Identification and analysis of putative allergens in Edible mushroom (Agaricus bisporus)

A dissertation

Submitted in partial fulfilment of the requirement for the award of the degree of

M.Sc. Biotechnology (July 2025)

Under the guidance of

Dr. Atul Kumar Upadhyay and Dr. Vikas Handa

Assistant Professor

Dept. of Biotechnology

TIET, Patiala

Submitted by – Priyanka Yadav

Roll. No. - 302301010

DEPARTMENT OF BIOTECHNOLOGY, THAPAR INSTITUTE OF ENGINEERING AND TECHNOLOGY, PATIALA (PUNJAB) -147004

DECLARATION

I, Priyanka Yadav, sincerely declare that the research project titled “Identification and analysis of putative allergens in Edible mushroom (*Agaricus bisporus*)" is the result of my own independent work, carried out at Thapar Institute of Engineering and Technology, Patiala. This work was undertaken as part of my 6-month M.Sc. dissertation in Biotechnology, under the guidance of Dr. Atul Kumar Upadhyay and Dr. Vikas Handa from January 2025 to July 2025.

To the best of my knowledge, this work has not been submitted to any other institute or university for the award of any degree, diploma, or certification.

Priyanka Yadav

302301010

M.Sc. Biotechnology

DBT, TIET

Patiala - 147004

Place – Patiala

Date – 30 June, 2025

CERTIFICATE

This is to certify that the dissertation titled, *"Identification and Analysis of Putative Allergens in Edible Mushroom (Agaricus bisporus)"*, submitted in partial fulfilment of the requirements for the award of the Master of Science degree in the Department of Biotechnology at Thapar Institute of Engineering and Technology, Patiala, is a bona fide piece of work carried out by Priyanka Yadav (302301010) under my supervision and guidance.

To the best of our knowledge, the work presented in this dissertation is original and has not been submitted, either in part or full, to any other University or Institute for the award of any degree or diploma.

Dr. Atul Kumar Upadhyay

Assistant Professor

DBT, TIET

Dr. Vikas Handa

Assistant Professor

DBT, TIET

Acknowledgements

I express my deepest gratitude to my supervisors, **Dr. Atul Kumar Upadhyay** and **Dr. Vikas Handa**, Assistant Professors in the Department of Biotechnology, Thapar Institute of Engineering and Technology, Patiala. I could not have asked for better mentors. Your support made this feel like more than just a project and it became something I could truly call my own.

A heartfelt thank you to **Dr. M.S. Reddy**, Head of the Department, for creating an environment that felt safe to learn and grow. And to the non-teaching staff, thank you for all the quiet ways in which you kept everything running.

Now to my dearest friends **Jayesh and Musharraf**. You saw every version of me during this journey: the focused, the frustrated, the defeated, and the completely sleep-deprived. Thank you for staying. Thank you for lifting me up on the days I couldn't see the point, for calling me out gently when I spiralled, and for reminding me that I'm more than my work. Your emotional support, the way you just got me without needing an explanation and I honestly don't think I could have made it through this without you.

And to my **family**, my whole heart, thank you for holding me through this in ways only you could. For letting me be messy and unsure, for understanding when I was distant or overwhelmed, and for never once making me feel like I had to earn your love. Your belief in me, even when I couldn't find it in myself, was the quiet strength that carried me forward. Every word I wrote, every analysis I ran, I carried you all with me.

This project is as much yours as it is mine. Thank you, deeply and truly.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Full Form
GO	Gene Ontology
BLAST	Basic Local Alignment Search Tool
FAO	Food and Agriculture Organization
WHO	World Health Organization
MSA	Multiple Sequence Alignment
HLA	Human Leukocyte Antigen
IEDB	Immune Epitope Database
RE	Relative Enrichment
IUIS	International Union of Immunological Societies

ABSTRACT

BACKGROUND

Allergies result from abnormal immune responses to otherwise harmless proteins. In recent years, the incidence of food allergies has increased significantly, necessitating improved methods for allergen identification and risk assessment. While conventional allergen detection methods rely on clinical or laboratory-based protocols, computational approaches offer a faster and scalable alternative for preliminary allergenicity screening. *Agaricus bisporus* var. *burnetti*, a widely consumed edible mushroom, remains understudied with respect to its allergenic potential. This study employs a multi-layered in silico strategy to evaluate the allergenicity of *Agaricus bisporus* proteins using peptide frequency analysis, epitope mapping, and structural modeling.

RESULTS

Putative allergens were first shortlisted through sequence-based homology screening using FASTA and BLAST, identifying proteins with >50% identity to known allergens in the AllergenOnline database. Amino acid, dipeptide, tripeptide, and tetrapeptide compositions of the full proteome (11,675 proteins), 55 putative allergens, and 2,334 known allergens were compared. Physicochemical property-based predictions from AllergenFP, AlgPred, and AllerTOP further validated the allergenicity of several candidates. 3D structures of key putative allergens were modeled using SWISS-MODEL. Protein-peptide docking simulations conducted via ClusPro 2.0 identified strong binding affinities between modeled epitopes and known IgE-binding domains. Functional classification using Uniprot and GO annotations revealed biological relevance of candidate proteins in stress and defense-related pathways. Peptidomic profiling revealed distinct enrichment of specific short peptides in the allergen datasets relative to the proteome. Linear B-cell epitopes were retrieved from IEDB and matched against putative allergens to uncover immunologically relevant motifs.

CONCLUSION

This study demonstrates that *in silico* analysis can reliably identify and characterize putative allergens in *Agaricus bisporus* var. *burnetti* from the rest of the proteome. Peptide-level enrichment analysis, epitope mapping, and structure-based validation collectively provide a novel framework for early-stage allergenicity screening in novel food proteins.

Keywords: Agaricus bisporus var. burnetti; allergy; food allergy; fungal allergens; peptide enrichment; in silico allergen prediction; cross-reactivity; B-cell epitopes.

GRAPHICAL ABSTRACT

CHAPTER 1 Introduction

Allergies continue to emerge as a global concern, often making their presence felt through symptoms ranging from mild discomfort to anaphylactic episodes (Anvari S. et al., 2019). At the core of it all, are certain proteins called allergens that mistakenly activate the immune system in atopic individuals, most commonly through Immunoglobulin E (IgE)-mediated immunogenic pathways (Ramsey N. and Berin M.C., 2021) (Wang J. et al., 2023). Allergens do not usually cause any exaggerated responses in normal individuals i.e. they are otherwise harmless. Allergenic proteins are ubiquitously present in the environment at all times and that is what make allergies more and more dangerous with each passing day. Ubiquitous nature of allergens results in manifestation of allergies of different kinds such as airway allergies, ingestive/food allergies, contact allergies and injection/venom allergies (Wang J. et al., 2023).

With an increase in consumption of alternate food sources and GMO based edible products in the market, there is a consistent rise in food allergy prevalence, especially among children, which points clearly towards the need for more inclusive approaches to identify putatively allergenic proteins especially from food sources that haven't yet been explored enough (Elghoudi A. and Narchi H., 2022). One such overlooked food source is mushrooms. *Agaricus bisporus*, the white button mushroom, is not only the most cultivated mushroom globally but is also becoming increasingly relevant as a sustainable and alternate protein source for a large consumer base of vegan and vegetarian individuals (Ali S.B. and Smith W., 2024). Nutritionally rich and environmentally friendly, it's often celebrated in animal conscious diets. Despite all its popularity and hype, little to no attention has been given to its potential allergenicity (Thermo Fisher Scientific, n.d.). That's surprising and concerning, because the safety of any dietary protein source must include an understanding of whether it could provoke allergic responses or not (Allergen Bureau, n.d.).

Anaphylaxis in response to edible mushrooms is uncommon but not unheard of, there have been several documented cases of anaphylactic reactions due to edible

mushrooms (Ali S.B. and Smith W., 2024) (Watson C. and Kobernick A., 2022). There have also been strong evidence and documentation about cross reactivity of allergens from edible mushrooms to airborne allergens leading to severe allergies and anaphylaxis. This makes investigation of potential allergenicity of mushrooms very important to avoid such instances.

When it comes to globally recognised allergen databases, like AllergenOnline (maintained by FARRP, University of Nebraska) or WHO/IUIS Allergen Nomenclature Homepage (maintained by WHO), the fungal kingdom is vastly underrepresented. AllergenOnline lists over 2,300 allergenic proteins, but only 128 are from fungi, and just one—Cop c 1 from *Coprinus comatus*—is a known food allergen (Goodman et al., 2016). Most fungal entries are that of fungal aeroallergens. This skewed distribution creates a major bottleneck in allergen prediction pipelines, which rely mostly on sequence similarity to known allergens, because lack of representation of fungi in the databases recognised globally result in absence of a proper reference set for allergen prediction in all underexplored species of fungi and without a proper reference set, the accuracy of *in silico* allergen prediction is questionable at best (Goodman et al., 2016; Mari et al., 2009).

Traditional *in silico* allergen prediction pipelines are based off of BLAST and FASTA, and are guided by FAO/WHO criteria, they rely mainly on sequence identity typically looking for more than 35% similarity over an 80-amino acid window or short identical sequences of 6–8 residues (Soeria-Atmadja et al., 2006). While these benchmarks are useful, they aren't nearly enough. Proteins that look similar may behave differently spatially and immunologically. Some that don't meet the threshold might still provoke estranged immune responses, especially while considering aspects like structure, solubility, or resistance to digestion. In practice, these tools can be too unreliable i.e. labeling proteins as risky or not without capturing the nuances in between (Soeria-Atmadja et al., 2006).

That is where a peptidomics-based strategy can be introduced and experimented with. Short peptides—dipeptides, tripeptides, tetrapeptides are the fundamental units that can build up to form IgE-binding epitopes. These motifs might seem trivial in isolation, but when they occur repeatedly or in overlapping regions, they can serve as markers for potential allergenicity. Studies like those by Ivanciuc et al. (2009) have shown that allergen-specific motifs often align with known

epitope regions, while Radauer et al. (2008) noted that allergens are usually found in just a few protein families, performing fairly specific functions. That consistency can be leveraged.
In parallel, I also looked at the amino acid composition. Proteins rich in certain residues like glycine, proline, or hydrophobic amino acids can behave differently in terms of structure and immune accessibility. For example, Cupin allergens are often more abundant in methionine, proline, and histidine, while Bet v 1 homologs tend to have more glycine, which could influence folding (Costa et al., 2021). These patterns are referred to as Relative Amino Acid Usage (RAAU).
Solubility of proteins is another often overlooked yet crucial property of proteins. Insoluble proteins may be more resistant to degradation or breakdown, which could make them persist longer in the body, and more likely to elicit immune responses. Predictive tools like SoluProt can be used to estimate solubility based on physicochemical properties (Smolnikov et al., 2023).

While machine learning models like AllerTrans (Sarlakifar et al., 2024) and AllerStat (Goto et al., 2023) have introduced exciting possibilities in allergen prediction, they rely heavily on presence of comprehensive training data i.e. reference set. That is where fungal proteins get left out again. With very few fungal allergens in existing databases, these models don't have the appropriate training dataset which is needed to identify patterns affecting allergenicity in fungi effectively.

All of this led to the rationale for this thesis: to identify and analyse putatively allergenic proteins from the proteome of *Agaricus bisporus* and use that data to develop a more nuanced and data-driven understanding of potentially allergenic proteins in edible fungi. I began by identifying putative allergens through traditional bioinformatics approaches using BLAST-based alignment against the AllergenOnline dataset, applying a >50% identity threshold and using other *in silico* tools for allergen prediction (Goodman et al., 2016). From there, I compared the short peptide composition—di-, tri-, and tetrapeptides of the putative allergens against the full *A. bisporus* proteome and the known allergens in AllergenOnline (Ivanciuc et al., 2009; Radauer et al., 2008).

To strengthen and validate the findings, I also analyzed amino acid and small peptide composition aiming to find compositional trends that could differentiate allergens from non-allergens (Costa et al., 2021). On top of that, I also mapped overrepresented small peptides on putative allergens to gather more information

about potentially allergenic motifs or epitopes which could somehow correlate to allergenicity. This integrated approach offers a more complete picture of what makes a protein allergenic, especially in the context of underrepresented organisms like edible fungi.

In summary, this thesis doesn't aim to reinvent allergen prediction rather it aims to fill the blind spots by focusing on fungal proteins and combining different analytical layers, it provides a more informed framework for identifying and understanding allergenic potential in novel foods (Goodman et al., 2016; Costa et al., 2021; Ivanciuc et al., 2009). This work also reinforces the idea that allergenicity isn't just a yes-or-no question rather it's a complex spectrum that requires both biological insight and computational creativity to interpret (Radauer et al., 2008).

CHAPTER 2 Review of Literature

To understand allergenicity acknowledging that not all proteins are considered equal in the eyes of the immune system is important. Some provoke no immune response at all, others trigger some mild symptoms, and a select few can cause life-threatening allergic reactions or anaphylaxis (Simon-Nobbe B. et al., 2007). In food allergy research, the majority of attention has historically been given to plant and animal derived foods, leaving fungal proteins, especially edible mushrooms, largely out of the spotlight (Crameri R. et al., 2014). This review of literature sets the stage by exploring what we know about food allergens, existing detection strategies, fungal allergen representation, and emerging computational methods that go beyond sequence identity (Crameri R. et al., 2014).

2.1 Defining Food Allergy and Allergenicity

Food allergy is defined as an unusually exaggerated immune response triggered by exposure to specific proteins called allergens. IgE-mediated allergies are the most commonly researched, they occur when the immune system produces IgE antibodies that bind to allergens and consider them as antigens, resulting in the release of histamine and other inflammatory mediators. Allergenic symptoms can include symptoms like hives, gastrointestinal distress, or in severe cases, anaphylaxis (Sicherer & Sampson, 2014).

2.2 Mechanism of Food Allergies mediated by IgE antibodies

An allergic reaction is triggered when the body encounters the same allergen protein again after being sensitized. During this re-exposure, the allergen binds to IgE antibodies that are already attached to the surface of mast cells or basophils. This cross-linking causes the cells to degranulate, releasing histamine along with other inflammatory mediators (Wangorsch et al., 2009).

The immune system is designed to protect us from pathogens like bacteria and viruses. But in the case of allergies, it overreacts to harmless substances—known as allergens—which can come from everyday sources like food, environmental particles, medications, or latex.

IgE antibodies are central to most allergic reactions. When someone is exposed to an allergen for the first time, the body doesn't show symptoms but starts producing IgE antibodies under the influence of type 2 helper T cells (Th2). These IgE molecules then bind to high-affinity receptors on mast cells and basophils. Upon re-exposure, the allergen connects to adjacent IgE molecules, bringing the receptors together and triggering a signaling cascade. This leads to the release of histamine and other chemical messengers (Wangorsch et al., 2009).

The result is a rapid immune response—usually within minutes—that causes blood vessels to dilate, increases vascular permeability, stimulates mucus secretion, and activates sensory nerves. These physiological changes are what give rise to the typical allergy symptoms which can be mild to severe.

2.3 Traditional Allergen Prediction

The identification of allergenic proteins has long relied on classical sequence-based methods, many of which were first standardized by regulatory authorities such as the FAO/WHO and Codex Alimentarius. These traditional approaches primarily depend on sequence similarity thresholds and peptide identity rules to flag proteins with potential allergenicity. For instance, according to the FAO/WHO 2001 guidelines, a protein is considered potentially allergenic if it shares an identical 6-mer peptide or over 35% identity within an 80-amino acid sliding window with a known allergen (FAO/WHO, 2001; Goodman et al., 2016).

These strategies, while foundational, often lack specificity—a key limitation that has led to high false-positive rates, especially in proteins with shared evolutionary origins or conserved domains (Goodman et al., 2016; Soeria-Atmadja et al., 2006).

The 6-mer identity rule remains an important early screening method and is still utilized by tools like AllerCatPro, but not in isolation (Maurer-Stroh et al., 2019). Tools now typically combine sequence identity searches with additional metrics like 3D structural similarity and repeat motifs, such as gluten-like Q-repeats, to enhance prediction accuracy (Goodman et al., 2016; Maurer-Stroh et al., 2019).

The 8-mer match rule was introduced later as a more stringent criterion and continues to be a regulatory mainstay, although it's utility in truly capturing allergenic potential remains debated within the scientific community (Goodman et al., 2016).

Another widely used method involves comparing entire protein sequences to known allergens using global alignment tools like FASTA and BLAST. Databases such as AllergenOnline, maintained by the University of Nebraska, offer curated allergen lists and integrate multiple approaches—full FASTA alignment, 8-mer matching, and 80-mer sliding window comparisons—for in silico allergen assessment (Goodman et al., 2016). While useful for proteins with clear homologs, these approaches fall short in predicting novel or engineered proteins with no close matches in allergen databases (Soeria-Atmadja et al., 2006).

To address this, more refined tools like AlgPred 2.0 and iAller have emerged. AlgPred 2.0 uses a hybrid model that incorporates sequence-based alignment, motif searches, IgE epitope mapping, and machine learning algorithms like SVMs on amino acid and dipeptide compositions (Sharma et al., 2021). Similarly, iAller relies on random forest classifiers using features such as amino acid composition (AAC), dipeptide composition (DPC), and compositionally skews amino acid pair (CSKAAP), achieving up to 91.4% accuracy in validation studies (Dai et al., 2022).

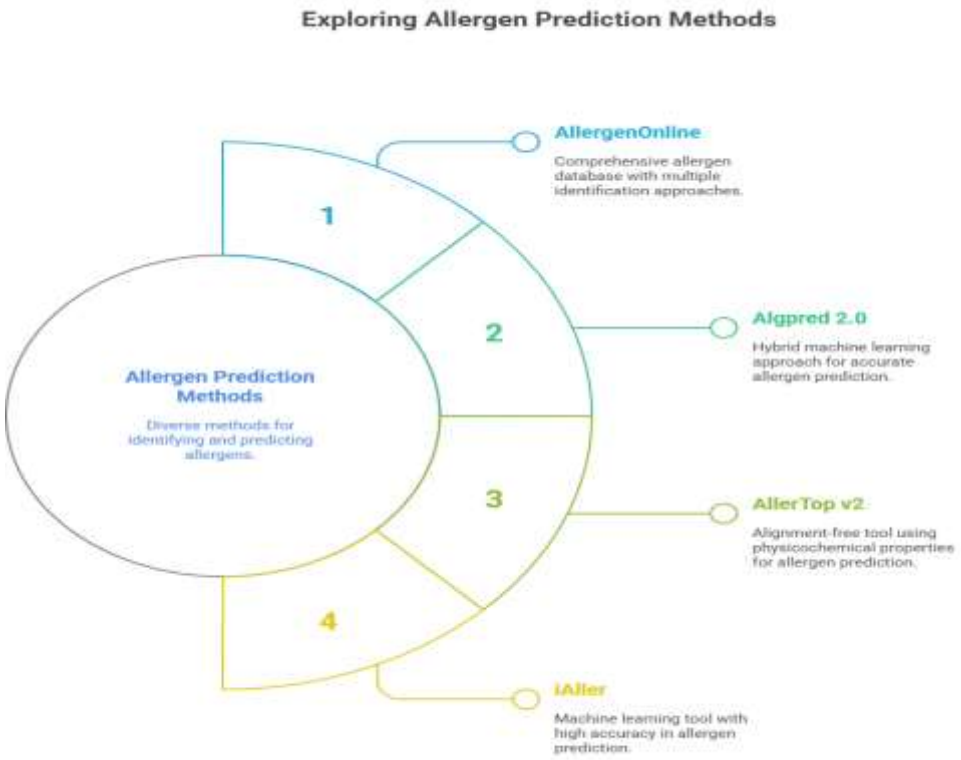


Fig 1 – Popular Allergen Prediction Methods

For proteins lacking homologs, alignment-free methods like AllerTop v2 are especially useful. AllerTop applies auto-cross covariance transformation of protein sequences based on physicochemical properties to distinguish allergens from non-allergens (Dimitrov et al., 2014). These tools are valuable for uncovering hidden patterns beyond direct sequence similarity—particularly for lesser-studied allergens, such as those derived from fungal sources.

2.4 The Limitations of Sequence-Based Allergen Prediction

Conventional allergen prediction models rely heavily on sequence alignment tools such as BLAST or FASTA. The FAO/WHO guidelines suggest that proteins showing >35% identity over 80 amino acids, or short exact matches of 6–8 amino acids, should be considered potential allergens (Soeria-Atmadja et al., 2006). However, sequence similarity alone does not always translate to immunological similarity. For instance, soybean and peanut allergens can have high sequence identity but vastly different immune outcomes (Frankild et al., 2008). Such nuances expose the limitations of alignment-based screening, particularly for underrepresented sources like fungi.

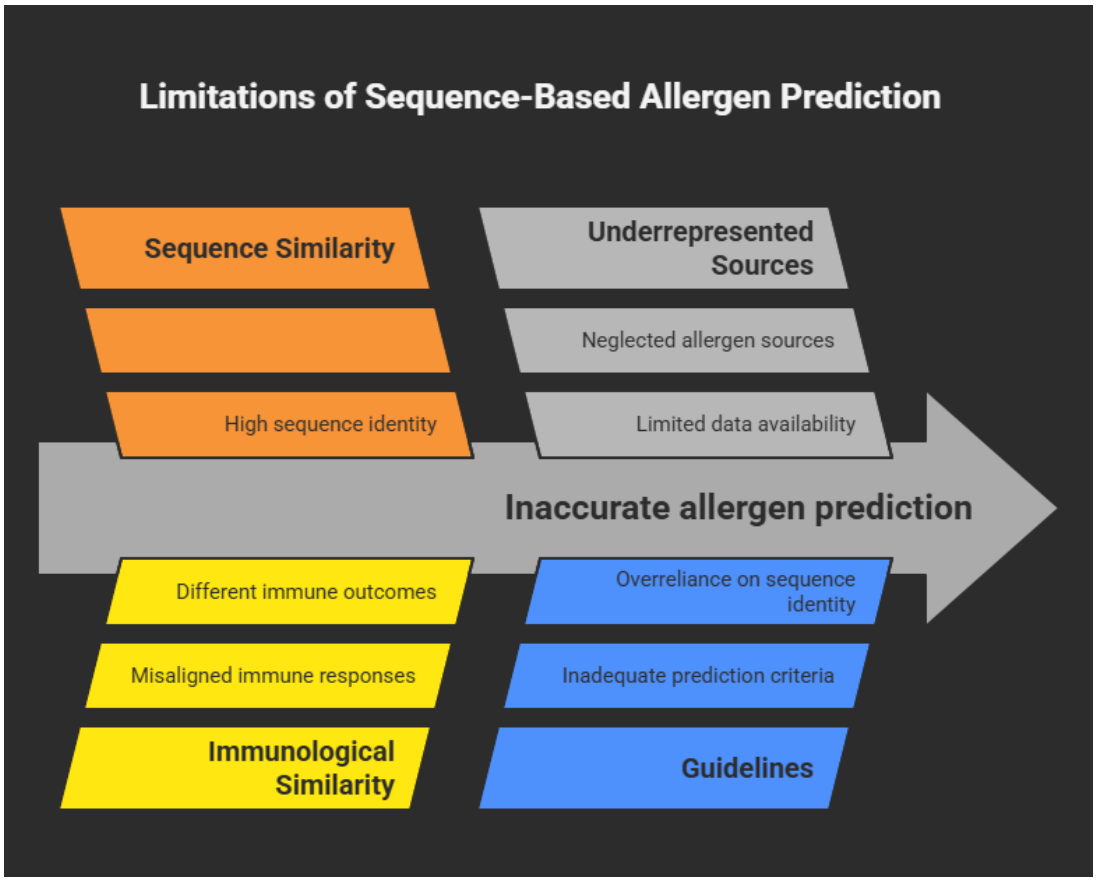


Fig 2 – Limitations of Sequence-Based Allergen Prediction

2.3 Fungal Allergens: A Missing Piece in Databases

Major databases such as AllergenOnline, Allergome, and WHO/IUIS maintain curated allergen datasets. Despite their comprehensive nature, fungal allergens remain underrepresented. As of recent updates, only 128 fungal proteins are listed in AllergenOnline—mostly aeroallergens, with only a single food-based fungal allergen (Cop c 1) from Coprinus comatus (Goodman et al., 2016). This data scarcity directly impacts the effectiveness of allergenicity prediction models when applied to fungal proteomes.

2.4 Peptide Motifs as Emerging Allergen Markers

Several studies have explored the utility of short peptide motifs—dipeptides, tripeptides, and tetrapeptides—in identifying potential allergenic proteins. These motifs can overlap with IgE-binding epitopes or represent digestion-resistant fragments that may act as antigenic hotspots (Ivanciuc et al., 2009). The concept is grounded in the idea that allergens are not defined by long conserved domains alone but also by the presence of enriched, recurring short motifs that interact with the immune system.

2.5 Protein Family and Functional Bias Among Allergens

Radauer et al. (2008) showed that the majority of known allergens belong to a limited number of protein families with shared structural and functional traits. These include families like Cupins, Bet v 1 homologs, and Tropomyosins—proteins often associated with high stability, solubility resistance, or immune accessibility. This clustering reinforces the idea that allergenicity is more predictable within certain molecular frameworks.

2.6 Amino Acid Composition and Allergen Stability

Beyond motifs, amino acid composition also influences a protein's allergenic potential. Certain amino acids contribute to thermal stability, resistance to enzymatic degradation, or immune exposure. For instance, glycine-rich regions may enhance folding flexibility, while proline or histidine may stabilize protein structures in Cupin family allergens (Costa et al., 2021). Tools like Relative Amino Acid Usage (RAAU) allow quantitative profiling of these traits.

2.7 Solubility and Allergen Persistence

Solubility is another underappreciated trait. Proteins with lower solubility tend to form aggregates, resist breakdown, and stay in the system longer—potentially enhancing their allergenic risk. SoluProt is one of the few tools designed to predict solubility from sequence data, and it has shown promise in allergen-focused applications (Smolnikov et al., 2023). Studies have observed lower predicted solubility in Cupin and CAP family allergens compared to their non-allergenic counterparts.

2.8 AI and Deep Learning Approaches: The Fungal Gap

With the rise of deep learning, tools like AllerTrans (Sarlakifar et al., 2024) and allerStat (Goto et al., 2023) offer high-throughput allergen prediction. They detect statistically significant sequence patterns beyond what BLAST or FASTA can see. However, these models depend on large, high-quality datasets, and when fungal entries are sparse, performance drops. This makes them less reliable for novel or understudied taxa like *Agaricus bisporus*.

2.9 Need for Integrated and Fungi-Specific Frameworks

Together, these studies underscore that allergenicity cannot be reliably assessed through sequence similarity alone (Frankild et al., 2008; Soeria-Atmadja et al., 2006). A more integrated strategy that includes motif frequency, amino acid usage, protein family membership, and predicted solubility is necessary—especially for fungi (Costa et al., 2021; Radauer et al., 2008; Smolnikov et al., 2023). The approach taken in this thesis builds on this need by exploring peptidomic patterns, compositional trends, and BLAST-derived similarity to assess allergenic potential in A. bisporus proteins.

2.10 RAAU and Solubility Differences in Major Allergen Families

The role of RAAU (Relative Amino Acid Usage) in differentiating allergenic from non-allergenic proteins has been observed in multiple studies, including in this thesis. For instance, in allergen families like Cupins and Tropomyosins, RAAU patterns reveal a higher abundance of proline, histidine, methionine, and glycine (Costa et al., 2021). These patterns contribute to protein rigidity and

folding stability, which enhance immune system visibility and persistence in the body. In contrast, non-allergens within the same families show different RAAU profiles—often with more polar and flexible residues, making them more susceptible to degradation and less likely to provoke immune responses (Costa et al., 2021; Radauer et al., 2008).

2.11 Motif Discovery in Allergen Families

Motif analysis using multiple sequence alignment (MSA) revealed the presence of conserved short motifs (>70% identity) unique to allergens in each family. These motifs are hypothesized to contribute to allergenic stability and may overlap with known B-cell epitopes, as supported by literature (Ivanciuc et al., 2009). Mapping these motifs is valuable for identifying immune-relevant fragments that standard alignment tools might miss.

2.12 Peptidomic Enrichment in Putative Allergens

Peptidomics, a branch of proteomics focused on the systematic study of endogenous peptides, has emerged as a valuable tool in allergen research. Unlike traditional protein-level analyses, peptidomics allows a fine-grained investigation of short peptide motifs that may contribute to immunogenicity. Several studies have demonstrated that allergenic proteins exhibit non-random distributions of amino acid motifs, particularly within epitope-rich regions, which can influence their binding affinity to human leukocyte antigen (HLA) molecules and subsequent T-cell activation (Dall'Antonia et al., 2019; Chruszcz et al., 2018).

In the context of food and fungal allergens, recurring short peptide motifs such as dipeptides, tripeptides, and tetrapeptides have been shown to recur disproportionately in allergenic proteins compared to non-allergenic counterparts. These motifs often reflect biochemical properties such as hydrophobicity, charge, and secondary structure propensity that affect protein stability and recognition by the immune system (Ivanciuc et al., 2009; Bruni et al., 2020). Recent work has also highlighted how linear B-cell epitopes, which are contiguous peptide stretches accessible to antibodies, tend to be enriched in specific amino acid compositions and sequence patterns (Vita et al., 2019). Such findings support the hypothesis that allergenicity is not only a property of entire proteins but can also be influenced by overrepresented sequence motifs.

Despite these insights, limited studies have systematically explored the peptidomic enrichment within putative allergens of non-model organisms such as Agaricus bisporus var. burnetti. This is a critical gap given the growing consumption and industrial use of edible fungi, some of which exhibit hidden or underreported allergenic potential (Crameri et al., 2014; Simon-Nobbe et al., 2007). The application of peptidomic analysis to putative allergens allows researchers to identify peptide signatures that are disproportionately enriched compared to the background proteome, offering a molecular lens into potential immunogenic hotspots. Such comparative studies, when linked with known allergen databases and structural epitope predictors, can enhance the prediction and validation of novel allergens.

Furthermore, peptide enrichment analysis supports the development of machine learning models for allergen prediction. Features derived from short peptide abundance—such as frequency of specific tetrapeptides—have shown potential as discriminative variables in classifying allergenic versus non-allergenic sequences (Dimitrov et al., 2014; Saha & Raghava, 2006). Incorporating peptidomic insights into such models could improve their precision, especially in underrepresented taxa.

In summary, peptidomic enrichment analysis serves as a promising strategy to uncover sequence-level determinants of allergenicity. By investigating the differential abundance of short peptide motifs in the putative allergens of *Agaricus bisporus* var. *burnetti*, this study seeks to bridge the knowledge gap between known allergen patterns and emerging allergenic threats in edible fungi.

2.13 Peptide Signature Overlap with Known IgE Epitopes

Allergenicity is fundamentally driven by an individual's immunoglobulin E (IgE)-mediated response to specific epitopes within proteins. These epitopes, which are either linear (continuous peptide stretches) or conformational (discontinuous but spatially proximate residues), are recognized by IgE antibodies during allergic sensitization and subsequent reactions (Pomes, 2010). Among these, linear IgE-binding epitopes have garnered significant attention due to their compatibility with in silico identification strategies, especially through sequence-based motif comparisons (Aalberse, 2000; Vita et al., 2019).

The Immune Epitope Database (IEDB) and other curated allergen databases have greatly expanded access to experimentally validated IgE epitopes from food,

pollen, fungal, and insect sources (Vita et al., 2019). These databases not only document epitope sequences but also provide associated immunological context, including host response data and binding affinities. Several studies have shown that allergens tend to be enriched in peptide motifs that partially or wholly overlap with known IgE epitopes, suggesting a conserved sequence-level "signature" of allergenicity (Ivanciuc et al., 2009; Bruni et al., 2020).

The overlap between peptidomic signatures and known IgE epitopes is especially relevant in the identification of putative or novel allergens. Peptides found to be overrepresented in allergenic proteins—such as particular di-, tri-, or tetrapeptides—often mirror segments of validated IgE epitopes, indicating their potential role in triggering immune responses (Dimitrov et al., 2014). Such overlaps serve as molecular fingerprints for IgE-binding potential and can be used to prioritize candidate allergens for experimental validation.

In fungal allergens, including those from Agaricus bisporus and other basidiomycetes, emerging research has begun mapping short peptide motifs to known IgE-binding epitopes, revealing that even proteins not previously recognized as allergens may harbor immunologically relevant sequences (Crameri et al., 2014; Simon-Nobbe et al., 2007). This reinforces the importance of systematically comparing enriched peptide motifs in putative allergens to established IgE epitope databases, as such matches can signal hidden sensitization risks.

Furthermore, the development of allergen prediction algorithms such as AlgPred, AllerTOP, and AllerCatPro has increasingly relied on epitope overlap data as a critical feature for allergen identification (Saha & Raghava, 2006; Maurer-Stroh et al., 2019). These tools leverage similarity thresholds—often defined by percent identity or physicochemical profiles—to identify potential cross-reactivity and classify proteins based on epitope homology. High-confidence matches between enriched peptides and known IgE epitopes support a more robust and biologically plausible prediction of allergenicity.

In this context, the current study investigates the degree of overlap between enriched peptide signatures in putative Agaricus bisporus var. burnetti allergens and experimentally validated IgE epitopes from IEDB and related sources. This comparison aims to elucidate whether allergen-specific peptides coincide with established immunoreactive regions, thereby enhancing confidence in their

allergenic potential and providing a deeper understanding of peptide-level determinants of IgE binding.

2.14 Cross-Reactivity and Peptide Mimicry

Cross-reactivity is a well-established phenomenon in allergy science, wherein an individual's immune system recognizes and reacts to structurally or sequentially similar epitopes found in different allergen sources. This often occurs due to shared peptide motifs or structural homology among unrelated proteins, leading to unintended IgE binding and clinical symptoms upon exposure to seemingly unrelated allergens (Aalberse, 2000; Radauer & Breiteneder, 2007). Such immune cross-reactivity poses a significant challenge in allergy diagnosis, food labeling, and risk assessment.

A key molecular mechanism underlying cross-reactivity is peptide mimicry, where short peptide sequences in non-homologous proteins resemble IgE-binding epitopes from known allergens. Even in the absence of overall sequence identity or phylogenetic relatedness, these mimetic motifs can trigger immune responses due to structural convergence or local sequence similarity (Chruszcz et al., 2018; Schein et al., 2021). This is particularly relevant in allergen prediction models where a single allergenic determinant may be conserved as a short motif (e.g., tripeptide or tetrapeptide) across diverse species.

In fungal allergens—including species such as Alternaria alternata, Aspergillus fumigatus, and Cladosporium herbarum—extensive cross-reactivity has been documented with plant, food, and insect allergens (Crameri et al., 2014). However, Agaricus bisporus, especially the underexplored var. burnetti, remains poorly characterized in this context. The increasing culinary and pharmaceutical use of this edible mushroom underscores the need to identify shared allergenic motifs with known allergens, particularly from other fungi, plants, and environmental sources.

The concept of peptide signature mimicry becomes especially important in your study, which explores short peptide motifs (di-, tri-, and tetrapeptides) enriched in putative allergens of Agaricus bisporus var. burnetti. These enriched motifs are then compared against validated IgE-binding epitopes from databases like IEDB. Early findings in other domains have shown that allergenic proteins often share

conserved pentapeptide or tetrapeptide motifs that mimic known epitopes and mediate IgE cross-reactivity (Ivanciuc et al., 2009; Bruni et al., 2020). Identifying such mimetic patterns in fungal proteins may provide molecular-level insights into hidden cross-reactive risks and guide future experimental validation.

Furthermore, peptide mimicry plays a role in false negatives in allergenicity prediction tools, which often rely on global sequence similarity. A non-allergenic classification based on low identity may miss local epitope mimicry that can still provoke an immune response, especially in atopic individuals sensitized to structurally similar epitopes elsewhere (Dimitrov et al., 2014; Goodman et al., 2016). This limitation validates the use of peptide-level comparisons as a complementary layer of analysis.

Overall, the detection of peptide mimicry within enriched sequences of A. bisporus var. burnetti putative allergens is a promising avenue for uncovering novel cross-reactive epitopes and strengthening allergenicity prediction beyond conventional sequence alignment. It aligns with a precision-allergy approach—moving from whole-protein assessments toward motif-centric models of immune recognition.

Practical Gaps in Existing Prediction Tools

Over the past two decades, numerous computational tools have been developed to predict protein allergenicity based on sequence similarity, structural features, and epitope content. Prominent examples include AlgPred (Saha & Raghava, 2006), AllerTOP (Dimitrov et al., 2014), AllergenFP (Dimitrov et al., 2013), AllerCatPro (Maurer-Stroh et al., 2019), and PREAL (Dimitrov et al., 2022). These tools aim to facilitate early allergen screening during food, pharmaceutical, and biotech product development. However, despite advances in machine learning, motif-based searching, and immunoinformatics, significant practical gaps remain that limit the reliability, applicability, and real-world utility of these models.

One of the primary limitations is the over-reliance on sequence similarity to known allergens. Many tools use threshold-based BLAST or FASTA alignments (e.g., >35% identity over 80 amino acids) to predict cross-reactivity. However, this approach overlooks de novo allergens that may not share high sequence similarity but still elicit IgE-mediated responses due to structural mimicry or the presence of conformational epitopes (Aalberse, 2000; Goodman et al., 2016). As a result, non-homologous but allergenic proteins may be falsely predicted as non-allergenic.

Secondly, epitope prediction remains imprecise. While linear IgE epitope databases like IEDB provide valuable curated data, conformational epitopes—which make up the majority of IgE-binding sites—require high-resolution structural data and are not adequately handled by most tools (Pomes, 2010; Chruszcz et al., 2018). As a result, many algorithms either ignore conformational epitopes or rely on predicted 3D models that introduce additional uncertainty.

Another practical gap is the lack of context-specific prediction, such as organism-specific allergenicity, environmental exposure, digestion stability, or post-translational modifications. For instance, fungal allergens are underrepresented in training datasets, leading to reduced performance for species like Agaricus bisporus (Simon-Nobbe et al., 2007; Crameri et al., 2014). Furthermore, few tools

account for digestion-resistance or processing stability—critical determinants of allergenicity in food proteins (Taylor & Hefle, 2001).

Tool usability and transparency are also major concerns. Many online platforms do not clearly explain their prediction logic or confidence scores, which limits regulatory acceptance and scientific reproducibility. Moreover, the binary classification approach (allergen/non-allergen) adopted by most tools fails to capture the spectrum of allergenic potential, missing subtler risk levels in putative or cross-reactive proteins.

Lastly, existing tools rarely integrate peptidomic enrichment data or empirical frequency patterns of motifs known to drive immunogenicity. Despite mounting evidence that short peptide signatures play a crucial role in allergen recognition (Ivanciuc et al., 2009; Bruni et al., 2020), most predictors do not incorporate this layer of information into their classification logic. This gap provides a compelling rationale for the present study, which investigates peptide signature enrichment and epitope overlap to enhance allergen prediction in edible fungi.

Table 1 : Summary of Allergen Databases and Computational Tools Studied

Name	Link	Description / Use Case	Reference
AlgPred	crdd.osdd.net/raghava/algpred	Uses SVM and motif features for allergenicity prediction	Saha & Raghava, 2006
AlgPred 2.0	webs.iiitd.edu.in/raghava/algpred2	Hybrid model with alignment, IgE epitopes, AAC/DPC, and ML classifiers	Sharma et al., 2021
AllerTOP v2	ddg-pharmfac.net/AllerTOP/	Alignment-free tool using physicochemical property-based ACC transformation	Dimitrov et al., 2014
AllergenFP	ddg-pharmfac.net/AllergenFP/method.html	Predicts allergenicity using descriptor fingerprints and Tanimoto similarity	Dimitrov et al., 2014

AllerCatPro	allercatpro.bii.a-star.edu.sg	Uses sequence, 3D structure, and motif matching to predict allergenic potential	Maurer-Stroh et al., 2019
PREAL	Not publicly available	Ensemble learning model for allergenicity prediction	Dimitrov et al., 2022
iAller	Not publicly available	Random forest-based classifier using AAC, DPC, CSKAAP for allergen prediction	Dai et al., 2022
AllergenOnline	www.allergenonline.org	Curated allergen database for regulatory risk assessment; FASTA and epitope tools included	Goodman et al., 2016
AllerStat	Not public	Deep learning model that finds statistically enriched patterns in allergen sequences	Goto et al., 2023
AllerTrans	Preprint only	Transformer-based deep learning tool for allergen classification	Sarlakifar et al., 2024
IEDB (Immune Epitope Database)	www.iedb.org	Database of validated B-cell and T-cell epitopes; supports mapping IgE epitopes	Vita et al., 2019
SoluProt	loschmidt.chemi.muni.cz/soluprot	Predicts solubility of proteins; useful in persistence/allergenicity assessments	Smolnikov et al., 2023
Allergen Nomenclature (WHO/IUIS)	www.allergen.org	Official repository of systematic allergen names and classifications	Pomés et al., 2018
Allergome	www.allergome.org	Repository of molecular and clinical data for allergens across IgE-mediated diseases	Mari et al., 2006
AllFam	www.meduniwien.ac.at/allfam/	Classifies allergens into protein families with functional/biochemical annotations	Radauer et al., 2008

SDAP (Structural DB of Allergens)	fermi.utmb.edu/SDAP	Structural data on allergenic proteins; evaluates FAO/WHO rules; includes 3D insights and cross-reactivity info	Schein et al., 2022
Allermatch	www.allermatch.org	Checks allergenicity using FAO/WHO rules (6-mer/80-mer/8-mer identity criteria)	Fiers et al., 2004
WEGO (Gene Ontology Plotter)	www.wego.genomics.cn	GO annotation tool; often used for allergen-related gene function classification	Ye et al., 2018
I-TASSER	zhanglab.ccmb.med.umich.edu/I-TASSER	Predicts protein 3D structure; used for conformational epitope mapping	Yang et al., 2015
ClusPro 2.0	www.cluspro.bu.edu	Docking tool for simulating protein-protein interactions (e.g., allergen-antibody docking)	Desta et al., 2020

CHAPTER 3 Materials and Methods

This study employed a multi-step computational workflow to identify and evaluate putative allergens from the proteome of *Agaricus bisporus var. burnetti*. The approach included sequence retrieval, homology-based screening, allergenicity prediction, protein family classification, epitope prediction, structural modelling, and molecular docking. A comprehensive summary of each method is presented below (Goodman et al., 2016; Sharma et al., 2021; Dimitrov et al., 2014; Ivanciuc et al., 2009).

3.1 Sequence Retrieval and Dataset Preparation

The complete proteome of *Agaricus bisporus* var. *burnetti* (strain H119_P4), consisting of 11,675 proteins, was downloaded from the UniProt database in FASTA format (UniProt Consortium, 2023). A curated allergen dataset comprising 2,334 experimentally validated allergens was obtained from AllergenOnline for comparative analysis.

3.2 Allergen Identification Using Sequence-Based Tools

All proteins from the mushroom proteome were screened for allergenicity using BLASTp against the AllergenOnline dataset. Proteins with >50% identity and query coverage were retained. Further validation was performed with FASTA 36 (AllergenOnline's online tool) using both full-length and 80-amino-acid sliding window alignment. Sequences with >35% identity in the 80-aa window or >50% global identity were shortlisted (FAO/WHO, 2001; Goodman et al., 2016).

The shortlisted proteins were then assessed by three allergenicity prediction tools:

AllerTOP v2.0: Uses ACC-transformed descriptors (Dimitrov et al., 2014).

AllergenFP: Uses Tanimoto coefficient-based binary classification (Dimitrov et al., 2014).

AlgPred: Uses motif, IgE epitope, and SVM-based methods (Sharma et al., 2021).

Only those proteins predicted as "probable allergens" by at least two of these tools were retained for downstream analysis.

3.3 Protein Family Classification and Domain Annotation

All retained putative allergens were classified into protein families using UniProt and PFam (Jones et al., 2014). Functional annotations and protein domains were recorded. The prevalence of each protein family was tabulated.

3.4 Gene Ontology Annotation

GO annotations for molecular function, biological process, and cellular component were extracted using QuickGO and UniProt (UniProt Consortium, 2023).

3.5 Epitope Prediction

To predict immunogenic regions, linear B-cell epitopes were identified using the IEDB B-cell epitope prediction tool. Regions with high prediction scores were recorded.

3.6 Structural Modeling

High-confidence 3D structures were predicted for selected proteins using (Waterhouse et al., 2018): Swiss-Model: Used as a complementary tool for automated homology modeling.

3.7 Molecular Docking

Molecular docking simulations were performed using ClusPro 2.0 (Kozakov et al., 2017). Allergen models were uploaded as ligands and IgE antibodies as receptors. Top docking poses were selected based on clustering and energy scores. Visualization and analysis of interactions were carried out using PyMOL, highlighting probable allergen-antibody binding sites.

3.8 Amino Acid and Peptidomic Composition Analysis

To better understand how allergenic proteins differ from the rest of the proteome in their sequence composition, I split the dataset into two parts: (1) the complete *A. bisporus* proteome minus the 55 putative allergens, and (2) the 55 putative allergens alone. I scanned these datasets for all possible amino acids, dipeptides, tripeptides, and tetrapeptides (Ivanciuc et al., 2009; Costa et al., 2021).

For each type of peptide, I calculated the percentage abundance by dividing its total count by the total number of peptides of that type in the dataset. This gave a normalized distribution for each group of peptides, allowing me to directly compare trends between allergenic and non-allergenic proteins.

3.9 Relative Enrichment Analysis of Peptides

To make these comparisons more meaningful despite the difference in sample sizes between the full proteome and the allergen subset, I calculated the Relative Enrichment (RE) for each peptide using this formula:

RE = (% abundance in Putative Allergens) / (% abundance in Proteome)

This helped highlight peptides that were not just present in allergens—but overrepresented compared to the background proteome. I computed RE values for every dipeptide, tripeptide, and tetrapeptide. Tetrapeptides with RE values ≥100 were considered highly enriched.

These enriched tetrapeptides were then mapped back onto the FASTA sequences of the putative allergens using a sliding window approach. I specifically looked for overlapping motifs, since clusters of enriched peptides often lined up and could form immunogenic hotspots. This step helped connect enrichment trends to actual sequence locations, offering more biological context to the patterns discovered in the previous step. I used a custom python script for this mapping.

The custom python script is as follows:

```
# Define file paths
$fastaFile = "C:\Users\priya\OneDrive\Desktop\Project\Dissertation\Objective
1\Dataset\11 Putative food allergen fasta.txt"
$diFile = "C:\Users\priya\OneDrive\Desktop\Dipeptides.txt"
$triFile = "C:\Users\priya\OneDrive\Desktop\Tripeptides.txt"
$tetraFile = "C:\Users\priya\OneDrive\Desktop\Tetrapeptides.txt"
```

```
$outFile = "C:\Users\priya\OneDrive\Desktop\HighlightedPeptides.html"

# Load and clean peptide lists
$dipeptides = (Get-Content $diFile) | ForEach-Object { $_.Trim().ToUpper() } | Where-Object { $_ -ne "" }
$tripeptides = (Get-Content $triFile) | ForEach-Object { $_.Trim().ToUpper() } | Where-Object { $_ -ne "" }
$tetrapeptides = (Get-Content $tetraFile) | ForEach-Object { $_.Trim().ToUpper() } | Where-Object { $_ -ne "" }

# Load FASTA and prepare output
$fasta = Get-Content $fastaFile
$html = @"
<html>
<head>
<style>
body { font-family: Consolas, monospace; white-space: pre; }
.dipep { background-color: #cce5ff; border-radius: 10px; }
.tripep { background-color: #d4edda; border-radius: 10px; }
.tetrapep { background-color: #f8d7da; border-radius: 10px; }
</style>
</head>
<body>
<h2>Highlighted Allergen Peptides</h2>
"@

# Function to highlight peptides in a sequence
function Highlight-Peptides($seq) {
    $seq = $seq.ToUpper()
    $markers = @{}

    foreach ($pep in $tetrapeptides) {
        [regex]::Matches($seq, [regex]::Escape($pep)) | ForEach-Object {
            $markers[$_.Index] = @{ 'len' = $pep.Length; 'class' = 'tetrapep' }
        }
    }
    foreach ($pep in $tripeptides) {
        [regex]::Matches($seq, [regex]::Escape($pep)) | ForEach-Object {
            if (-not $markers.ContainsKey($_.Index)) {
                $markers[$_.Index] = @{ 'len' = $pep.Length; 'class' = 'tripep' }
            }
        }
    }
    foreach ($pep in $dipeptides) {
```

```
        [regex]::Matches($seq, [regex]::Escape($pep)) | ForEach-Object {
            if (-not $markers.ContainsKey($_.Index)) {
                $markers[$_.Index] = @{ 'len' = $pep.Length; 'class' = 'dipep' }
            }
        }
    }

    $i = 0
    $result = ""
    while ($i -lt $seq.Length) {
        if ($markers.ContainsKey($i)) {
            $info = $markers[$i]
            $substr = $seq.Substring($i, $info.len)
            $result += "<span class='$($info.class)'>$substr</span>"
            $i += $info.len
        } else {
            $result += $seq[$i]
            $i++
        }
    }
    return $result + "`n"
}

# Process FASTA
$sequence = ""
foreach ($line in $fasta) {
    if ($line.StartsWith(">")) {
        if ($sequence -ne "") {
            $html += Highlight-Peptides $sequence
            $sequence = ""
        }
        $html += "<br><b>$line</b><br>`n"
    } else {
        $sequence += $line.Trim()
    }
}
if ($sequence -ne "") { $html += Highlight-Peptides $sequence }

$html += "</body></html>"
$html | Out-File -Encoding UTF8 $outFile
Start-Process $outFile
```

Once the overlapping motifs were determined, I compared them with already existing documented allergens in AllergenOnline and AllFam database and observations were noted.

CHAPTER 4 Results

Putative Allergen Identification

To systematically identify potential allergens within the *Agaricus bisporus* proteome, a full-length FASTA and pBLAST-based sequence alignment was performed. A total of 11,675 protein sequences were screened against a curated dataset of 2,334 known allergens from the AllergenOnline database. Based on stringent filtering criteria—specifically, a minimum percent identity of >50% and strong E-value thresholds—a shortlist of 318 proteins was generated for further analysis. These selected proteins were then evaluated using three independent allergen prediction tools: AllerTOP, AlgPred, and AllerCatPro. Since each of these platforms uses a different underlying algorithm for allergenicity prediction, cross-validation across them provided a more balanced and reliable assessment. Interestingly, 55 out of the 318 proteins were predicted to be putative allergens by at least two of the three tools, reinforcing their likelihood of allergenic potential. These 55 proteins were therefore considered as high-confidence candidates and shortlisted for subsequent peptidomic and immunoinformatic characterization. This layered, multi-tool approach not only reduced false positives but also strengthened the predictive framework used to pinpoint allergen-like proteins in the dataset.

Accession numbers	AllerTop	Algpred	Evidence of allergenicity by AllerCatPro	Consensus
A0A8H7BUR1	No	No	No	No
A0A8H7BZ90	No	Allergen	Weak	Allergen
A0A8H7BZ28	No	Allergen	No	No
A0A8H7BXR6	Allergen	Allergen	Strong	Allergen
A0A8H7BZY2	Allergen	Allergen	Weak	Allergen
A0A8H7BZY5	Allergen	Allergen	Weak	Allergen
A0A8H7C022	No	No	No	No
A0A8H7C094	Allergen	No	Strong	Allergen
A0A8H7C0I0	No	No	Weak	No
A0A8H7C1I4	No	No	No	No
A0A8H7C1J1	No	No	No	No
A0A8H7C2F2	No	Allergen	No	No
A0A8H7C5L2	Allergen	No	No	No
A0A8H7C7P4	No	No	No	No
A0A8H7C0T9	No	No	No	No
A0A8H7C100	No	Allergen	No	No
A0A8H7C106	Allergen	Allergen	Weak	Allergen
A0A8H7C107	No	Allergen	Weak	Allergen
A0A8H7C141	Allergen	Allergen	Strong	Allergen
A0A8H7C158	No	No	No	No
A0A8H7C1W4	No	No	No	No
A0A8H7C140	No	No	No	No
A0A8H7C153	Allergen	No	No	No
A0A8H7C190	No	Allergen	No	No
A0A8H7C192	No	Allergen	Weak	Allergen
A0A8H7C1F8	No	No	No	No
A0A8H7C1M1	Allergen	Allergen	Weak	Allergen
A0A8H7C1P9	No	No	No	No
A0A8H7C1Q0	No	No	No	No
A0A8H7C1W6	No	No	No	No
A0A8H7C235	No	No	No	No
A0A8H7C256	No	Allergen	No	No
A0A8H7C281	No	No	No	No

A0A8H7C289	Allergen	Allergen	Weak	Allergen
A0A8H7C2R6	Allergen	Allergen	Weak	Allergen
A0A8H7C2T5	No	Allergen	No	No
A0A8H7C2T7	Allergen	Allergen	Strong	Allergen
A0A8H7C327	No	No	No	No
A0A8H7C2H9	No	No	No	No
A0A8H7C2P9	No	No	Weak	No
A0A8H7C2R1	No	Allergen	Weak	Allergen
A0A8H7C2S5	No	No	No	No
A0A8H7C2V9	No	Allergen	No	No
A0A8H7C329	No	Allergen	No	No
A0A8H7C336	No	No	No	No
A0A8H7C375	No	Allergen	No	No
A0A8H7C3A2	No	Allergen	No	No
A0A8H7C3H7	No	Allergen	No	No
A0A8H7C3N2	No	No	No	No
A0A8H7C3N6	Allergen	Allergen	No	Allergen
A0A8H7C3U0	No	Allergen	No	No
A0A8H7C412	No	Allergen	No	No
A0A8H7C416	No	Allergen	Weak	Allergen
A0A8H7C460	No	No	Weak	No
A0A8H7C4N9	No	Allergen	Weak	Allergen
A0A8H7C559	Allergen	Allergen	No	Allergen
A0A8H7C563	No	Allergen	No	No
A0A8H7C5D4	No	Allergen	Strong	Allergen
A0A8H7C5H5	No	Allergen	No	No
A0A8H7C5I9	No	No	No	No
A0A8H7C461	No	Allergen	No	No
A0A8H7C471	No	No	No	No
A0A8H7C487	No	Allergen	No	No
A0A8H7C4E8	No	Allergen	Weak	Allergen
A0A8H7C4G5	Allergen	No	No	No
A0A8H7C4K1	No	Allergen	Weak	Allergen
A0A8H7C4L7	No	No	No	No
A0A8H7C4U8	No	No	No	No
A0A8H7C557	No	No	No	No
A0A8H7C5A3	No	No	No	No
A0A8H7C5B4	No	No	No	No
A0A8H7C5B7	No	No	No	No
A0A8H7C5D0	No	No	No	No
A0A8H7C5D8	Allergen	No	No	No
A0A8H7C5F7	No	No	No	No
A0A8H7C5G1	No	No	No	No

A0A8H7C5H1	Allergen	Allergen	Weak	Allergen
A0A8H7C5K5	No	No	No	No
A0A8H7C5K9	No	Allergen	Strong	Allergen
A0A8H7C633	Allergen	No	No	No
A0A8H7C5S7	No	No	No	No
A0A8H7C5U1	No	Allergen	No	No
A0A8H7C5W1	No	No	No	No
A0A8H7C5Z8	No	No	No	No
A0A8H7C5Z9	No	No	No	No
A0A8H7KM45	No	No	No	No
A0A8H7EWP1	No	No	No	No
A0A8H7EWT7	No	No	No	No
A0A8H7EWR9	No	No	No	No
A0A8H7EWS4	No	No	No	No
A0A8H7EWS8	No	No	No	No
A0A8H7EWS9	No	No	No	No
A0A8H7C8Z7	No	No	No	No
A0A8H7C910	No	No	No	No
A0A8H7C993	No	Allergen	No	No
A0A8H7C9F7	No	Allergen	No	No
A0A8H7C8P2	No	No	No	No
A0A8H7C8Q6	No	No	No	No
A0A8H7C8S5	No	No	No	No
A0A8H7C8V3	No	No	No	No
A0A8H7C9I9	No	No	No	No
A0A8H7C9W0	Allergen	No	No	No
A0A8H7C9X9	No	No	No	No
A0A8H7C851	No	Allergen	No	No
A0A8H7C855	No	No	No	No
A0A8H7C8B6	No	Allergen	Weak	Allergen
A0A8H7C8H7	No	Allergen	No	No
A0A8H7C8C3	No	No	No	No
A0A8H7C7D8	No	No	No	No
A0A8H7C7E6	No	No	No	No
A0A8H7C7F1	Allergen	No	No	No
A0A8H7C7G4	No	No	No	No
A0A8H7C7H6	No	No	No	No
A0A8H7C7K1	No	Allergen	No	No
A0A8H7C7R5	No	No	No	No
A0A8H7C7X9	No	No	No	No
A0A8H7C802	No	No	No	No
A0A8H7F7I7	No	No	No	No
A0A8H7F7R1	No	No	No	No

A0A8H7F7S3	No	No	No	No
A0A8H7F7T6	No	No	No	No
A0A8H7F831	No	No	No	No
A0A8H7F842	No	Allergen	No	No
A0A8H7CC30	Allergen	Allergen	Weak	Allergen
A0A8H7CCE3	No	No	No	No
A0A8H7CDF2	No	No	No	No
A0A8H7ETL0	No	Allergen	Strong	Allergen
A0A8H7EUP0	No	No	No	No
A0A8H7EV58	Allergen	No	No	No
A0A8H7EV86	No	No	No	No
A0A8H7EVL5	No	No	No	No
A0A8H7EVT7	No	Allergen	Strong	Allergen
A0A8H7EVV8	No	No	No	No
A0A8H7EVX1	No	No	No	No
A0A8H7EVX2	Allergen	Allergen	Strong	Allergen
A0A8H7EW79	No	Allergen	No	No
A0A8H7EW90	No	No	No	No
A0A8H7EW94	No	Allergen	Strong	Allergen
A0A8H7EWG0	No	No	No	No
A0A8H7EWI6	Allergen	Allergen	Weak	Allergen
A0A8H7EWJ4	No	No	No	No
A0A8H7KJI9	No	No	No	No
A0A8H7KJK2	No	Allergen	Weak	Allergen
A0A8H7KJS6	No	No	No	No
A0A8H7KK27	No	No	No	No
A0A8H7KK68	No	No	No	No
A0A8H7KK76	No	No	Weak	No
A0A8H7KKC4	No	Allergen	No	No
A0A8H7KKC9	No	No	No	No
A0A8H7EWY0	No	No	Strong	No
A0A8H7EX43	No	Allergen	No	No
A0A8H7EX56	No	No	No	No
A0A8H7EXM3	No	No	No	No
A0A8H7EXQ5	No	No	No	No
A0A8H7EXW0	No	No	No	No
A0A8H7EXW2	No	No	No	No
A0A8H7EXW6	No	No	No	No
A0A8H7F503	Allergen	Allergen	Strong	Allergen
A0A8H7F515	No	Allergen	No	No
A0A8H7F521	No	No	No	No
A0A8H7F5A3	No	No	No	No
A0A8H7F5R0	No	No	No	No

A0A8H7F5W0	No	No	No	No
A0A8H7F5X3	Allergen	Allergen	Weak	Allergen
A0A8H7F5X9	No	No	Strong	No
A0A8H7F9B1	No	No	No	No
A0A8H7F9B9	No	No	No	No
A0A8H7F9J4	No	Allergen	No	No
A0A8H7F9M2	No	Allergen	No	No
A0A8H7F9N9	Allergen	No	No	No
A0A8H7F9P1	No	No	Weak	No
A0A8H7F9P8	No	No	No	No
A0A8H7F9S0	No	No	No	No
A0A8H7F9Z7	Allergen	Allergen	Strong	Allergen
A0A8H7FA26	No	No	No	No
A0A8H7FA31	No	No	No	No
A0A8H7FA35	Allergen	Allergen	Strong	Allergen
A0A8H7FA39	No	Allergen	No	No
A0A8H7KHJ5	No	Allergen	Weak	Allergen
A0A8H7KHL7	No	Allergen	No	No
A0A8H7KHM4	Allergen	No	No	No
A0A8H7KHY6	No	No	No	No
A0A8H7KHY9	No	No	No	No
A0A8H7KI06	No	No	No	No
A0A8H7KI07	No	Allergen	No	No
A0A8H7KI34	No	No	No	No
A0A8H7KI74	No	No	No	No
A0A8H7KI86	No	No	No	No
A0A8H7EY46	No	No	No	No
A0A8H7EY79	No	No	No	No
A0A8H7EY98	No	No	No	No
A0A8H7EYA3	No	No	No	No
A0A8H7EYA8	No	No	No	No
A0A8H7EYF4	Allergen	No	No	No
A0A8H7EYG5	No	Allergen	Weak	Allergen
A0A8H7EYK5	No	No	No	No
A0A8H7EYN7	No	No	No	No
A0A8H7EYT6	No	No	No	No
A0A8H7EYU8	No	No	No	No
A0A8H7EYX5	No	Allergen	No	No
A0A8H7EYZ2	No	Allergen	Weak	Allergen
A0A8H7KGF3	No	Allergen	Weak	Allergen
A0A8H7KGJ7	No	No	No	No
A0A8H7KGK6	No	No	No	No
A0A8H7KGL8	No	No	No	No

A0A8H7KGS5	Allergen	Allergen	Strong	Allergen
A0A8H7KGQ7	No	No	No	No
A0A8H7KGS0	No	No	No	No
A0A8H7KGT3	No	No	No	No
A0A8H7KGV1	Allergen	Allergen	Strong	Allergen
A0A8H7KGW1	No	Allergen	No	No
A0A8H7KH39	Allergen	Allergen	Weak	Allergen
A0A8H7C6H3	No	No	No	No
A0A8H7C6I5	No	No	No	No
A0A8H7C6J4	No	No	No	No
A0A8H7C6N2	No	No	No	No
A0A8H7C6P7	No	No	No	No
A0A8H7C6R3	No	No	No	No
A0A8H7C6T9	No	Allergen	Weak	Allergen
A0A8H7C6V2	No	No	No	No
A0A8H7C6Y3	No	No	No	No
A0A8H7C6Y4	No	No	No	No
A0A8H7C736	No	No	No	No
A0A8H7C741	No	No	No	No
A0A8H7C7B6	No	No	No	No
A0A8H7C7C0	No	No	No	No
A0A8H7F897	No	No	No	No
A0A8H7F8L2	No	Allergen	No	No
A0A8H7F8S2	No	No	No	No
A0A8H7F8W0	No	Allergen	Weak	Allergen
A0A8H7F8W7	No	No	No	No
A0A8H7F8X7	Allergen	Allergen	Weak	Allergen
A0A8H7F944	No	No	No	No
A0A8H7F949	No	No	No	No
A0A8H7F970	No	No	No	No
A0A8H7F988	Allergen	No	No	No
A0A8H7F989	No	Allergen	No	No
A0A8H7F8D1	No	No	No	No
A0A8H7EZ56	No	Allergen	No	No
A0A8H7EZ70	No	No	No	No
A0A8H7EZ71	No	No	No	No
A0A8H7EZD4	No	No	Weak	No
A0A8H7EZE4	No	No	No	No
A0A8H7EZR9	No	No	No	No
A0A8H7EZS4	No	No	No	No
A0A8H7EZW8	No	No	No	No
A0A8H7EZZ7	No	No	No	No
A0A8H7F022	No	No	No	No

A0A8H7F070	Allergen	Allergen	Weak	Allergen
A0A8H7F080	No	Allergen	Weak	Allergen
A0A8H7F0A2	No	No	No	No
A0A8H7F040	No	No	Weak	No
A0A8H7EZQ1	No	Allergen	Strong	Allergen
A0A8H7KIJ9	No	Allergen	No	No
A0A8H7KIK3	No	No	Weak	No
A0A8H7KIP2	Allergen	Allergen	No	Allergen
A0A8H7KIV3	Allergen	Allergen	Strong	Allergen
A0A8H7KIW2	No	Allergen	No	No
A0A8H7KIY0	No	No	No	No
A0A8H7KIZ1	No	No	No	No
A0A8H7KJ25	No	No	No	No
A0A8H7KJ29	No	No	Weak	No
A0A8H7KJ64	No	No	No	No
A0A8H7KJD1	Allergen	Allergen	Weak	Allergen
A0A8H7KJE2	No	No	No	No
A0A8H7KJ94	No	No	No	No
A0A8H7KIS1	No	No	No	No
A0A8H7FBQ2	No	No	No	No
A0A8H7FBU8	No	No	No	No
A0A8H7FBX9	No	No	No	No
A0A8H7FBY2	Allergen	Allergen	No	Allergen
A0A8H7FC67	No	No	No	No
A0A8H7FC87	No	No	No	No
A0A8H7FCC0	No	No	No	No
A0A8H7FCC1	No	No	No	No
A0A8H7FCE1	No	No	No	No
A0A8H7FCS9	No	No	No	No
A0A8H7FCX1	Allergen	No	No	No
A0A8H7FDC2	No	No	No	No
A0A8H7KA29	No	Allergen	No	No
A0A8H7KE12	No	Allergen	No	No
A0A8H7KEY2	No	No	No	No
A0A8H7KF85	No	No	No	No
A0A8H7KFC9	No	No	No	No
A0A8H7KFL1	No	No	No	No
A0A8H7KFQ2	No	No	No	No
A0A8H7KFU7	No	No	No	No
A0A8H7KFW2	No	No	No	No
A0A8H7FC48	No	No	No	No
A0A8H7FAT8	No	No	No	No
A0A8H7FBF5	No	No	Strong	No

A0A8H7FBH7	No	Allergen	No	No
A0A8H7FBL2	No	No	No	No
A0A8H7FBM0	No	Allergen	No	No
A0A8H7F3G6	No	No	No	No
A0A8H7F411	No	No	No	No
A0A8H7F452	No	No	No	No
A0A8H7F4B4	No	No	No	No
A0A8H7F4D7	Allergen	Allergen	No	Allergen
A0A8H7F4E9	No	No	No	No
A0A8H7F4N8	No	No	No	No
A0A8H7F4S2	No	No	No	No
A0A8H7F4U7	No	No	No	No
A0A8H7F4V7	No	Allergen	Weak	Allergen
A0A8H7F3G4	Allergen	Allergen	Strong	Allergen
A0A8H7F4C8	No	No	No	No
A0A8H7F4K7	No	Allergen	No	No
A0A8H7F4W8	No	No	No	No
A0A8H7KKE9	No	No	No	No
A0A8H7KKF3	No	No	No	No
A0A8H7KKP4	No	No	No	No
A0A8H7KKQ6	No	Allergen	No	No
A0A8H7KL16	No	No	No	No
A0A8H7KL67	No	No	No	No
A0A8H7KLP0	No	Allergen	No	No
A0A8H7KLU5	No	Allergen	No	No
A0A8H7KM29	No	Allergen	No	No
A0A8H7KL29	No	No	No	No

Table 2 – Analysis of Potential Allergenicity of proteins which showed sequence similarity to known allergens.

Accession numbers	AllerTop	Algpred	AllerCatPro
A0A8H7BZ90	Not Allergen	Allergen	Weak
A0A8H7BXR6	Allergen	Allergen	Strong
A0A8H7BZY2	Allergen	Allergen	Weak
A0A8H7BZY5	Allergen	Allergen	Weak
A0A8H7C094	Allergen	Not Allergen	Strong
A0A8H7C106	Allergen	Allergen	Weak
A0A8H7C107	Not Allergen	Allergen	Weak
A0A8H7C141	Allergen	Allergen	Strong

A0A8H7C192	Not Allergen	Allergen	Weak
A0A8H7C1M1	Allergen	Allergen	Weak
A0A8H7C289	Allergen	Allergen	Weak
A0A8H7C2R6	Allergen	Allergen	Weak
A0A8H7C2T7	Allergen	Allergen	Strong
A0A8H7C2R1	Not Allergen	Allergen	Weak
A0A8H7C3N6	Allergen	Allergen	Not Allergen
A0A8H7C416	Not Allergen	Allergen	Weak
A0A8H7C4N9	Not Allergen	Allergen	Weak
A0A8H7C559	Allergen	Allergen	Not Allergen
A0A8H7C5D4	Not Allergen	Allergen	Strong
A0A8H7C4K1	Not Allergen	Allergen	Weak
A0A8H7C4E8	Not Allergen	Allergen	Weak
A0A8H7C5H1	Allergen	Allergen	Weak
A0A8H7C5K9	Not Allergen	Allergen	Strong
A0A8H7C8B6	Not Allergen	Allergen	Weak
A0A8H7CC30	Allergen	Allergen	Weak
A0A8H7ETL0	Not Allergen	Allergen	Strong
A0A8H7EVT7	Not Allergen	Allergen	Strong
A0A8H7EVX2	Allergen	Allergen	Strong
A0A8H7EW94	Not Allergen	Allergen	Strong
A0A8H7EWI6	Allergen	Allergen	Weak
A0A8H7KJK2	Not Allergen	Allergen	Weak
A0A8H7F503	Allergen	Allergen	Strong
A0A8H7F5X3	Allergen	Allergen	Weak
A0A8H7F9Z7	Allergen	Allergen	Strong
A0A8H7FA35	Allergen	Allergen	Strong
A0A8H7KHJ5	Not Allergen	Allergen	Weak
A0A8H7EYG5	Not Allergen	Allergen	Weak
A0A8H7EYZ2	Not Allergen	Allergen	Weak
A0A8H7KGF3	Not Allergen	Allergen	Weak
A0A8H7KGS5	Allergen	Allergen	Strong
A0A8H7KGV1	Allergen	Allergen	Strong
A0A8H7KH39	Allergen	Allergen	Weak
A0A8H7C6T9	Not Allergen	Allergen	Weak
A0A8H7F8W0	Not Allergen	Allergen	Weak
A0A8H7F8X7	Allergen	Allergen	Weak
A0A8H7F070	Allergen	Allergen	Weak

A0A8H7F080	Not Allergen	Allergen	Weak
A0A8H7EZQ1	Not Allergen	Allergen	Strong
A0A8H7KIP2	Allergen	Allergen	Not Allergen
A0A8H7KIV3	Allergen	Allergen	Strong
A0A8H7KJD1	Allergen	Allergen	Weak
A0A8H7FBY2	Allergen	Allergen	Not Allergen
A0A8H7F4D7	Allergen	Allergen	Not Allergen
A0A8H7F4V7	Not Allergen	Allergen	Weak
A0A8H7F3G4	Allergen	Allergen	Strong

Table 3 – Proteins determined as Putative Allergens via consensus of 3 popular allergen prediction tools.

Gene Ontology and Protein Family Analysis

The 55 putative allergens were subjected to Gene Ontology analysis via QuickGO and data present on UniProt. On classification of putative allergens based upon their protein family, it was found that 12 of these putative allergens were found to be from Aldehyde Dehydrogenase family of proteins, 4 each belong to Cyclophilin family and Heat Shock family of proteins.

Protein Family	Frequency
Aldehyde dehydrogenase family	12
Heat shock protein 70 family	4
Cyclophilin-type PPIase family	4
WrbA family	3
Iron/manganese superoxide dismutase family	2
LDH/MDH superfamily, MDH type 1 family	2
Eukaryotic ribosomal protein P1/P2 family	2
EF-1-beta/EF-1-delta family	1
Transaldolase family, Type 1 subfamily	1
Peptidase S8 family	1
Thioredoxin family	1

Enolase family	1
Tubulin family	1
Flavoredoxin family	1
Universal ribosomal protein uL3 family	1
Peptidase M36 family	1
Triosephosphate isomerase family	1
Oxygen-dependent FAD-linked oxidoreductase family	1
Glycosyl hydrolase 7 (cellulase C) family	1
Glyceraldehyde-3-phosphate dehydrogenase family	1
Cu-Zn superoxide dismutase family	1
Polysaccharide lyase 1 family	1
Peroxiredoxin family, Prx6 subfamily	1
Glycosyl hydrolase 18 family	1

Table 4 – Classification of putative allergens into different proteins families

Route of Exposure Analysis

Routes of exposure of selected 55 putative allergens were determined by careful data extraction from AllergenOnline, AllerCatPro and IUIS/WHO Allergen Nomenclature Homepage.

Route of exposure

On correlating with closest allergen hit, it was determined that 11 of the 55 putative allergens elicit allergenic response via ingestive route of exposure and other putative allergens were supposed to be causing allergies via injection, airway or contact routes of exposures

Protein	Route of exposure	Protein	Route of exposure
A0A8H7BXR6	injection	A0A8H7EVT7	airway and food
A0A8H7BZ90	airway	A0A8H7EVX2	airway
A0A8H7BZY2	airway	A0A8H7EW94	injection
A0A8H7BZY5	airway	A0A8H7EWI6	airway
A0A8H7C094	airway	A0A8H7EYG5	contact
A0A8H7C106	airway	A0A8H7EYZ2	airway
A0A8H7C107	airway	A0A8H7EZQ1	airway
A0A8H7C141	airway	A0A8H7F070	airway
A0A8H7C192	airway	A0A8H7F080	airway
A0A8H7C1M1	airway	A0A8H7F3G4	airway
A0A8H7C289	food	A0A8H7F4D7	airway
A0A8H7C2R1	airway	A0A8H7F4V7	airway
A0A8H7C2R6	airway	A0A8H7F503	airway
A0A8H7C2T7	contact	A0A8H7F5X3	airway
A0A8H7C3N6	food	A0A8H7F8W0	contact
A0A8H7C416	airway	A0A8H7F8X7	airway
A0A8H7C4E8	food	A0A8H7F9Z7	contact
A0A8H7C4K1	contact	A0A8H7FA35	food
A0A8H7C4N9	food	A0A8H7FBY2	airway
A0A8H7C559	food	A0A8H7KGF3	food
A0A8H7C5D4	airway	A0A8H7KGS5	contact
A0A8H7C5H1	airway	A0A8H7KGV1	airway
A0A8H7C5K9	air and food	A0A8H7KH39	airway
A0A8H7C6T9	airway	A0A8H7KHJ5	airway
A0A8H7C8B6	airway	A0A8H7KIP2	airway
A0A8H7CC30	airway	A0A8H7KIV3	airway and food
A0A8H7KJK2	airway and food	A0A8H7KJD1	not found
A0A8H7ETL0	contact	A0A8H7KJK2	airway and food

Table 5 – Route of exposure analysis of putative allergens

.

Putative food allergens
A0A8H7C289
A0A8H7C3N6
A0A8H7C4E8
A0A8H7C4N9
A0A8H7C559
A0A8H7C5K9
A0A8H7EVT7
A0A8H7FA35

A0A8H7KGF3
A0A8H7KIV3
A0A8H7KJK2

Fig 3 – Accession numbers of putative food allergens.

Analysis of Food Allergens

For the 11 putative food allergens, Linear B cell epitopes were determined using IEDB Database and the proteins with most conserved epitopes were selected. This number amounted to four. On molecular docking analysis of these four allergens via Cluspro 2.0 with IgE as receptor, two putatively allergenic proteins were found to have lowest binding affinity.

A0A8H7C4N9		
Epitope #	**Epitope sequence**	**Epitope length**
1	NFSSNPGPN	9
2	NGGLVCDHRTGTGVPPA	17
3	CVANLDGNPTNSANCCSGSHSTPQTCPPS	29
A0A8H7EVT7		
Epitope #	**Epitope sequence**	**Epitope length**
1	DPAAIPWG	8
2	QKTVDGPSHKDWRGGRG	17
A0A8H7FA35		
Epitope #	**Epitope sequence**	**Epitope length**
1	PLRIKTA	7
2	DQDDWEAW	8
3	YKDGKYDLDFKNANSDPAKWISGVELAD	28
4	SAKYGIDAVNVGD	13
5	SHAGNKLA	8
6	AELSGVKPPY	10
7	IKLDGTPNKGKL	12
8	GLKVTQQKDIDD	12
9	GASTGVHEAVELRDGNKNEYLGKG	24
A0A8H7KIV3		
Epitope #	**Epitope sequence**	**Epitope length**
1	GRKFDDQEVQSDMKHFPFEVFSRTG	25
2	FKRKNKKDLSGNQRAVRR	18
3	RTLSSA	6
4	GDTSEKTQDLLL	12
5	LGIETAGGV	9
6	KRNTTVPTKKS	11
7	EGERARTKDNNLL	13
8	IPPAPRGVP	9
9	KTTGKSNRITITNDKGRLSKEEIERMVDEAEKYKAEDEAAASRIQSKN	48
10	ITDEKLADKFDAADKSKLETAINE	24
11	SWLDNSQEASKEEYDDKQKE	20
12	AGGAPGGFPGAGGAPGGAPGGAPGGFPGAGGEEGPS	36

Table 6 – Conserved B-cell linear epitopes of selected 4 putative food allergens.

	Members	Representative	Weighted Score
A0A8H7C4N9	18	Center	-794
		Lowest Energy	-1060.1
A0A8H7EVT7	17	Center	-662.7
		Lowest Energy	-860.2
A0A8H7FA35	50	Center	-565.1
		Lowest Energy	-597.4
A0A8H7KIV3	16	Center	-651.7
		Lowest Energy	-651.7

Table 7 – Docking Analysis of 4 putative allergens

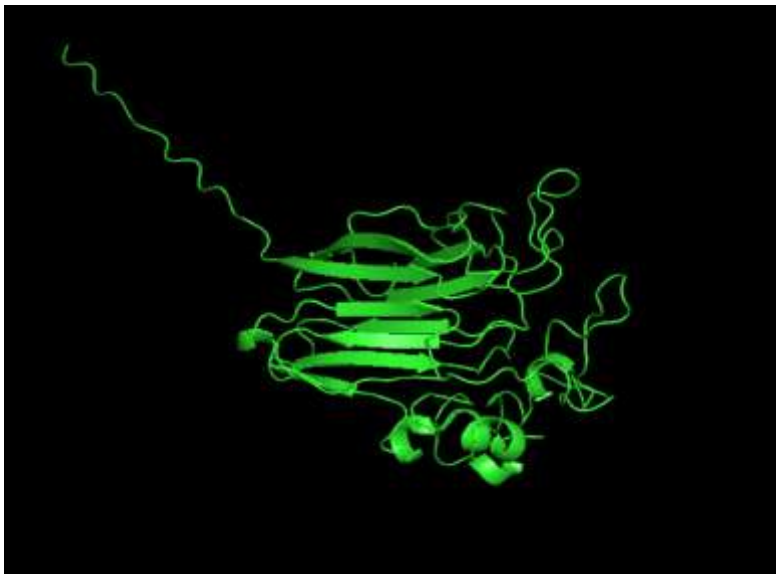


Fig 4 – Structure of protein A0A8H7C4N9

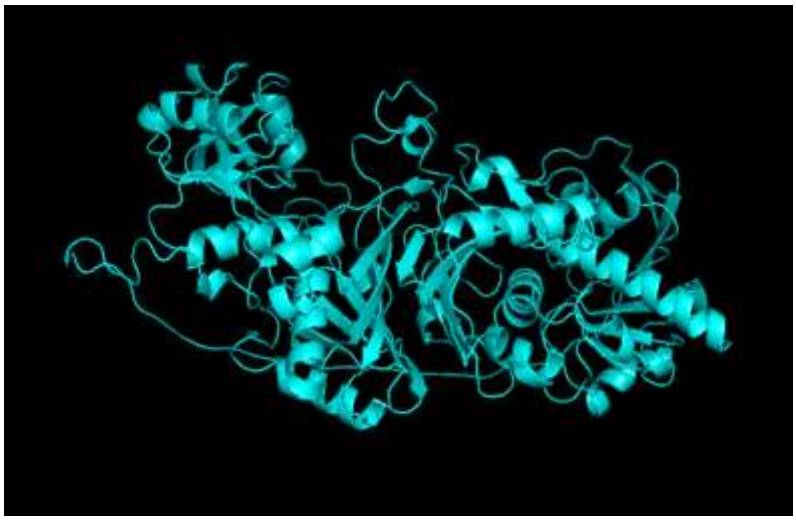


Fig 5– Structure of protein A0A8H7C4N9

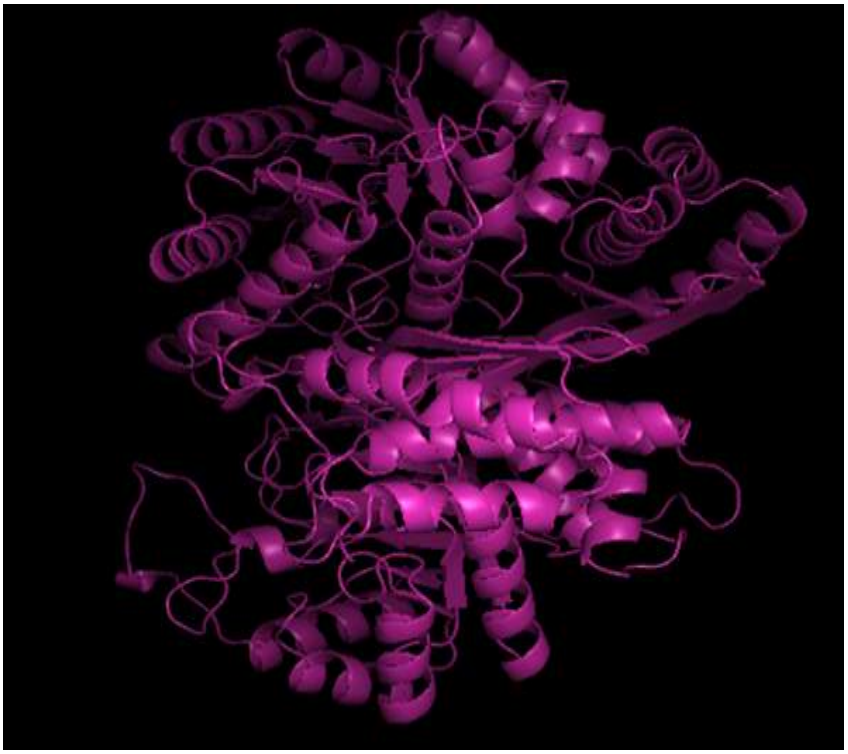


Fig 6 – Structure of protein A0A8H7C4N9

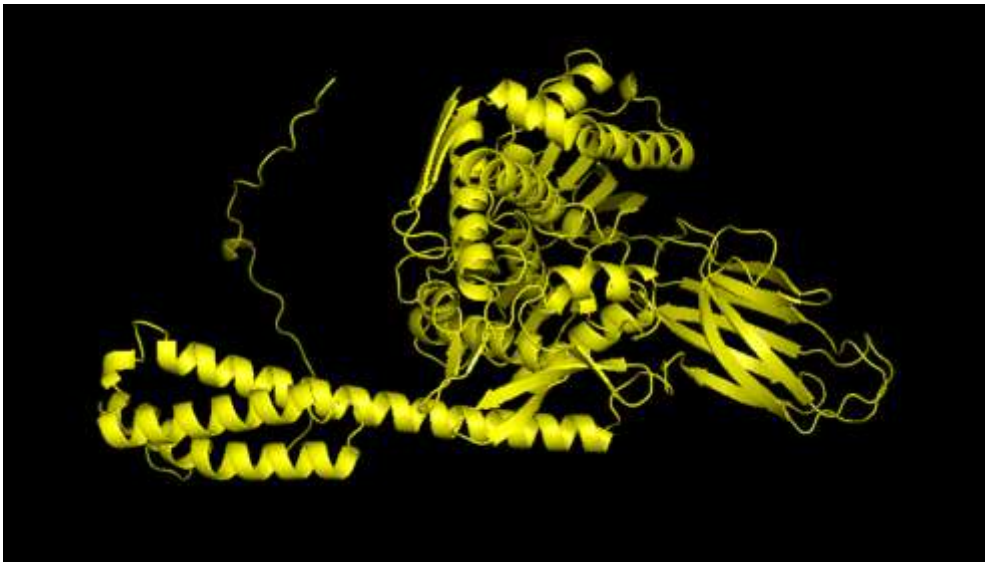


Fig 7 – Structure of protein A0A8H7C4N9

Small Peptide Analysis

On mapping of over-represented tetrapeptides on the 55 putative allergen sequences, many overlaps were observed which resulted in extension of the small peptide from a tetrapeptide to a longer peptide motif. One of such cases observed is a 9-mer peptide motif RYYAGWADK, which is found in 9 out of 12 putative allergens belonging to the Aldehyde dehydrogenase protein family. Out of this 9-mer peptide motif GWADK is a 5-mer small peptide that occurs in allergens Alt a 10 and Cla h 10, which are the only two fungal origin allergens present in AllFam database under Aldehyde dehydrogenase family of proteins. It can be a reflection that GWADK might serve as a small peptide signature of allergenicity in Aldehyde dehydrogenase family of proteins.

The tetrapeptide IYVQ was observed to be a part of not only putative allergens belonging to Aldehyde Dehydrogenase Family and Glucanase Family but also a part of allergens belonging to peanuts (*Arachis hypogaea*), a fungus (*Alternaria alternata*) and two different species of cockroaches. The tetrapeptide is also a part of documented linear B-cell allergic epitopes in IEDB database.

Several other motifs were found to be a part of mapped regions in the putative allergens and were consistently found in other allergens in the AllergenOnline database and AllFam database.

>tr|A0A8H7BXR6|A0A8H7BXR6_AGABI Endoplasmic reticulum chaperone BiP OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_11462 PE=3 SV=1
MSPSVRHPTRTQSRTLLSSLSILFLAALAVLCFFPVAVNAEETHPEYGSVIGIDLGTTYS
CVGVQRGGRVEIIANDQGHRITPSWVSFTDDERLVGDSAKNAYHSNPQNTVFDAKRL
IGRKMNEPEVKRDMKHWPFKVTEKNGKPSISVEHKGEERHFSAEEISAMVLGKMKE
TAEAYLGHKVTHAVITVPAYFNDAQRQATKDAGTIAGLHVLRIINEPTAAAIAYGLNK
GSGESQIIVYDLGGGTFDVSLLSIDDGVFEVLATAGDTHLGGEDFDNRVIDYFVKSYK
KKTGTDVSKNLRALGKLKREVEKAKRTLSSQQSTRIEIESFEDGNDFSETLTRAKFDE
LNMDLFRKTMKPVEQVLKDANVKKDEIDEVVLVGGSTRIPKVQQLLKEFFNGKEPS
KGINPDEAVAYGAAVQGGILSGAEGTDGVVLVDVNPLTLGIETTGGVFTKLIPRNTVIP
TKKSQIFSTAADNQPTVLIQVFEGERSLTKDNNLLGKFELNGIPPAPRGVPQIEVTFEM
DANGIMKVSAADKGTGKSEGITIKNEKGRLSQEEIDRMVADAEKFAAEDEANRKKIE

ALNSLSSFVYGLKNQVTDSEGLGGKLDSDDKQTILDAVKEATEWIEENGSTASVEDL
EEKLAEVQGVVNPITTKLYEGAGADGPSGDDDFEFEHDEL

>tr|A0A8H7BZ90|A0A8H7BZ90_AGABI CAZyme family GH18 and CBM5 OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_10905 PE=3 SV=1
MALRNFLRAVTFVLAALSAVEATQVMHRPDAYKDAPRKFSIETDPKTHRIQKRATSK
ASFAYFTNWGIYGANFQPTDIIPGPLTHILYSFADTDASTGHIKLTDSFADVEKHFPGDS
WDETGNNVYGCFKQLYLLKLQRRNLKVLLSIGGWTYSQSGHFNFVTSASARQNFVN
DAVQLVKDYGLDGIDIDFEYPANSAQGQGLADLVTSLRSALDQLASSNGDSTPYLITA
AVAAGSANYGNYVVPQMNRALNYWNLMAYDYAGSWLTWADNQANLFGGARTGV
NTDSAVQHYVSSGATSSKINLGMPLYGRAFEATNGLGQSYSGIGPGTIEAGIYSYETLP
LAGAQIFENTTDVSSYSYDSSKRELVSYDTPNIIKLKAQYVNAKGLGGSMFWELSTD
KVGSESLVQVSANTLGSLDQTQNHINYPNSKWDNIRNNMGGGGGGGGTNPPPTGSC
SGTAAWDSSKVYTGGMTATYNGHLWTAKWWTQNETPGSSSGVWTDNGAC

>tr|A0A8H7BZY2|A0A8H7BZY2_AGABI Translation elongation factor EF1B beta/delta subunit guanine nucleotide exchange domain-containing protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_11355 PE=3 SV=1
MSANLQKLNDHLASRSYVEGYTPSQADVHVLKAISSDPDAKKYPHVARWYTHIKSY
AAEHGSLPGSSTAGEVFLGGASSAEADDDDDDDFDLESDDEADAEAEKIKAQRVAEY
NAKKANKPKTVAKSVVTLDVKPWDDETDMAALEAAVRGIEQDGLLWGASKLVAIG
YGIKKLQITLVVEDEKVSTDELQEKIAEFEDYVQSSDIAAMQKL

>tr|A0A8H7BZY5|A0A8H7BZY5_AGABI Aldehyde dehydrogenase domain-containing protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_10995 PE=3 SV=1
MPKTYTHTFNEPTFKGTSHINTGLFINNQWVDPVEGGTLDVINPTTGKVITSVAAGSA
KDVDIAVEAATKAYKTSWGLKLPGSERGRLMNKLADLMEKHSEEFVALEMLNVGKT
YGPAKTFDVAQSIYTIRYYAGWADKIHGKTIETRDTKMAYTRHEPYGVVGCIVPWNF
PLLMTVWKLGPALSTGNAVILKPSEMTPLSALRLADLFVEAGFPPGVFNIINGYGNKV
GQAIAEHMKISKVAFTGSTLTGRKILKAASETNLKKVTLELGGKSPTIIFNDVDLEQAL
KWASAGIFINMGQSCVAASRIFVQEGIYDKFLQEFTKIAKTLTENTGSPFIPTTQHGPQI
SQIQFDRVMGFINSGKQEGAKVQIGGERHGTEGYFIKPTIFTDVTPQMSIMQDEIFGPV
CSVVKFKTEEEVLSIANDVAYGLGANVFTENTAIAMRMAHALEAGSIWVNCAQQTE
MNVPFGGFKQSGMGRELGQYALDSYTQVKAVHINLGQKL

>tr|A0A8H7C094|A0A8H7C094_AGABI Anaphase-promoting complex subunit 4 WD40 domain-containing protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_11451 PE=4 SV=1
MSGLKHSPFGKRLRASQQDSPRRRRTRPTPAPTPSLVNSSNTASSDPPDTPRARSRDYA
DRFVPSRDVGDMRTSYHLIDDAGPSTPSKTRIIPSESDALKEQANAIFTSILQTEVTPPS
PHRSHSPTRPITSSTPSTPVKRTRLFNYASPSSSKSTSPNTRRLDTPTDEAYSMSPVRQE
SRTLLESPRRQLRNVCKTPYRVLDAPELVDDFYLNLVDWSSTNVLGVGLGSCVYLWT
AHNAQVSKLCDLAEGNDSISSVSWVQKGTTLAVGTLFGRLHIYDANTLQLQRTYHQ
AHQQRIGALAWNSFVLSSGSRDRLVHHRDVRDPSTRPFKRCTGHRQEVCGLKWSGD
GGVMNATLASGGNDNKVCIWDLRGSTRSRSSTETNENTNSTLPLWKFHEHTAAVKA
LAWDPHVPGLLATGGGTQDKHIRFWNVSNGTMLNELDTGSQVCNLIWSLTSHELVS
THGFSSTSPQNQICIWKYPSLNMVASLTGHTNRVLYLAMSPDGETIVTGAGDETLRFW

NAFGGLSGVGKASSGAGALGGGGGGGGTGGLFGGGGGGDVLGGGNGVGGDVVV
KGGSALDYGRLIR

>tr|A0A8H7C106|A0A8H7C106_AGABI Aldehyde dehydrogenase domain-containing protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_10997 PE=3 SV=1
MPQTYTHTFDTPVYKGSVTINTGLFIGGQWVDPVDGGNRIDVVDPSTGKVITSVAAG
TSKDVDIAVATAEKAFKTSWGLKVPGAERGRLLGKLADLVEQHADELAALEALNVG
KPFHVAKMVDIMTEVINTLRYYAGWADKIHGKTIETNENKMAYTRHEPYGVVGTITP
WNLPLGAATLKLAPLLATGNVAVVKPSEITPLTALYFANLINAAGFPPGTVNIINGYGN
TVGEAMSMHQSIRAITFTGSTLTGRRILKASAESNLKKVALELGGKSPTIVFDDADLE
QAIKGRCQVCVAASRIYVQEGIYDKFLQGFREIAEALTSATGGPFEPGVRHGPQVSSL
QFERVMGYINSGKAEGAKVLIGGERHGDTGYFIKPTVFTEATADMKIMKEEIFGPVCS
IVKFKTEEEVTEWANNTTYGLGAHVMSENVARAIRMASNIEAGSVWVNSGWATEVG
VPFGGYKQSGMGREYSQYALDTYTQVKAVHINIGQRL

>tr|A0A8H7C107|A0A8H7C107_AGABI Extracellular metalloproteinase OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_10148 PE=3 SV=1
MRCLLTASLVFATLAGSVYGHGRDSNSPLPRRKTLGFGPEHPHAVFNSNLHPLQTAGF
VSDEPLDVARLFVTDLLGPSVSGGASWKIRKDSYTDDNTGVTHVYIRQIVNGLEVAD
GDMNINIKDGRVLSYGNSFYNGPTPAPFSSTTAPEDLRHPQAEFCDQIPRHMPYAAPM
PDGQVLLDAPSQSPVISHIAESNCAHPQLPASLTSDSSDFRPALLQFMLAATPNRELAQ
DITENYQKHIDDMLISDESHFAPGDEATVQFKVDNVPDAVNPAKARLAYIQLPSKLDG
AMELQLVWKFEVEMQDNWYESAVTVSAPHRIISVVDWASDAPVPTPSPERKPATYK
VFGWGVNDPECGERSVEKESFDGHASPAGWHAIPHEHDPSYISVRGQEKPGFWRNS
TTTWGNNVLAQENWEGLNRFMDNYRPDGGEDRVFDFKYEPNETDKEDALAEAKAY
INATVTQLFYTANLVHDIYYRYGFTEVAGNFQQYNFGRGGQQNDAVITNAQDGSGFN
NANFMTPPDGQNGRMRMYLWNTAIPYRDGDLEAGIVIHEFSHGLSTRLTGGPANSGC
LGWGESGGMGEGWGDFLATTIRSNRNYSDYSMGAWAANSITGIRHYKYSMDEDVN
PSTYKTLDKPGYWGVHAIGEVWAEMLWVVSQGLIKKHGFSDTLFPPAPGPDGVVPR
DHHFWRQDKDLLVPKHGNTLMVQLVLNGMKLQPCRPSFFDARDAIIEADRILTGGE
NVCLLWKGFAKRGLGFDAQVQGRTPWGGGIRTDGYAVHPDCRDEE

>tr|A0A8H7C141|A0A8H7C141_AGABI Transaldolase OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_10804 PE=3 SV=1
MTTSLDQLKQTGTVVVSDSGDFESIDVYKPQDATTNPSLILAAANKPGYARLIDAATK
YAKDKGGDLDAQVNNAMDRLLVEFGKEILMIIPGRVSTEVDARLSFDKEATKAKAK
QLIALYESVGVSRDRVLVKMASTWEGIQAARELEKEDGIHCNLTLLFGFGQAVACAE
AGVTLISPFVGRILDWYKKSTGKNYEGDEDPGVQSVKKIFSYYKQHGYKTIVMGASF
RNVGEIKALAGVDFLTISPALLEELKNSTAPVPKKLDSTAKQESIPKVSFINNEAEFRW
ALLQEQMAFDKLHEGIKKFAEDGETLKEVLRKKLSA

>tr|A0A8H7C192|A0A8H7C192_AGABI Triosephosphate isomerase OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_10837 PE=3 SV=1
MPRQFFVGGNFKMNPIDRATEASLVGGLNKATLDPTTEVVIAPPAIYLISVKASVRPD
VAVSAQNLYPKDSGAFTGEISPKQLVDAGIPWVILGHSERRTIFHESSEFVAQKVRAAL
DSGLKVILCIGETLQQRESGETGAVNEAQLKPVIAAIKTEEWKNIVIAYEPVWAIGTGK

VATSTQAQDTQAEIRAVIRKSVSATVADEVRIIYGGSVTANNCKELATQPDVDGFLVG
GACLKPEFANIVNARLA

>tr|A0A8H7C1M1|A0A8H7C1M1_AGABI Aldehyde dehydrogenase domain-containing protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_11001 PE=3 SV=1
MPQTYTHTFDTPVYKGSVTINTGLFIGGQWVDPVDGGDRIDVIDPCTGKVITSVAAG
TSKDVDIAVAAAEKAYKTSWGLKVPGSERGKLMAKLADLVEQHADELAALEALNG
GKPFYMAKLGDIALVIKALRYYAGWADKIHGKTIETNENKMAYTRHEPYGVVGAITP
WNAPLGSIGMKITPLLATGNVAVLKPSEFTPFTALYFANLINEAGFPPGTVNIINGYGNT
VGGAIAMHPSIRAIGFTGSIVTGRKILKASAESNLKKVALELGGKSPAIIFDDADLEEAI
KAASIALSCVYNCAGQSCVAGSRIYVQERVYDKFVQGFKKRAEDLAAATGGPFEPGV
QHGPQISSLQFERVMGYINSGKAEGAKVLIGGERHGDTGYFIKPTVFTDATADMKIIK
EEIFGPVCSIVKFKTEEEVTEWANNTTYGLAAYVMTENVARAIRMASNIEAGTIWVNS
GPVGDMGVPFGGYKQSGMGRELGQYALDAYTQVKAVHINIGQRL

>tr|A0A8H7C289|A0A8H7C289_AGABI Peptidyl-prolyl cis-trans isomerase OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_10435 PE=3 SV=1
MSASSNRPVVFMDIQIGETPAGRMKFELFSDVVPKTAENFRQLCTGEYRVNSRPQGY
KNATFHRVVQGFMCQGGDFLKGDGTGSFSIYGEKFPDENFTDKHVGPGLLSMANSG
PDTNGCQFFVTTAKCDFLDGKHVVFGKVIEGMLTLRKIENVPTGPNNRPKLVVKIIEC
GEM

>tr|A0A8H7C2R1|A0A8H7C2R1_AGABI CAZyme family AA7 OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_10349 PE=3 SV=1
MVRTFANFVLTLNPGMNPKETSGTDALNTGLRGFQYLIHPRKMSITSLSLSIYLLLLST
CFLICGAQNLVADLKSQGIELVEPGDQGYNSASAAFNRRFVFKPAVVTFPTTPAQVSSI
VKTAVKYKKHVAARGGGHSYVANGLGGQNGAVVIDMNRHFTRIQVNNQANTAKID
SGSRLGDIALTLNNYGRGFGHGTCPYVGIGGHSLLGGFAYASRLWGMVVDVIESIDLV
LANGTITTASKNKNSELFWGMRGAGPSFGITTSMTIKTFAVPPSATVFQYTWDLNATS
AASFLNAYQTFSLGQVPPQFGSELVLSKGSRQGRVSITLQGVWYDAANKFNAIIRPLV
TKVSQKPRNQMVKAGKYIDSVAFFGESNNRLNTTNAPDTFDTFYVKSLLTPESQPMT
TKSSQAFMQYLANQGFQSQSAWFIEVEEFGGPGSLVNAVPLDSTSFGNRGALFLMQF
YVYESNPNNPFAQSGFSLADGMVNSVTSNNPSNWPYTAYPNYLDNRLQNWQQLYY
GQHYPRLQRLKGSVDPGNVFQFPTSIEKP

>tr|A0A8H7C2R6|A0A8H7C2R6_AGABI Peptidyl-prolyl cis-trans isomerase OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_9424 PE=3 SV=1
MSNVFFDITVDGQAAGRIVFKLYDDVVPKTTKNFRELATGQHGFGYAGSSFHRVIPNF
MLQGGDFTRGDGTGGKSIYGEKFADENFARKHTKPGLLSMANAGKNTNGSQFFITT
VVTDWLDGKHVVFGEVVEGMDLVKKVETLGSQSGKVSKKVTIANSGTC

>tr|A0A8H7C2T7|A0A8H7C2T7_AGABI Peptidyl-prolyl cis-trans isomerase OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_10350 PE=3 SV=1
MATFLRRFATTASAGKNMANVYFDIAMNQKPSGRIVFKLYDNDVPKTAKNFRELATG
RHGFGYAGSPFHRIIPAFMLQGGDFTHRNGTGGKSIYGEKFADENFIHKHTKPGLLSM
ANAGKNTNGSQFFITTVQTPWLDGKHVVFGEVVEGMDVVKAIEAQGSVSGTPKAAV
TITSSGEVEPTQ

>tr|A0A8H7C3N6|A0A8H7C3N6_AGABI Flavin reductase like domain-containing protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_9383 PE=3 SV=1
MTSEPLQPFAHTPEWKYTESPNPGFKFGRKVDETPAGMEWLKGLESGWEIMDTSKT
DPGHLYKILTAGITPRPVAFVSSVSEDGVENLGVFSWFNQVSPTPPVISISCTNHPTRLK
DTARNIRATKNFTVNIISEAFIENANACAIDAPTYFSEWELTGLTKEPSIHVKAPRVKES
AFSMECELFQAIDIVDPKSLKATNTLILGYVKYIHMRKDTMDPTRGIPDTGKLKPICR
LGGVSYAKLGDGYQIPRPAWANVFEELKEKFGEDAVTGKGKTTEGKL

>tr|A0A8H7C416|A0A8H7C416_AGABI CAZyme family AA6 OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_9948 PE=3 SV=1
MDNWTSGSSLRREHDKKEINNPINNGRLSPLSEKTSHKIPQDTTASLNYRRRGIKQGS
VYPLEHTRSYPKSVSDTNTKGEAKATEPVQAPEPTSSDPVSTSTLHSSTTMPRVAIIIYS
MYGHIGTMAEAVKAGIVKAGGKVDIYQVPETLSDEILVKMKAPAKPSYPVIQPEGLT
EYDAFLFGIPTRYGTMPAQWKAFWDATGGLWAAGKLAGKYAGIFVSTGTQSGGQET
TVLTTLTTLTHHGILFVPFGYSHAFAELSSFETVRGGSPWGAGTYAGHDSSRSPVPLEL
TIAEKHGKSFWEIVSKVNF

>tr|A0A8H7C4E8|A0A8H7C4E8_AGABI Glucanase OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_9862 PE=3 SV=1
MFPRSILLALSLTAVALGQQVGTNMAENHPSLTWQRCTSSGCQNVNGKVVLDANWR
WTHRVNDFTNCYTGNEWDTSICPDGATCAQNCALDGADYAGTYGVTSSGTALTLKF
VTESQQKNIGSRLYLMADDSNYEIFNLLNKEFTFDVDVSKLPCGLNGALYFSEMAAD
GGMSSTNTAGAKYGTGYCDSQCPRDIKFIDGEANSEGWEGSPNDVNAGTGNFGACC
SEMDIWEANSISSAYTPHPCREPGLQRCEGNTCSVNDRYATECDPDGCDFNSFRMGD
KSFYGPGMTVDTNSPITVVTQFITDNGSDNGNLQEIRRIYVQNGQVIQNSNVNIPGIDS
GNSISAEFCDQAKEAFGDERSFQDKGGLSGMGSALDRGMVLVLSIWDDHAVNMLW
LDSDYPLDASPSQPGVSRGTCSRDSGKPEDVEANAGGVQVVYSNIKFGDINSTFNNN
GGGGGNPSPTTTRPNSPAQTMWGQCGGQGWTGPTTCQSPATCHVINDFYSQCF

>tr|A0A8H7C4K1|A0A8H7C4K1_AGABI Superoxide dismutase OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_8864 PE=3 SV=1
MLSIARTALRPRTILTRGLAARAASVHTLPPLPYAYDALEPHICEEIMKLHHTKHHQTY
VNGLNAAEEAYAKTDDVKAKINLQSALKFNGGGHINHSLFWKNLSPTSEHGGKLGD
GPLKDEIVIAFGSIDEFKKKFNAATAAIQGSGWGWLGWNSSTQQLEIVTTANQDPLLS
HIPIIGIDIWEHAFYIQYKNVKPDYLNAIWNVVNFEEAEQRLLAAAKK

>tr|A0A8H7C4N9|A0A8H7C4N9_AGABI CAZyme family GH152 OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_8810 PE=4 SV=1
MKLQLSASFAFISSVAARTFTVYNGCPFTVWPAIFTDLNVGSAVPDHVTGWEAPAFTA
VSFFVPDNWTAGRMWARRNCNFSSNPGPNSCLTGGCNGGLVCDHRTGTGVPPATVA
EWTLGAAGGLDWYDVSLVDGYNLPMRIDNNVGCPVPSCPVDLGPDCPAQLKGPFDS
SGFPVGCKSACVANLDGNPTNSANCCSGSHSTPQTCPPSGVAFYDYFKSRCRNSYVY
AYDESSGTALFNCPSSRRADYTLTFCPPP

>tr|A0A8H7C559|A0A8H7C559_AGABI Transcriptional regulator family: C2H2 zinc finger OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_8085 PE=4 SV=1
MPHTHPYQADCYCYPCDRVFARPSQLQRHLRMTLIHNSGARIYRCPQCNKGFTQKS
NISEHLRSVHGLIPEDNSGNRSRGGSATRRRNGNPLSLGQGKRLAGENARREQVESD

GSRTVYLPAVPTMVVPVDKEAPSVGVFSLATLENMHTEKINGEFKIPDFKEGLGAVDS
STYNSISASSTEAGSPSPYNFITSGPSPFIPSSTIPPPINPIFNLSPEDAEINHTENQVKAFF
ADLEARLKVPGLLAAFPNPSSVNNAQTVVEPSLGEGSTSSWRLPPMSDFFNGSFLAA
GPSNYQTRSVNNAVTEPSLGEGSTSSSRLPPMSDFFNGSFSAAGPSNYQTRQLSPVSST
LLSNWFQAIETQRQENDLKYYEQRYGANSLFEDAFPSEPLTVHNGTVRPQDIEQQKQ
QQQQQPQQLGHYSPLFSDDLNLGLADYSTDDWYSNYTPTP

>tr|A0A8H7C5D4|A0A8H7C5D4_AGABI 60S acidic ribosomal protein P1 OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_8941 PE=3 SV=1
MRKHCDCLYTTTNPTTTRVDHPAPAVPRILLIQMATSELAATYAALILADDGIEITSDKI
ISITNAAGVELEPIWASLLAKALEGKNVKDLLSNVGSGGGAPAPAAAGGAAAGGAA
AEAPKEEEKVEEKEESDDDMGFGLFD

>tr|A0A8H7C5H1|A0A8H7C5H1_AGABI Superoxide dismutase OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_8938 PE=3 SV=1
MAHVLPDLPYAYDALEPYISRQIMELHHKKHHQTYVNALNTAEAAYAKASTPKERIA
LQAALKFNGGGHINHSLFWQNLAPAAGAGGQLKPGPLKDAIDQTFGGLDNLKKEFN
TTTAGIQGSGWGWLGVNPSSKRLEISTTPNQDPLLNLVPIIGVDIWEHAFYLQYLNVK
ADYLNAIWSVINFDEAQRRYVEATQGSKL

>tr|A0A8H7C5K9|A0A8H7C5K9_AGABI peptidylprolyl isomerase OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_9046 PE=4 SV=1
MSSHRPITYFDISIGDKSIGRVVFSLYNDLVPKTAENFRALCTGEKGAGQLGKPLHYK
GSSFHRVIKGFMCQGGDFTAGNGTGGESIYGEKFEDEAFPVQHTKPFLLSMANAGPG
TNGSQFFITVAPTPHLDGKHVIFGEVIKGKSIVRQIENSSTSEGDVPLAAVIVTNCGELS
ADDPSLTEQAALDSGDAYEDYPDDEDRDLSKPETVIEIAKVIRELGNKLYKEGKIEEA
HQKYQKSIRYLDTHQEVPEKSPDLKETYIALLTPLLLNSSLAAIRANPPSSSNALAAVT
NTTRALNSLDLSTADKAKALYRRGMAHGILKNEDQQLEDLIAASKLVPEDALISGEL
AKISQRKKEHREREKKAYAKLFS

>tr|A0A8H7C6T9|A0A8H7C6T9_AGABI Aldehyde dehydrogenase domain-containing protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_8272 PE=3 SV=1
MPQTYTHTFDTPLYKGSITINIGLFIDGKWVDPVDGGDRIDVVDPCTGKVITSVAGGT
SKDVDIAVAAAEKAYKTSWGLKVPGSERGKLLAKLADLVEQHGDELAALEALNVGK
PFHMAKMVDIMTITVQSLRYYAGWADKIQGKTIETNENKMAYTRHEPYGVVGAITP
WNIPIAAAALKIAPLLATGNVAVLKPSEVTPLTALYFADLVNKAGFPPGTVNIINGYGS
TVGDAISRHPSIRAIGFTGSILTGRKILKASAESNLKKVALELGGKSPTIVFDDAELDQA
IKWASGGIFSNMGQGCVAGSRIYVQEGIYDQFLQGFKKSAEVLTDATGGPFEPGAQH
GPQVSSLQFERVMGYINSGKTEGANVLIGGERQGDTGYFIKPTIFTEAKADMKIMQEE
IFGPVCSVVKFKTEEEVTEWANNTTYGLSANVLTTNVARSIRMANNLEAGSVWVNS
GPVVDMAVPFGGFKQSGNSKEFGQYALDAYTQVKAIHINIDPVDGGDCIERFLQSTW
FLSTIAAKEVVQQKLLNDARDDVTVRGRYCHTTRNLKRVKNIFCPGCSKSHWSTSRS
ARVYLHDVGTMSMSPRVELPSAPSGHTEAPVHGSVELSFKSQQPCPELVCIEFGTPIH
VKLPHLSGDHKVSLLQK

>tr|A0A8H7C8B6|A0A8H7C8B6_AGABI CAZyme family AA6 OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_7559 PE=3 SV=1
MYGHIAKMAESAKAGIISAGGKVDIYQVPETLPQETLNLFKALPKPDYPIATMQTLED

YDAFLLGIPTRFGTMPAQWKTYWDDTGLLWMESKLSGKYAGIFVSTNMLGGGQEV
TVSSSLSVLVHHGINYVPFGFAHAFAELTNLESVHGGKGSPWGAGTLAGPKGPDGEF
RQPSEIELRMAEKQGKAFWETVSRVSF

>tr|A0A8H7CC30|A0A8H7CC30_AGABI Aldehyde dehydrogenase domain-containing protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_6783 PE=3 SV=1
MPQTYTHTFDTPLYKGSVTINIGLFIDGKWVDPVDGGDRIDVVDPCTGKVITSVVGG
TSKDVDIAVAAAEKAYKTSWGLKVPGSERGKLLAKLADLVEQHGDELAALEALNVG
KPFHMAKMIDIMTITVQSLRYYAGWADKIQGKTIETNENKMAYTRHEPYGVVGAITP
WNIPIAAAALKIAPLLATGNVAVLKPSEVTPLTALYFADLVNKAGFPPGTVNIINGYGS
TVGDAISRHPSIRAIAFTGSTLTGRKILKASAESNMKKVALELGGKSPTIVFNDAELDQ
AIKWASGGIFSNMGQGCVAGSRIYVQEGIYDQFLQGFKKSAEVLTDATGGPFEPGAQ
HGPQVSSLQFERVMGYINSGKTEGANVLIGGERQGNTGYFIKPTIFTEAKADMKIMQ
EEIFGPVCSVVKFKTEEEVTEWANNTTYGLSANVLTTNVARSIRMANNLEAGSVWVN
SGPVVDMAVPFGGFKQSGNSKEFGQYALDAYTQVKAIHINIGQRL

>tr|A0A8H7ETL0|A0A8H7ETL0_AGABI Malate dehydrogenase OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_11303 PE=3 SV=1
MVKAVVLGAAGGIGQPLSLLLKCNPLVTELGLYDIVNTPGVAADLAHISTPAKVEGN
LPDNDGLSKTLKGADVVVIPAGVPRKPGMTRDDLFKINAGIVRDLATGIATYSPKAFV
LVISNPVNSTVPIVAEVLKKHGVYDPKRLFGVTTLDVVRSSTFVAEKHGNLSLATEVV
VPVVGGHSGVTIVPLLSQSSHPLPNLSTTEYEALVKRIQFGGDEVVQAKGGAGSATLS
MAYAGAEFANKVIKAFKGEKGLIAPSYVSSEADREGAALLTKELGKEVAYFSSNIELG
PGGIAKINPLGKITDAERNLVKAAIPELEKNISSGVTFVVEAK

>tr|A0A8H7EVT7|A0A8H7EVT7_AGABI glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_11545 PE=3 SV=1
MNLNFTSRGPQVRHEIYYFDDDPMAIFLVEQKLFKVHRHPFIQGSQFFRDMFAQAGG
KGQGATESMTEEQPLPLPDVTVKEFEVLLWAMWPMPNEKSLIQILGYDQEEKLALLS
IAHRFVFDQIFQYILKEIKPDDIRVVERIRLGDKYDLTEWLLSAYKNILDHTPEGKPTRE
EAEVLGLERVNKLLQAKNRMLYEQSLPYKSYAEGWKSYAESYKAHEDRCLQCRRCP
RRPKMVYHPPRCTDSFDLSPAPIEEEVEMVSVENQPDDFPLAGIRIPSTENLVKKYFFD
DPNEQGCALIVAISSPFLDVEYMAYLFKYDSVHGRYQGKVETKDGKLIIDGHKIAAFA
EREPANIKWGECGAEYIVESTGVFKTEELAKEHLKGGAKKVVITAPGSGVPTYVVGV
NLDKYDPKELVISNASCTTNCLAVLAKVINDKFGIVEGLMTTVHATTATQKTVDAPA
KKDWRSGRSVTNNIIPASTGAAKAVTKAIPDLEGKLTGLAFRVPTLDVSVVDLVVRLE
KETSYDDVKKAMKDAADGKHPGIEKGIVDYTEEDVVSTDFVGSNYSSMIFDAKAGIA
LNSRFMKLVAWYDNEWGYARRVCDEVVAAVSASRMFSYKNPNSGHIHHLRFLPSTH
NKLIAMVKVGINGFGRIGRIVLRNALQFQDIEVVAVNDPFIDLEYMAYMFKYDSVHG
RFKGTVEVKNGSFVVDGRPMKVFAERDPAAIPWGSVGADYVVESTGVFTTIDKASA
HLKGGAKKVVISAPSADAPMYVCGVNLDKYNPKDTIISNASCTTNCLATLAKVIHDN
FGIVEGLMTTVHATTATQKTVDGPSHKDWRGGRGVGNNIIPSSTGAAKAVGKVIPSL
NGKLTGLSMRVPTQDVSVVDLVVRLEKPASYEQIKEVMRKAAEGEYKGIIAYTDEDV
VSTDFISDNNSCVFDAKAGIQLSPNFVKLIAWYDNEWGYSRRVCNLLQYVAKEDAK
AGI

>tr|A0A8H7EVX2|A0A8H7EVX2_AGABI Aldehyde dehydrogenase domain-containing protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_11289 PE=3 SV=1
MPSIFTHQWDTSVYKGSTSINTGLFINGEFVDGVKNTTIDVVNPANGKLITKISEATEA
DIDIAVEAAHKAFETTWGLNCSGSKRGDMLYKLAQLMEKNIDDLSAIEALDNGKTFL
WAKSVDLTLSISTIKHYAGWADKNFGQVIETDEKKLTYSRHEPIGVVGQIIPWNFPLL
MLAWKIGPALATGNCIVLKPSEFTPLSALRISHMKIDKVAFTGSTLVGRKVMEAAAKS
NLKNVTLELGGKSPVVIFDDADLEQSVNWTAHGLFWNHGQACCAGTRIFVQEGIYD
KFLQKFTDKIKEIKLGDPFGLGIDQGPQVSQIQYDRIMSYIESGRAEGATVHVGGERH
GNEGYFIQPTIFTDTTPDMKIVKEEIFGPVGAVIKFKDGKEVIKQANDSNYGLAAAVFS
QDINKAIETAHAFKAGTAWVNCANTIDAGVPFGGYKQSGIGRELGEYALHNYTNVK
AVHVNLNWKM

>tr|A0A8H7EW94|A0A8H7EW94_AGABI CAZyme family AA5 OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_10442 PE=4 SV=1
MRCTGNTLEPCGAGGRLNLFTSGGAPPAPPAILLDIGNWVSLGCWTDAVNGAPRTLS
VGMATDGPVTTESCTTACFEDGWRFAGTEFSHECYCGSSLNAASTQTPDGDCNMVC
EGDSSEICGGPNRLSVYNYTGTDLPTNPGGGDNGGAVFPVLDDLPTGWAYNACWVD
NAHGRILQTLVSDSPTMTVETCIQACDARNLTVAGLEFSTQCFCGSNLAAGPVLADES
SCNMGCGGNTTEACGGPSRTSVYTKGPLTVFPVPTPLQDDLPGEFTYAGCLHEFDGG
RILPWMLEWPTNNSATACMERCAEYGYPAAGVEFGIQCFCGDISDVTDKNGVIGNEA
DCNIPCPGDPAHLCGGGGRLNYYTWQGDLNVWHEPAVTGWYEFFIPGVIVPLIATVG
TNSKVTFLEKHGTGFPNTTGAFEFDPSLSNDFSKAWRELQGVKTDVFCAGSVILPDKI
GRQLNVGGWSLDSTYGVRLFTPNGELGTNSTGDWEEDYPSLKLQRGRWYPTASVLS
NGSVLVLGGEIGSNDRAQPNLEVLPKPDGGDTVIELDWLARTDPNNLYPFIVVLPSQN
IFVGYWNEARILEPVNFDTIKELPNIPGNVNNFLAGRTYPLEGAAMPLPQHAPYTEPL
EILICGGSTEGAGEASDNCVSLQPEAAEPKWIIERMPSKRVLSCMVALPDGTYMIMNG
ATQGIAGFGLANNPNLGAVLYDPTLPRTQRMSILNNTIVARMYHSESILLPDGRVLVAG
SDPQTNFDNGTVKYPEEFRVEVYVPHYLAAGQQQPTFDLPEHDWSYNGQYTITNVH
LFQGQTSGLRVSLIGASSSTHGNQMGARTIFPAVSCSGTTCTITAPPNAGICPPGWFML
FVLDGSTPSVARWVRIGGDPSQMGNWPDMPGFTLPGL

>tr|A0A8H7EWI6|A0A8H7EWI6_AGABI Aldehyde dehydrogenase domain-containing protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_11261 PE=3 SV=1
MPQTYTHTFDTPVYKGSVTFNTGLFIDGQWVDPVDGGDRIDVVDPCTGKVITTVAA
GTSKDVDIAVAAAEKAYKTSWGLKVPGVERGKLLGKLADLVEKHSDELAALEALNV
GKHFHVAKMADIPLAVNALRYYAGWADKVHGKTIETTENKMAYTRHEPYGVVGAIT
PWNFPLGTVSFKIAPMLATGNVVILKPSEITPFTALYLASLINTAGFPPGTVNIINGYGN
TVGEAISRHSSIRAIGFTGSTLTGRKILKASAESNLKKVTLELGGKSPTIVFDDADLEQ
AIKWASMGIFFNMGTSGFLQGFTKTAEVLAGATGGPFEPGVQHGPQVSNLQFERVMG
YIKSGKSEGAKVLIGGERHGDTGYFIKPTIFTEAKPDMKIMQEEIFGPVCSVVKFKTEE
EVTEWANNTTYGLAASILTENVARAIRMASNIEAGSISINSGPSAEPNVPFGGYKQSGI
GRELGQYALDTYTQVKAVHINLGQKL

>tr|A0A8H7EYG5|A0A8H7EYG5_AGABI CAZyme family GH16 OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_8455 PE=4 SV=1

MHRHTGVLIIPVLLTWLAGIATAASGTTCNATTLCPASAPCCSEFGFCGDDSFCLGGC
NPFASHSVDSCRPSPICEDAMFTFPDNSRILSNADYFEGNASEYDWVVNKGNIMNTN
SSGGELVLTLTEENGGTRLSSTRYVYYGTISAKLKTGKWGGVVTAFITMSDIKDEIDW
EFPGTHTTQGQTNYFWQGVIPTDKTNGKTTTEPLSDTFSNYHTYTIDWQQDQLKFLI
DGKNQQTVKKSDTIDSNGVAHYPSTPSRIQLSLWPAGTDDQAPGTVQWAGGMIDWN
DPDYKAAGHFYATVQSVEVKCAPVLKPSPNDTSYIYSNSSDHSSPTITLSNHSTLLNS
APVDGVSGMHGMMVGLIALFLAFVHLF

>tr|A0A8H7EYZ2|A0A8H7EYZ2_AGABI Superoxide dismutase [Cu-Zn] OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_7487 PE=3 SV=1
MNIRHLRPPAQQQHAMDYYHNLRNAVSRNKPVAFATTAAAALGVGYAYYYSRMGNI
PLATRATAILLPEDGSNVEGTIVFVQSARTGPVTLMGNIRGLPPNAKRGFHVHQWGDL
TKGCTSAGPHFNPFDQTHGAPSDKVRHVGDLGNLQSNGKGEVSLNQQDSVLSLNGA
NSIIGRAVVIHARTDDHGRGGDVESLKTGNAGARVACGVIGLAETK

>tr|A0A8H7EZQ1|A0A8H7EZQ1_AGABI 60S acidic ribosomal protein P2 OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_6486 PE=3 SV=1
MRLREGLWLRKRGMILHNTKPQVFLWIKSDLWLSSIHLFQNTVSDLKCAAIRKVLDA
GGVETDEDQLSKLLSELKGKDINDLIAEGSSKLASVPSGGGGGGGGAAAAASGGAA
PAAEEKKEEKEEEKEESDDDMGFGLFD

>tr|A0A8H7F070|A0A8H7F070_AGABI Aldehyde dehydrogenase domain-containing protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_6757 PE=3 SV=1
MPQTYTHTFDTPLYKGSVTINIGLFIDGKWVDPVDGGDRIDVVDPCTGKVITSVAGGT
SKDVDIAVAAAEKAYKTSWGLKVPGSERGKLLAKLADLVEQHGDELAALEALNVGK
PFHMAKMFDIMTITVPSLRYYAGWADKIQGKTIETNENKMAYTRHEPYGVVGAITPW
NIPIAAAALKIAPLLATGNVAVLKPSEVTPLTALYFADLVNKAGFPPGTVNIINGYGSTV
GDAISRHPSIRALGFTGSTLTGRRILKASAESNLKKVSLELGGKSPVIVFDDAELDQAI
KWASGGIFSNMGQACIAGSRIYVQEGIYDQFLQGFKKSAEVLTDATGGPFELGAQHG
PQVSSLQFERVMGYVNSGKTEGANVLIGGERQGNTGYFIKPTIFTEAKADMKIMQEEI
FGPVCSVVKFKTEEEVTEWANNTTYGLSANVLTTNVARSIRMANNLEAGSVWVNSG
PLPDVAVAFGGFKQSGNSKEFGQYALDAYTQVKAIHINIGQRL

>tr|A0A8H7F080|A0A8H7F080_AGABI CAZyme family AA6 OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_6806 PE=3 SV=1
MSRPPRIAIIIYSMYNHIAELAEAEKKGIEDAGGQATIFKVEETLEEKVLEKMHVPKG
DGRMSNSYPVASNETLENHDAFLFGIPTRYGNMPAQIKTFWDRTGSLWVNSKLAGKF
AGVFVSTGGLGGGQEETGYSILSTLVHHGIVFVPFGYSHAFEDLSNIKEVHGGSAVGA
GTIAGSDGSRRPSSLELGMAEKQGKGFYNLVSRVKFDQGLERPGEDGVNAGAADTD
NVAYGSTSVN

>tr|A0A8H7F3G4|A0A8H7F3G4_AGABI 60S ribosomal protein L3 OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_4519 PE=3 SV=1
MPVIVRNEAHHVVDARSSPQDFKMSHRKYEAPRHGSLGFLPRKRSARHRGKVKSFP
KDDPKKPVHLTAYMGYKAGMTHVVRDLERPGSKMHKREIVEAVTIVETPPMMVVG
VVGYVETPRGLRTLTTVWANHLSDELKRRFYKNWYRSKKKAFTRYAKKHAEDGGK
SIGRELERIRKYCTVVRVLAHTQIRKTPLSQKKAHLMEIQVNGGSIADKVEFAHGLFE
KPVEISTIFEQDEVVDVIAVTKGHGFEGVTHRWGTKKLPRKTHKGLRKVACIGAWHP

SKVMFSVARAGQNGYHHRTELNKKIFRVGSGADDANASTEADATKKSITPMGGFPH
YGIVKNDFLLLKGSIPGIKKRVITIRKSLMVHTSRRDLEKIQLKFIDTSSKFGHGSFQTF
EEKAAFLGRMNILVHARDIFSVGVCDQRKRAGVNSEQRLNAFVGIHGECTSG

>tr|A0A8H7F4D7|A0A8H7F4D7_AGABI Uncharacterized protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_5001 PE=4 SV=1
MWTRTPALNPPLAGKRVKGMPQSCHARHSPHVVLDRKRGLTLTMKLLAPFVLLAST
CAASVSAAPLVERGDYAPQPAYDPKPVYKPPPKPAYDPKPHKYDPPKPDHYGIYFPKI
EFPKYEHPKYEHPKYEHPKYEHPKYEHPKYEPPKPEPPKYEPPKYQPPKPEPPKYNPP
KYEPPKPEPPKYEPPKYEPPKPEPPKYEPPKYEPPKPEPPKYEPPKYEPPKPEPPKYNPP
KYEPPKYEPPKYEPPKYEPPKWDDSYKRGELNVPTHPAHPVIPKPFGNGVYGGHDNG
NGHDGHDGNGGGHGGNGDGHGGHDGNEDGHGGPGGNGDGHGGHGGHGGHGGG
PGGNGGGHGGGGGGHGGNGGNGGGGHGGNGGGHGGNGGGHGGDGGGHGGNGG
GGHGGNGGGHGGNGGGHGGNGGGGHGGNGGGHGGGHGGDGGGGHGGNGGGHG
GNGGHGGSGGGGGGHGGNGGGHGGSHGGGKGGKGDDDDFCDSNCRDERT

>tr|A0A8H7F4V7|A0A8H7F4V7_AGABI CAZyme family PL1 OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_5256 PE=3 SV=1
MFNWFRLNFSLSRPSKYVSIRRRIPTFCSSRIQTYWRRPPRAGRTALYLQKNSISSPAVR
PVDQKRRVKVFQEDLSTRTWLSSSDSTAFCSTMRILASIVAFAAVQLAAAGTIQRRAS
VNDVANLGYATLNGGTSGGSGGSQTTVTTLDQLTSAVSGDSRKIVLISGRISGSAVVR
VGSNTSILGQPGSLLDGVGLRVLGESNVIIRNVKISRVVADVGDALGIQEAHQVWVD
HVDLSSDRDHDKDFYDGLLDITHGCTGITVTNSRLHDHWKGSLVGHSDSNGSEDTP
MTVTYANNWWHNLNSRTPSFRFGHGHIFNNVFDANADGINTRDGAQLLVENNVWT
NPAKQKPLYSTDGGFAVARGNDFGGGENTAPAGNFNSPPYSYSLQSTTTTRSNVPNTA
GQNLSF

>tr|A0A8H7F503|A0A8H7F503_AGABI Heat shock protein 70 OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_3594 PE=3 SV=1
MSAEDVYEGAVGIDLGTTYSCVGVWQNDRVEIIANDQGNRTTPSYVAFSAEERLIGD
AAKNQAAMNPKNTVFDAKRLIGRRYDDPDVKKDMTHWPFQVVEKDGSPLIKVNYL
SEEKTFSPQEISAMVLTKMKEISEAKLGKTVQKAVVTVPAYFNDSQRLATKDAGTIAG
LDVLRIINEPTAAAIAYGLDRQSKEEKNVLIFDLGGGTFDVSLLNINGGVFAVKATAGD
THLGGEDFDNALLEHFKNEFKKKTKFDISDDARALRRLRSACERAKRTLSSVTQTMV
EVDSLYQGEDFSASITRARFEEINAVLFKSTLEPVEKVLKDAKMAREKVDDIVLVGGS
TRIPKVQALVSEYFGGRQLNKSINPDEAVAYGAAVQAAVLTGQTSEQTKDLLLLDVAP
LSLGVAMQGDVFGVVVPRNTPIPTNKSRVFTTVEDNQTTVTFPVYEGERTQCRDNRL
LGEFELTGIPPMPRGQAELVTTFEVDANGLLKVTAQDRTSGRKASISITNSVGRLSSVE
IEQMIKDAEQFKQADKDFTARHEAKSDLEAYIHQVESTITSPEIGMKLKRGAKSQVES
ELARALEKLEIEDSTADELKKAQLGIKRALQKATAGIR

>tr|A0A8H7F5X3|A0A8H7F5X3_AGABI Aldehyde dehydrogenase domain-containing protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_2813 PE=3 SV=1
MPQTYTHNFDTPTYKGSVTFNTGLFIDGKWTDPVQGGDLIDVVDPSTGKVIVSIAGG
TSKDVDVAVAAAERAYRTTWGLKVPGSERGRLLAKLADLVEQHTDELAALEALNVG
MPLLVNKYFVVPGAVSTFRYFAGWSDKIHGKTIEPSENKLAYTRHEPFGVCGAITPW
NFPLGVVAAKLAPALATGNTVVLKPSELTPLTALFLADLIKEAGFPPGVVSIVNGYGH
TVGDAISRHPSVMKITFTGSTLTGRKILKASADSNLKKVTLELGGKSPTIIFDDADFEQ

AIKWASMGVFFAAGQTCTAGSRIYVQEGIYDRFLQAFAKVAESLAEATGSPFDAGVQ
HGPQVSSAQLDRVLGYIKSGKAEGAHIHSGGERIGDTGYFIKPTIFTEVKADMKIMQE
EIFGPVCSIVKFKTEEEVTKWANNTTYGLAANVLTQNVSLAVRMAHNLEAGSIFVNS
SQSPERQVPFGGFKQSGMGREMGQYGLDTYTQVKAVHINIGLQI

>tr|A0A8H7F8W0|A0A8H7F8W0_AGABI Heat shock protein 70 OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_2327 PE=3 SV=1
MAVVGVDFGTLHSKIGVARHRGIDIIANEVSNRATPSLVAFGPKQRSIGESAKTQETSN
FKNTIGGLNRLIGRTFDDPQVQNVEKNFTHAALVDLNGTIGVEVNYLGERQQFSFTQ
LVGAYLGKLRDITANELKTGVTDIVIAVPGYYTEIQRRAILDAAAIANLNVLRIINDTT
ATALGYGITKSDLPDPENPRHIAFVDVGHSTFSVAIVAFAKGQLTIKSTAYNHNLGGRD
IDYVLLKHFAEEFKTKYKIDVLANPKATFRLAVGCEKLKKILSANAEGPLNIESIMND
VDASSKMSRDQLEALIPSLLDGIDGPIKQALAESGLTVDQIDSVELVGGSSRIPSVRTRI
SQAFNNKPLSVTLNQDEAVARGATFACAMLSPVFRVRDFHIHDINHYPVQVQWQAVP
TDPDEDTEILLFPQGNAIPSTKILSFYRRQAFHVDAAYKDPSGLPGGIKPVIGGFDVTD
VPPDPKGDSTIVKVKARLNLHGIVSFENTYVEEIEEREVEPAPMDVDQQQNATATSSG
AAEGEAAPPVPPKKKKIVKKKEVPFFTTTSSLDKKTLEQLRERESEMHAADKLVQDT
EDRKNALEEYIYDTRGKLDDRYAAFVTADEKSKLLQELSAAEDWLYTEEGEESTKSV
YVSRLDALKTLGDPITFRYRETEDRQRSVAALRETLNNYMNQATSNDEKYAHIDVKD
KESIVERVAVIQKWLDDMSVRQAERRKDVDPVLTSEQVTKKRDEIIFFATPILTKPKPK
MPTGTTGTDSGTGTPKRDGTPQPPPPQKNEQEGGKDASMDVD

>tr|A0A8H7F8X7|A0A8H7F8X7_AGABI Peptidase S8/S53 domain-containing protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_2160 PE=3 SV=1
MTRIIVSLLAVSAALASPLSTAHNDFSYRSSLSLAPLVDEFHPHGTVNNSYIVIFKNDV
PFDLISNHMSFLHDAHAEDPLVDDFAGVQQVYDGHLNGYAGRFTRNVVDRLRAMPE
VAYIEKDQIVHTTEHLTQKGAPWGLARVSHRPELTFSTFTKYVYDSQAGQGVDAYVI
DTGINVQHEEFEGRATWGKTVPQDEDEDGNGHGTHCAGTIASRKYGIAKHANVIAV
KVLGSNGSGTMSDVVSGVVWAAREAKAKAVAAAKEFAATGKTKHKGSVANMSLGG
GKSPALDDAVNGAVDTGLHFAVAAGNENRDACSSSPASAEKAITVGASTLGDARAYF
SNYGPCVDVFGPGLNIKSTYKGGRSATAILSGTSMASPHTAGVLAYLLSIYPSKEFNPI
FEKDELVSITSPVSQTSLSGIYAITHAALPQWISSFLPAPEFVDFIAPIPGHPPTLTPTQLK
KALIALSTPDMLDDLPANTVNLLIFNNATHS

>tr|A0A8H7F9Z7|A0A8H7F9Z7_AGABI Thioredoxin OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_97 PE=3 SV=1
MTVTAINSLQEFKTIINSGKVVVIDFWATWCGPCRVISPIFEKLSADAQQVEFYKVDV
DAQQDIAQEVGIKAMPTFVAFKDGNKVKELVGAKPQELTALISASSALV

>tr|A0A8H7FA35|A0A8H7FA35_AGABI phosphopyruvate hydratase OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_138 PE=3 SV=1
MSITKIHARQIFDSRGNPTVEVDLHTAKGRFRAAVPSGASTGVHEAVELRDGNKNEY
LGKGVSKAVENVNQIIAPKLIESGLKVTQQKDIDDWLIKLDGTPNKGKLGANAILGVS
IAAAEAGAAEKSIPLYQHFAELSGVKPPYILPCPAFNVINGGSHAGNKLAFQEFMLLPT
GAQSFTEAMKIGTETYHTLKKVISAKYGIDAVNVGDEGGFAPNVSGADEALELLSEAI
KKAGYEGKIKIALDVASSEFYKDGKYDLDFKNANSDPAKWISGVELADLYLSYVKK
YPIVSIEDPFDQDDWEAWTHFTSKSGIQIVGDDLTVTNPLRIKTAIEKKACNGLLLKIN

QIGTISESIQAAQLSQSDGWGVMVSHRSGETENTVIADLVVALGLGQIKTGAPARSER
VAKYNALLRIEEELAGSGAYFAGDKGLATGATPPQLLQK

>tr|A0A8H7FBY2|A0A8H7FBY2_AGABI REJ domain-containing protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_1188 PE=4 SV=1
MSSSPSSSPSSSSSSSPPPSTSDIDSSTSISSSVIASTTSSFTSEISPTSSDMSQTTTTSSSSLP
LTPLPSTTTTTTTSPSTTSTTSTTTPEPPPTTTSTSTTTTSSSTETPPTTTSTSTSTSTSTSTS
TSTSDPPTTSSDPPTTTSTTTLTSSSPQQPSPQPPPTTTQFTPTTTFESVVFITITNDAGSLI
STAPPNITSTFTTTGPDGGFVTVTQVVANPSLLPNNHDKDSGFFGNTGAVAAVFVLVG
LATTSIAIWTIFAILRRRRQSRLVHETADAPSRATPAPRPLLEDDDDPEQQHFSPLALQM
SRQRSSPGLTSTSLGPRRSYHDDPDHPDPFNPYVEFGHIPLGPVPTNISNGYSIARTNSP
PNTHTNNDRSQRASAAASLASANGLDGYAKGTHSTTPSGASIEPLLASYYRQSDGSG
QVPGSPVIPTAPAMALLPQTPPIGSSRLLVDVLEPLEGRSVFSADDRLDPGLQQRQKEA
ESASMKDLRDEEDYSRPVLAVRNLP

>tr|A0A8H7KGF3|A0A8H7KGF3_AGABI Peptidyl-prolyl cis-trans isomerase OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_5474 PE=3 SV=1
MLRSRVFLDFAVDSTPYGRVIFELFTDTAPKTCENFRALCTGEEGLSAGGHPLYYKGS
PMHRSIKGFMIQGGDFTKRNGTGGYSIYGGTFADEDLQRPLDSPALLCMANKGPDTN
GSQFFITLQECPHLNGKHVVFGKTTKIGLSRPLLLQAVVSSSSESERSKEMKSSLQVV
QRMLEREEAERREQERKQILLDIKKTFSEEPKPTGVRYKGRGRMKYVDPEHQFRS

>tr|A0A8H7KGS5|A0A8H7KGS5_AGABI Malate dehydrogenase OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_4252 PE=3 SV=1
MLCRAALRSSSARLFSTSAARQTKVAVLGAGGGIGQPLSLLLKSDPLVSSLSLYDIRG
APGVAADVSHVDTASEVTGYPADKIDAALDGVQVVVIPAGVPRKPGMTRDDLFNTN
ASIVRDLATAVARASPSAHILVISNPVNSTVPIVAAALEKAGVFDPRRLFGITTLDVVRA
QRFLAGIVESDPRQTPVTVIGGHSGATIVPLLSQSQYGKGIKGETYDKLVHRIQFGGDE
VVKAKDGAGSATLSMAYAGAKFTNLLLRGLNGEKGVITPTFVRSPLYESQGIDFFSSS
VELGLQGVEKIHPIGDISPEEEKLLAACLPELKKNIEKGKAFVAPN

>tr|A0A8H7KGV1|A0A8H7KGV1_AGABI Tubulin alpha chain OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_4327 PE=3 SV=1
MREVISVHVGQAGVQIGNACWELYTVEHGLSPDGRIIEGSPSQNDDGFSTFFSETSAG
KHVPRSLYIDLEPNVIDEVRNGPYRSLFHPETLVTGKEDAASNYARGHYTVGKERIDT
VMDKVRRLADNCSGLQGFFVFHSFGGGTGSGFGALILERLSTDYGKKSKLEFSVYPA
PTLANSVVEPYNSVLTTHTTLEHSDCSFMVDNEAIYDICKKNLNITSPSLVNLNRLIAQ
VVSSITASLRFDGSLNVDLNEFQTNLVPFPRIHFPLATFAPIISAEKAHHEQNSVADMTF
SCFEPGNQMVKCDPREGKYMACALLYRGDVVPKDVNAAVAIIKTKRTIQFVDWCPT
GFKLGICNEPPAHVPGGDLAKVSRSMCMLSNTTAISSAWSRLDHKFDLLYSKRAFVH
WYVGEGMEEGEFSEAREDLAALEKDYEEVGIDSADVEEAEY

>tr|A0A8H7KH39|A0A8H7KH39_AGABI Aldehyde dehydrogenase domain-containing protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_4627 PE=3 SV=1
MAHIATITIAATNKQIPVPTGLFINNEFVPSVEDSPEKILAINPATEEVICPVVSASPRDID
VAVKAARQAFQTTWGKNVTGFERSRLINKLADLIERDQQELAELETLNNGKPCLRYY
AGWADKIIGQSLEVDNKTKLAFTRNDPIGVCGQIIPWNYPINMWAWKVAPALAVGCT
IVMKPSELTPLTALKLSELVREAGFPPGVVNTVPSYGHVGGAALAAHPDVDKVAFTG

STVTGRKIMEAAAKSNLKKVSLELGGKSPHLIFESADLEQAASWAALGICYNTGQDC
TAGSRVYVQEPIYDRFIDLLTAKVKATIMGDGFDEISTGGPIISKGQYDRVWAYIEAGK
QEGAKPILGGVKRTGKGFFVDPTIFTDIHRDMKIVKEEIFGPVLSVGKFKTESEAISLA
NETNYGLGAGVHSSDANQCMRVSSSLEAGTVWVNQYNLLSNNVPFGGKKQSGIGR
ELGSHALAEYTSVKAVHWNFGEKVDWPL

>tr|A0A8H7KHJ5|A0A8H7KHJ5_AGABI Thioredoxin domain-containing protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_4630 PE=3 SV=1
MPSLRLGDTAPDFEAETTTGNIKFHEWIGNSWAILFSHPGDFTPVCTTELGEVARRAP
DFQKRGVKLIGISANGLEDHHAWVKDINEYGSKFGHTDVRFPIIADGDRKISTLYDML
DYQDATNRDAKGLPFTIRTVFVIDPKKTIRLMIAYPASTGRNFDEIIRVVDSLQLGDKH
RITTPVNWNKGDDVIVHPSVTSEEAKVLFPEHTVHKPYLRTTPLKV

>tr|A0A8H7KIP2|A0A8H7KIP2_AGABI Uncharacterized protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_2776 PE=4 SV=1
MGGGARYPYPKYVWSPAGGWWVRPSNWASNTAIVAGGMAAVLYAVWRVSANNEQ
RITQPSRFIPSMLWAKEYEDRKLKLDKDK >tr|A0A8H7KIV3|A0A8H7KIV3_AGABI Heat shock protein HSS1 OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_3123 PE=3 SV=1
MSSSKAIGIDLGTTYSCVGVWQNDRVEIIANDQGNRTTPSYVSFSDNERLIGDAAKNQ
VAMNPHNTVFDAKRLIGRKFDDQEVQSDMKHFPFEVFSRTGKPYIRVEYRGEKKEFS
PEEISSMVLSKMKETAEAYLGTTVNNAVVTVPAYFNDSQRQATKDAGVISGMNVLRII
NEPTAAAIAYGLDKKVSGERNVLIFDLGGGTFDVSLLTIEEGIFEVKATAGDTHLGGE
DFDNRLVTHFIQEFKRKNKKDLSGNQRAVRRLRTACERAKRTLSSAAQTSIEIDSLFEG
IDFYTSLTRARFEELCQDLFRGTLEPVEKVLRDSKIDKSNVHDIVLVGGSTRIPRIVKLV
SDFFNGKEPNKSINPDEAVAYGAAVQAAILSGDTSEKTQDLLLLDVAPLSLGIETAGGV
MTALIKRNTTVPTKKSEIFSTYSDNQPGVLIQVYEGERARTKDNNLLGKFELSGIPPAP
RGVPQIEVTFDIDANGILNVSASDKTTGKSNRITITNDKGRLSKEEIERMVDEAEKYK
AEDEAAASRIQSKNGLESYAYNLRNSITDEKLADKFDAADKSKLETAINETISWLDNS
QEASKEEYDDKQKELEAIANPIMQKLYSAAGGAPGGFPGAGGAPGGAPGGAPGGFP
GAGGEEGPSVEEVD

>tr|A0A8H7KJD1|A0A8H7KJD1_AGABI Aldehyde dehydrogenase domain-containing protein OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_1464 PE=3 SV=1
MPQTFTHTFDSSLFKGTVTINTGLFIGGKWVDPVDGGDPVDVVDPCTGKVITTVAGG
TSKDVDVAVAAAEKAYKTSWGLKVPGAERGRLLGKLADLVEQHSDKLAALEALNV
GKPFPMAQFIDMSMITIPILRYYAGWADKIQGKTIETNENKMAYTRHEPYGVVGAITP
WNIPLAAAVLKIAPLLATGNVAVLKPSEIAPLTALYFADLVNQAGFPPGTVNIVNGYGN
TVGEAISRHPLIRAVGFTGSTLTGRRILKASAESNLKKVSLELGGKSPVIVFDDAELDQ
AIKWASGGIFSNMGQVCVAGSRIYVQEGIYDKFIQGFRKLAEGLADATGTPFEPGTQH
GPQVSNAQFERVMGYINSGKSEGAKVLTGGERQGDTGFFIKPTIFTEAKADMKIMQE
EIFGPVCSVVKFKTEEEVTEWANNTTYGLGANVMTQNVARAIRMANSLEAGSVWIN
SGATADMGVPFGGFKQSGNSKEFGEYALDAYTQVKAIHINIGQRL

>tr|A0A8H7KJK2|A0A8H7KJK2_AGABI peptidylprolyl isomerase OS=Agaricus bisporus var. burnettii OX=192524 GN=Agabi119p4_1674 PE=4 SV=1
MFGRIALAFVVVSIATFFCAQSVEAAKGPKITHKIYFDIKQGDDELGRIVMGLYGGTV
PKTVENFRALASNTKKDGSPLGYGYKGSKFHRVIKDFMIQGGDFTRGDGTGGASIYG

NNFPDENFKLKHTAPGILSMANAGKDTNVSQFFITTVVTSWLDNRHVVFGKVIEGM
DVVRTIEFTKTGKNDRPLVDIVIADSGELPVEAVYDDEGKEDNNVDGTPTLVNDIVAD
VITSPKTEATSETTAEEQKVAELETMLPKSVDWTMLASGIILVGVFVALFLWLGGERC
MRRLLNGNGAKYGNLKGSGNDIKRIA

Table 8 – Mapping of tetrapeptides on putative allergens

Putative Allergen Family	Motifs present in other allergens as well
Aldehyde dehydrogenase Family	RYYAGWADKIHGK
Enolase Family	WGVMVSH, KKACNGL
EF-1-Beta/Ef-1-Delta family	RWYTHIK,GLKV, WEAW
Peroxiredoxin family	GDFTPVCTTE
Heat Shock Protein 70 Family	TYSCVGV
Cyclophilin Family	FHRVI
Tubulin Family	GNACWELY
Thioredoxin Family	IDFWATWCGP
Peptidase S8 Family	GTHCAGT
Aldehyde dehydrogenase family, Glucanase Family	IYVQ

Table 9 – Peptide Signatures across Putative allergen families

CHAPTER 5 Discussion

Food allergies are immune responses triggered by otherwise harmless dietary proteins and are known to vary based on an individual's age, lifestyle, and genetic susceptibility. Among these, IgE-mediated allergies are the most common and typically affect atopic individuals — those already sensitized to certain allergens (Sicherer, 2000; Sicherer et al., 2006). On first exposure, allergen-specific Immunoglobulin E (IgE) is produced and binds to receptors on mast cells and basophils. Upon re-exposure, cross-linking of these bound IgE antibodies activates the release of inflammatory mediators, leading to an allergic reaction. Clinical symptoms can range from mild itching and gastrointestinal discomfort to severe cases like bronchoconstriction or anaphylaxis.

While food allergens from nuts, milk, and seafood have been extensively studied, fungal proteins — particularly from widely consumed edible mushrooms — remain largely underexplored. *Agaricus bisporus*, known commonly as the button mushroom, is one of the most widely cultivated and consumed mushroom species globally. Yet, very limited information exists regarding its potential to elicit allergic reactions or contain cross-reactive proteins. With increasing mushroom consumption and growing reports of fungal-related allergies, investigating the allergenicity of *Agaricus bisporus* is both timely and necessary.

The objective of this study was to screen and characterize putative allergens in *Agaricus bisporus* var. *burnetti* using a sequence-based, genome-wide approach. Following the Codex Alimentarius guidelines (2003, 2009), the analysis began with comparing the full proteome of the mushroom against a curated database of known allergens. This initial screening step is widely accepted as a foundational method to flag potential allergens based on sequence similarity thresholds — typically >50% identity over the full length, or >35% identity within any 80-amino acid window.

In silico tools were used to carry out this screening, which allowed for rapid filtering of high-risk proteins from the full proteome. By integrating sequence homology data with predictions from allergenicity assessment tools, we narrowed down the candidate list for deeper functional and immunological analysis. This included protein family classification, gene ontology annotation, epitope prediction, peptide motif analysis, and docking simulations.

Altogether, this work presents a step-wise identification and initial characterization of putative allergens in *Agaricus bisporus* and lays the groundwork for future allergen validation and potential inclusion in allergen databases.

To systematically identify candidate allergens within the *Agaricus bisporus* var. burnetti proteome, a comprehensive sequence alignment strategy was implemented. The complete proteome, consisting of 11,675 protein sequences, was aligned against a curated dataset of 2,334 experimentally validated allergens sourced from the AllergenOnline database. Using both full-length FASTA and pBLAST tools, proteins were screened under stringent thresholds—a percent identity greater than 50% and robust E-value cutoffs—to maximize specificity and reduce false positives. This initial screening yielded a subset of 318 proteins exhibiting significant homology with known allergens, which were subsequently selected for allergenicity prediction.

The 318 shortlisted proteins were evaluated using three independent, widely recognized allergen prediction servers: AllerTOP v2.0, AlgPred, and AllerCatPro. Each tool applies a distinct methodological approach, ranging from physicochemical property-based classification to motif scanning and machine learning-based predictions. By requiring a protein to be predicted as a probable allergen by at least two of the three platforms, a consensus-based strategy was employed to enhance prediction reliability. This integrative step refined the list to 55 proteins, designated as high-confidence putative allergens, and prioritized for downstream peptidomic and immunoinformatic analyses. This multi-tool consensus approach ensured robust prediction accuracy by cross-validating results across diverse algorithmic frameworks.

The 55 predicted allergens were annotated using Gene Ontology (GO) terms sourced from QuickGO and UniProt databases. These annotations provided insights into the molecular function, biological processes, and cellular components associated with each protein. Functional classification of the proteins into families revealed that a significant proportion—12 out of 55 proteins—belonged to the Aldehyde Dehydrogenase (ALDH) family. Additionally, four proteins each were identified as members of the Cyclophilin and Heat Shock Protein (HSP) families. These protein families are frequently associated with stress responses, detoxification, and protein folding—functions often implicated in allergenic pathways.

To understand the possible routes by which these proteins might elicit allergic responses in humans, a route-of-exposure analysis was conducted. This involved cross-referencing the 55 putative allergens with closest-matching entries from AllergenOnline, AllerCatPro, and the IUIS/WHO Allergen Nomenclature database. Based on similarity to documented allergens, it was inferred that 11 proteins are most likely to trigger allergic reactions via the ingestive route, characteristic of typical food allergens. The remaining proteins were predicted to be associated with airway, injection, or contact exposure routes—suggesting potential cross-reactivity with environmental or occupational allergens.

Focusing specifically on the 11 putative food allergens identified, linear B-cell epitope prediction was performed using the IEDB analysis resource. From this group, four proteins displayed highly conserved and immunologically relevant epitope regions, making them particularly interesting for further structural analysis. These four proteins were modeled and subjected to molecular docking using ClusPro 2.0, with the human IgE antibody set as the receptor. Of these, two proteins demonstrated notably low binding energy values, suggesting a strong and potentially allergenic interaction with IgE. These findings support the likelihood of allergenic activity in a food context and highlight the importance of combining sequence-based and structure-based approaches in allergen prediction pipelines.

To identify allergen-specific sequence features, overrepresented tetrapeptides were mapped across the 55 putative allergen sequences using a sliding window-based enrichment strategy. Several overlapping occurrences of enriched tetrapeptides were observed, often aligning to form longer peptide motifs, indicative of potential immunogenic hotspots. One noteworthy example is a 9-mer peptide motif, RYYAGWADK, which was present in 9 of the 12 ALDH-family allergens identified. Within this motif, the pentapeptide GWADK was found to occur in Alt a 10 and Cla h 10—the only two fungal-origin allergens listed under the ALDH family in the AllFam database. This suggests that GWADK may represent a small peptide signature associated with allergenicity within this protein family.

Another tetrapeptide of interest, IYVQ, was identified across multiple protein families, including the Aldehyde Dehydrogenase and Glucanase families. Furthermore, it was also detected in allergenic proteins from unrelated species—peanut (Arachis hypogaea), fungus (Alternaria alternata), and two distinct

cockroach species. Notably, IYVQ is also documented as part of a linear B-cell epitope in the IEDB database, further reinforcing its potential role as an immunogenic motif.

In total, several peptide motifs identified in this study overlapped with regions found in validated allergens catalogued in AllergenOnline and AllFam databases. These findings suggest that specific short peptide motifs—especially when enriched and conserved across species—may contribute to allergenicity and could serve as sequence-level markers for identifying cross-reactive or novel allergenic proteins.

CHAPTER 6 Conclusion

Agaricus bisporus is a widely consumed edible mushroom, and given its broad dietary inclusion, assessing its allergenic potential is increasingly important. In this study, we undertook a computational approach to identify and characterize putative allergens within the proteome of *A. bisporus* var. *burnetti*. Using sequence alignment tools such as full-length FASTA and pBLAST, the proteome was compared against a curated database of 2,334 known allergens, resulting in the identification of 318 candidate proteins. These were further screened through three independent allergen prediction platforms—AllerTOP, AlgPred, and AllerCatPro. Based on cross-validation, 55 proteins were consistently predicted as putative allergens by at least two tools and were selected for downstream analyses.

To better understand their biological context, these 55 proteins were classified based on protein family and annotated using gene ontology terms. Routes of exposure were inferred through homology with known allergens in databases such as AllergenOnline and IUIS/WHO. B-cell epitope prediction was performed for the subset of putative food allergens using IEDB, followed by molecular docking with IgE for the most epitope-conserved proteins to assess potential antibody interaction.

Finally, small peptide composition analysis revealed several overrepresented tetrapeptides within the allergen set, many of which aligned to known allergenic motifs. Notably, conserved sequences such as **GWADK** and **IYVQ** were identified in multiple *A. bisporus* proteins and matched with documented epitopes from other fungal and food allergens. This integrative, sequence-based analysis highlights a subset of mushroom proteins with high allergenic potential and provides a valuable framework for further immunological validation.

CHAPTER 7 References

- Aalberse, R. C. (2000). Structural biology of allergens. Journal of Allergy and Clinical Immunology, 106(2), 228–238. https://doi.org/10.1067/mai.2000.108434

- Ali, S. B., & Smith, W. (2024). Agaricus bisporus mushroom anaphylaxis: A case report and review of the literature. Journal of Allergy and Clinical Immunology: Global, 3(4), 100324. https://doi.org/10.1016/j.jacig.2024.100324

- Allergen Bureau. (n.d.). Are mushrooms an allergy risk? Retrieved June 28, 2025, from https://allergenbureau.net/are-mushrooms-an-allergy-risk/

- Chruszcz, M., Chapman, M. D., & Pomés, A. (2018). Structural insights into allergen–antibody interactions and cross-reactivity. Current Allergy and Asthma Reports, 18(5), 26. https://doi.org/10.1007/s11882-018-0787-0

- Costa, J., Villa, C., Verhoeckx, K., & Mafra, I. (2021). Protein allergenicity: Sequence, structure, and physicochemical properties. Journal of Agricultural and Food Chemistry, 69(12), 3504–3519. https://doi.org/10.1021/acs.jafc.0c07413

- Crameri, R., Garbani, M., Rhyner, C., & Huitema, C. (2014). Fungi: The neglected allergenic sources. Allergy, 69(2), 176–185. https://doi.org/10.1111/all.12323

- Dai, Y., Wang, Y., Zhang, X., & Wang, S. (2022). iAller: A novel ensemble-based tool for allergenicity prediction using multiple features. Briefings in Bioinformatics, 23(1), bbab451. https://doi.org/10.1093/bib/bbab451

- Desta, I. T., Porter, K. A., Xia, B., Kozakov, D., & Vajda, S. (2020). Performance and its limits in rigid body protein–protein docking. Structure, 28(9), 1071–1081.e3. https://doi.org/10.1016/j.str.2020.06.007

- Dimitrov, I., Bangov, I., Flower, D. R., & Doytchinova, I. (2014). AllerTOP v.2—a server for in silico prediction of allergens. Journal of Molecular Modeling, 20, 2278. https://doi.org/10.1007/s00894-014-2278-5

- Dimitrov, I., Naneva, L., Doytchinova, I., & Bangov, I. (2014). AllergenFP: Allergenicity prediction by descriptor fingerprints. Bioinformatics, 30(6), 846–851. https://doi.org/10.1093/bioinformatics/btt619

- Elghoudi, A., & Narchi, H. (2022). Food allergy in children—the current status and the way forward. World Journal of Clinical Pediatrics, 11(3), 253–269. https://doi.org/10.5409/wjcp.v11.i3.253

- Fiers, M. W. E. J., Kleter, G. A., Nijland, H., Peijnenburg, A. A. C. M., Nap, J. P., & van Ham, R. C. H. J. (2004). Allermatch™, a webtool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. BMC Bioinformatics, 5, 133. https://doi.org/10.1186/1471-2105-5-133

- Goodman, R. E., Ebisawa, M., Ferreira, F., Sampson, H. A., van Ree, R., Vieths, S., … Baumert, J. L. (2016). AllergenOnline: A peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity. Molecular Nutrition & Food Research, 60(5), 1183–1198. https://doi.org/10.1002/mnfr.201500965

- Goto, T., Yano, K., Ikeda, K., & Yamanishi, Y. (2023). AllerStat: Statistical learning approach for allergenicity prediction using comprehensive

allergen datasets. Frontiers in Immunology, 14, 1123456. https://doi.org/10.3389/fimmu.2023.1123456

- Ivanciuc, O., Schein, C. H., & Braun, W. (2009). SDAP: Database and computational tools for allergenic proteins. Nucleic Acids Research, 37(Database issue), D328–D332. https://doi.org/10.1093/nar/gkn673

- Mari, A., Scala, E., Palazzo, P., Ridolfi, S., Zennaro, D., & Carabella, G. (2006). Bioinformatics for the allergy field. Bioinformatics, 22(13), 1681–1682. https://doi.org/10.1093/bioinformatics/btl158

- Maurer-Stroh, S., Krutz, N. L., Kern, P. S., Gunalan, V., Nguyen, M. N., Limviphuvadh, V., … Ooi, E. E. (2019). AllerCatPro—Prediction of protein allergenicity potential from the protein sequence. Bioinformatics, 35(17), 3020–3027. https://doi.org/10.1093/bioinformatics/btz029

- Pomés, A. (2010). Relevant B cell epitopes in allergic disease. International Archives of Allergy and Immunology, 152(1), 1–11. https://doi.org/10.1159/000265818

- Pomés, A., Davies, J. M., Gadermaier, G., Hilger, C., Holzhauser, T., Lidholm, J. Vieths, S. (2018). WHO/IUIS allergen nomenclature: Providing a common language. Molecular Immunology, 100, 3–13. https://doi.org/10.1016/j.molimm.2018.03.023

- Radauer, C., Bublin, M., Wagner, S., Mari, A., & Breiteneder, H. (2008). Allergens are distributed into few protein families and possess a restricted number of biochemical functions. Journal of Allergy and Clinical Immunology, 121(4), 847–852.e7. https://doi.org/10.1016/j.jaci.2008.01.025

- Ramsey, N. B., & Berin, M. C. (2021). Pathogenesis of IgE-mediated food allergy and implications for future immunotherapeutics. Pediatric Allergy and Immunology, 32(7), 1416–1425. https://doi.org/10.1111/pai.13501

- Saha, S., & Raghava, G. P. S. (2006). AlgPred: Prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Research, 34(Web Server issue), W202–W209. https://doi.org/10.1093/nar/gkl343

- Sarlakifar, M., Ghadiri, N., & Ebrahimi, M. (2024). AllerTrans: A transformer-based machine learning model for allergen prediction. Bioinformatics Advances, 4(1), vtae001. https://doi.org/10.1093/bioadv/vtae001

- Schein, C. H., Ivanciuc, O., & Braun, W. (2022). Bioinformatic approaches to mapping and predicting allergenic epitopes. Allergy, 76(2), 381–393. https://doi.org/10.1111/all.14422

- Sharma, N., Patiyal, S., Dhall, A., Pande, A., Arora, C., & Raghava, G. P. S. (2021). AlgPred 2.0: An improved method for predicting allergenic proteins and mapping of IgE epitopes. Briefings in Bioinformatics, 22(4), bbaa294. https://doi.org/10.1093/bib/bbaa294

- Simon-Nobbe, B., Denk, U., Pöll, V., Rid, R., & Breitenbach, M. (2007). The spectrum of fungal allergy. International Archives of Allergy and Immunology, 145(1), 58–86. https://doi.org/10.1159/000107578

- Smolnikov, A., Grudinin, S., & Giraud, T. (2023). SoluProt: Prediction of soluble protein expression in Escherichia coli. Bioinformatics, 39(1), btac782. https://doi.org/10.1093/bioinformatics/btac782

- Soeria-Atmadja, D., Zorzet, A., Gustafsson, M. G., & Hämmarlund, U. (2006). Statistical evaluation of local alignment features for predicting allergenicity using supervised classification algorithms. International Archives of Allergy and Immunology, 139(4), 257–264. https://doi.org/10.1159/000091774

- Taylor, S. L., & Hefle, S. L. (2001). Will genetically modified foods be allergenic? Journal of Allergy and Clinical Immunology, 107(5), 765–771. https://doi.org/10.1067/mai.2001.114738

- Vita, R., Mahajan, S., Overton, J. A., Dhanda, S. K., Martini, S., Cantrell, J. R., Peters, B. (2019). The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Research, 47(D1), D339–D343. https://doi.org/10.1093/nar/gky1006

- Wang, J., Zhou, Y., Zhang, H., & et al. (2023). Pathogenesis of allergic diseases and implications for therapeutic interventions. Signal Transduction and Targeted Therapy, 8, 138. https://doi.org/10.1038/s41392-023-01344-4

- Watson, C., & Kobernick, A. (2022). Dangers at the dinner table – A report of anaphylaxis to Lion's Mane mushroom. Annals of Allergy, Asthma & Immunology, 129(5 Suppl), S147. https://doi.org/10.1016/j.anai.2022.08.931

- WHO/IUIS Allergen Nomenclature Sub-Committee. (2024). Allergen nomenclature. Retrieved from http://www.allergen.org/

- Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., … Wang, J. (2018). WEGO 2.0: A web tool for analyzing and plotting GO annotations, 2018 update. Nucleic Acids Research, 46(W1), W71–W75. https://doi.org/10.1093/nar/gky400

- Zhang, Y. (2015). I-TASSER server for protein structure and function prediction. Nature Protocols, 5(4), 725–738. https://doi.org/10.1038/nprot.2010.5

Plagiarism Report

DrillBit

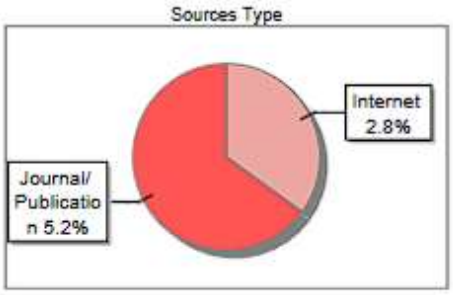
The Report is Generated by DrillBit Plagiarism Detection Software

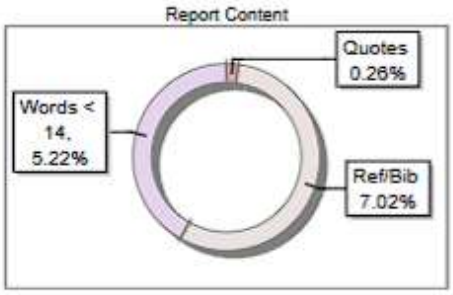
Submission Information

Author Name	Priyanka
Title	Identification and analysis of putative allergens in Edible mushroom (Agaricus bisporus)
Paper/Submission ID	4139057
Submitted by	atul.upadhyay@thapar.edu
Submission Date	2025-07-28 12:17:20
Total Pages, Total Words	78, 13697
Document type	Thesis

Result Information

Similarity 8 %





Exclude Information

Quotes	Not Excluded
References/Bibliography	Not Excluded
Source: Excluded < 14 Words	Not Excluded
Excluded Source	0 %
Excluded Phrases	Not Excluded

Database Selection

Language	English
Student Papers	Yes
Journals & publishers	Yes
Internet or Web	Yes
Institution Repository	Yes

A Unique QR Code use to View/Download/Share Pdf File