū, meile, už visą palaikymą.

Jūsų visų dėka galėsiu užbaigti šią erą ir keliauti toliau. Ačiū.

NOTES

Vilniaus universiteto leidykla
Saulėtekio al. 9, III rūmai, LT-10222 Vilnius
El. p. info@leidykla.vu.lt, www.leidykla.vu.lt
bookshop.vu.lt, journals.vu.lt
Tiražas 15 egz.