DEPARTMENT OF AGRICULTURAL, FOOD AND ENVIRONMENTAL SCIENCES

DEGREE COURSE: FOOD AND BEVERAGE INNOVATION AND MANAGEMENT

UNDERSTANDING THE MOLECULAR BASIS OF HERMETIA ILLUCENS ALLERGENICITY IN FOOD AND FEED APPLICATIONS

TYPE OF DISSERTATION: Research

Student:
ELENA URBISAGLIA

Supervisor:
PROF. MICHELE CIANCI

ACADEMIC YEAR 2023-2024

Every change in life begins with a clear and unequivocal decision:
do something or stop doing it.

(Anonymous author)

LIST OF TABLES

LIST OF FIGURE

ACRONYMS AND ABBREVIATIONS

BSF	BLACK SOLDIER FLY
BSFL	BLACK SOLDIER FLY LARVA
AK	ARGININE KINASE
H. ILLUCENS	HERMETIA ILLUCENS

CHAPTER 1
AIM OF THE THESIS

Insects are an alternative source of protein and recognized as potential novel food for human consumption.
However, as many other types of food, insects also can induce allergic reaction due to their ingestion.
Arginine kinase (AK) is an allergen present in crustaceans and it is also present in many different species of insects like the black soldier fly larvae (BSFL).

The AK of BSFL could cause a possible cross-reactivity of insect proteins in people allergic to crustacean protein due to the similarities between the insect proteins and crustacean proteins because both insects and crustaceans belongs to the same phylum Arthropoda, which means they share a common evolutionary lineage.

The description of the tridimensional structure of an allergen enable to predict cross-reactivity. In fact, similarities in tridimensional structure and repeated positioning of epitopes among allergens correlate with its allergenic activity. In case of proteins, like gluten and crustacean proteins, which do not share structural similarities, the cross-reactivity will be unlikely.

In this thesis we studied the tridimensional structure of AK from to BSFL locate the epitopes which can trigger allergic reactions.

Starting from the crystallization of the protein, it has been possible to rebuild its structure and in the end check the epitopes, their position to determine if it's more or not likely to cause allergy.

CHAPTER 2
INTRODUCTION

The Earth population is approaching nine billion of people with the likelihood of food and water shortages under all the problematics of climate change.

One way of ensuring food security could be the consumption of edible insects which will provide high quality of nutrients while requiring less water consumption and having less environmental impact compared to the conventional animal rearing.

Insects are already used as human food and animal feed in about 80 countries across Asia, Africa and America by 2 billion of people, from the estimated of FAO, food and agriculture organization of the united nations (Fao, 2013a).

In Europe, the consumption of insect products is still uncommon, but it is gaining attention. For this reason it is essential to understand the possible cross-reactivity of insect proteins in people allergic to crustacean proteins.

The reason is the similarity between the insect proteins and crustacean proteins because both insects and crustaceans belongs to the same phylum Arthropoda, which means they share a common evolutionary lineage.

In people with shellfish allergies, their immune system produces antibodies (typically IgE) against arginine kinase; since the arginine kinase protein in insects and crustaceans are very similar in structure, the immune system can by mistake recognise insect arginine kinase as the same allergen, triggering an allergic reaction.

2.1 ***Hermetia Illucens***

Hermetia illucens, known as "Black Soldier Fly" (BSF), belongs to the family Stratiomyidae, subfamily Hermetiinae, genus Hermetia.

Hermetia illucens is an American native insect, but now it has a worldwide distribution, occupying habitats between 45°N and 40°S. In Europe, this species was first recorded in Malta in 1926 and it currently inhabits mainly the Mediterranean area.

The species has three generations per year in subtropical regions and one generation per year in temperate regions (Gligorescu et al., 2019).

It can be distinguished from other Stratiomyidae for its size of 15-20 mm, and for its predominantly black colour, with blue metallic highlights on the thorax and sometimes a reddish end of the abdomen, Figure 1.

Figure 1: Adult of Hermetia illucens

The antennae are about twice as long as the head. The wings are membranous and at rest are folded horizontally over the abdomen. Females have a longer abdomen, more white hairs on the head and larger wing size compared to males, as well as a longer and translucent abdominal area from which the presence/absence of eggs can be seen (Marzouk, 2016)

2.1.1 *Reproductive cycle*

The life cycle of the Soldier Fly consists of five main stages: egg, larva, prepupae, pupa and adult, Figure 2.

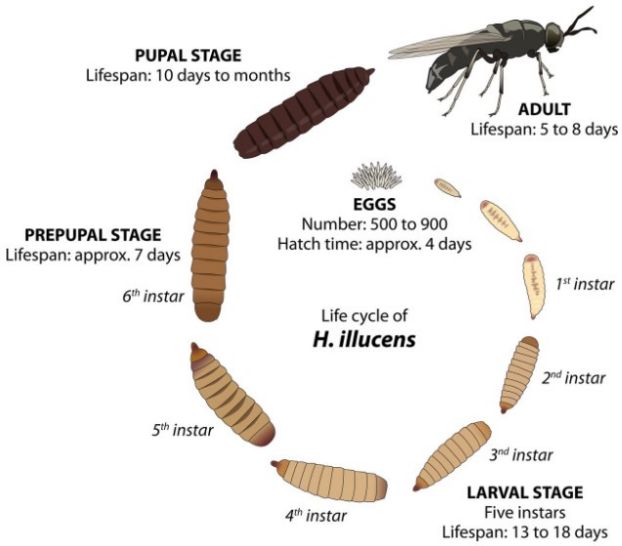

Figure 2: Life cycle of H. illucens*(De Smet et al., 2018)*

The eggs, which are light yellow and about 1 mm long are laid on the organic substrate and hatch in about 4 days, if conditions are optimal.

The larva initially it is about 1-2 mm long until fully mature, and after three stages, which last 20-30 days, it reaches a length of about 2 cm. During these days it feeds continuously. They have a cylindrical, markedly segmented body, a protruding head containing chewing mouthparts and the colour varies from white to very deep brown.

Black soldier fly larvae (BSFL) are reported as feeding on an immense variety of organic material and have already been used in small-scale waste management purposes using substrates such as manure, rice straw, food waste, distillers' grains, fecal sludge, animal offal, kitchen waste.

Figure 3: Black soldier fly larvae

H. illucens larvae have a short development time of about 20 days at 27°C before going into dry areas, where they pupate for about 18 days (Gligorescu et al., 2019).

After reaching the optimal maturity, the larva withdraws inside the last layer of skin, called the puparium, which, thanks to a process of keratinisation hardens. At this point, the pupa enters a state dormancy waiting for optimal conditions for the transition to the adult stage

In nature, pupation, which is the transition from larva to pupa, can occur in a time frame ranging from 9 days to 5 months (Oliveira et al., 2015).

The adult can reach 2 cm in length, it is black and has wings with very dense veins covering the entire membrane. Adults are short-lived, in fact they survive only 5-8 days, during which they mate about two days after flickering and do not feed because they exploit the fat stored in the larval stage (Diclaro & Kaufman, 2009).

Adults consume nothing but water, do not approach humans, do not bite or sting, and do not vector or disseminate any specific diseases (Wang & Shelomi, 2017). Adult flies only serve reproductive purposes and live about 8–20 days depending on temperature, sex, larval food quality and larval development time. The female can lay up to 500 eggs throughout their lifetime.

2.1.2 *Application of Hermetia illucens in feed*

BSFL meal and oil are already considered to be an animal-grade alternative to fish meal and fish oil used to feed carnivorous fish and in other animal diets, due to their high protein and lipid contents even when fed plant-based waste streams. BSFL has also been used in poultry feed as a partial replacement for maize or soy-based feeds. BSFL supplementation (50%) or total replacement of soybean cake in the diets of laying hens had no impact on hen health or performance and little to no effect on the eggs themselves (Wang & Shelomi, 2017).

Hermetia illucens (black soldier fly, BSF) larvae are already used as ingredient in aquafeed and can be considered one of the most promising insect for human consumption due to its ability to convert food waste into a source of proteins rich in essential amino acid, fats, vitamins, and minerals. (Delfino et al., 2024)

An added benefit is the recycling of the waste itself, as the management of organic wastes such as manure, leachates, and food waste is both costly and a growing environmental concern. Black soldier flies are also known to reduce the mass and nutrient content of swine manure at efficiencies similar to poultry manure, with benefits for improved farm hygiene, reduced pest fly populations, and reduced nutrient pollution in runoff. (Wang & Shelomi, 2017).

It is therefore well established that BSFL can be used to feed many vertebrates and can use various vertebrate wastes as a substrate, with no effects on the palatability of the BSFL-fed

meats for humans and with significant implications for sustainable and lower-input agriculture in the developing world.

Regarding microbial contamination, studies have shown BSFL's ability to significantly reduce Enterobacteriaceae colonies and Salmonella spp. in a variety of different feed sources ranging from manure to fecal matter. This is because in nature BSFL are innately decomposers and they have a multitude of mechanisms in place to reduce the microbial load in their feed. Due to the high pH of their gut, gut enzymatic reactions and competitive gut bacteria, some bacterial species (e.g., E. coli, Salmonella spp.) from the feed are not able to survive (Bessa et al., 2021).

2.1.3 *Application of Hermetia illucens in food*

Entomophagy, meaning consumption of insects, has been practiced by humans on every inhabited continent, not only historically, but also up until the present day. Today we found that insect are rich sources of proteins, good fats, and certain trace elements . Compared to other animal meat, *Hermetia illucens* has a lower environmental impact, produce fewer greenhouse gases and lower ammonia emissions than any conventional livestock.

Industrial-scale insect farms need less water and land space than pasture, can have a lower water footprint per gram of protein than any conventional livestock or even milk and eggs (Wang & Shelomi, 2017).

Insects provide high-quality protein and nutrients comparable with meat and fish, they are also rich in fibre and micronutrients such as copper, iron, magnesium, manganese, phosphorous, selenium and zinc (Fao, 2013b).

More specifically, *Hermetia* illucens is a good source of proteins and lipids, averaging 40.8 ± 3.8% protein and 28.6 ± 8.6% fat (Wang & Shelomi, 2017).

About the safety issued, the Hazard Analysis Critical Control Points (HACCP) system, a science-based and systematic tool, identifies specific hazards and establishes control systems to ensure the safety of food, throughout the insect supply chain will be a determining factor in the success and development of the edible insect sector.

Also traditional processing methods, such as boiling, roasting and frying, are often applied to improve the taste and palatability of edible insects, like *Hermetia illucens,* and have the added advantage of ensuring a safe food product (Fao, 2013b).

Insects may have associated micro-organisms that can influence their safety as food. Both insects collected in nature and insects raised on farms may be infected with pathogenic micro-organisms, including bacteria, virus, fungi, protozoa and others. In general, insect pathogens

are taxonomically separate from vertebrate pathogens and can be regarded as harmless to humans (Mézes, 2018).

As regards *Hermetia illucens,* it has a natural stronger resistance to microorganism, including bacteria, fungi and viruses due to its natural rearing environment like manure, compost, less likely to infection of the insect

According to the European Food Safety Authority (EFSA), the chemical contaminants in *Hermetia illucens* of greatest concern are heavy metals such as cadmium, mercury, lead, and arsenic, as well as the accumulated pollutants from the environment such as hormones and pesticides (Pan et al., 2022). This risk can be reduced by controlling the feed quality of the insect and the rearing condition.

2.2 **Food allergens**

The food allergens are described in the Regulation (EU) No 1169/2011 on the provision of food information to consumers, as any food or ingredients that can cause a food allergy, a reaction from the immunoglobulin Es, IgEs, of the immune system against the allergens, considered foreign by the body.

The allergens are generally protein or peptides. They are labelled following the IUIS systematic nomenclature, designated by the first three letters of the genus and the first letter of the species according to the Linnaean taxonomic system, with an increasing number with the chronological order in which they were identified and characterized.

The clinical manifestations of insect food allergies range from a mild local reaction to severe systemic clinical manifestations such as an anaphylactic shock. Reported symptoms can be divided into skin (e.g., hives, itching, rash, flushing, angioedema), gastrointestinal (e.g., abdominal pain, nausea, vomiting, diarrhea), and respiratory (e.g., asthma, dysp nea). Moreover, the duration of symptoms can range from a few minutes to 6 h and they can be life-threatening leading to: constricted airways in the lungs, severe lowering of blood pressure and shock (“anaphylactic shock”), suffocation by swelling of the throat and larynx www.FDA.org

Therefore, many effective strategies have been proposed to reduce the sensitization of insect proteins, such as heating, microwaving, glycosylation, high pressurization and enzymatic hydrolysis. Heat treatment may reduce this risk, but does not eliminate it (Pan et al., 2022).

2.2.1 *The 14 class of allergens*

From the Regulation (EU) No 1169, the Annex II list all the food allergens: gluten, crustaceans, eggs, fish, peanuts, soy, milk, nuts, celery, mustard, sesame, sulphur dioxide, lupins and molluscs.

According to the Regulation (EU) No 1169/2011, the food business operator (FBO) must inform customers about the presence of food allergens by highlighting it on the label of the packaging or in the brochure for the presentation of food.

2.2.2 *Allergens from edible insects*

The edible insect are defined as Novel Food, from the Regulation **(**UE) 2015/2283 on Novel Food. A Novel Food is any food not used to a significant extent for human consumption in the Union before 15 May 1997.

The edible insect can be the whole insect or a part of it.

Today the insects approved in the Regulation **(**UE) 2015/2283 are:

- *Acheta domesticus*, house cricket, as frozen, dried and powdered or partially defatted powder form;
- *Locusta migratoria* as frozen, dried and powdered form;
- *Tenebrio molitor* larva (yellow mealworm) as dried, frozen and powdered form;
- *Alphitobius diaperinus* (lesser mealworm) larvae as frost, paste, dried and powder form.

Insects and crustaceans (e.g., crayfish, shrimp) belong to the family arthropods (which have an exoskeleton and segmented bodies) and allergies to shellfish are quite common and potentially serious, cross-reactivity of insect proteins with crustacean proteins is perhaps the major concern for food allergies (R-biopharm, 2022).

Tropomyosin kinase (EC 2.7.11.28) and Arginine kinase (EC 2.7.3.3), proteins have been identified as the major allergenic proteins within insects that can trigger an allergic response when consumed by sensitized individuals (R-biopharm, 2022).

These allergens share similar physicochemical characteristics with most food allergens in that they are glycoproteins, are relatively stable to heat and acid−alkali, and retain their immunological activity after digestion (Zhang et al., 2015).

Tropomyosin kinase catalyze the reaction between ATP and tropomyosin with the production of ADP and O-phosphotropomyosin.

Arginine kinase is a phosphagen kinase participating in cell metabolism which catalyses the reversible transfer of a phosphoric group from Mg^{2+} ATP to L-arginine, leading to

N^{ω}-phospho-L-arginine and Mg2+ ADP, and plays an important role in cellular energy metabolism in invertebrates (Shen et al., 2011).

Myosin light chain (MLC) from *L. vannamei*, the North Sea shrimp (*Crangon crangon*), and *P. monodon* is recognized as a novel shrimp allergen is recognized as a novel shrimp allergen also identified in BSFL (Zhang et al., 2015)(Bessa et al., 2021).

Chitinases (EC 3.2.1.14) catalyze the hydrolysis of the β-1,4-N-acetyl-D-glucosamine linkages in chitin polymers, has been identified in edible insect like silkworm pupae (*Bombyx mori* L.); even after they are eaten fried or boiled, they can retain their allergenicity (Leoni et al., 2019).

Recent studies have shown the ability of IgE from patients with food allergy to shrimp and/or inhalant allergy to mites to cross-react with BSF tropomyosin variants (Leni et al., 2020) while recent data has been reported in the literature for the allergenic potential of AK from BSF

Overall, these results indicate H. illucens as a possible source of allergenic AK for individuals sensitized to crustaceans or mites, even if more studies are needed to evaluate the AK stability in food matrices under the different typical processing conditions used by food industry (Delfino et al., 2024)

The fact that IgE from shrimp allergic patients bindsing to proteins from different insects is not surprising since crustaceans and insects both belong to the same phylum (Arthropoda) (Broekman et al., 2017).

2.3 **Arginine kinase**

In invertebrate cells, Arginine Kinase (AK), together with other enzymes belonging to the phosphagen kinase family, performs the same biochemical function as creatine kinase in vertebrates acting as an energy modulator. Most of the AKs are monomers composed of two distinct structural domains, the globular α N-terminal domain involved in the arginine binding and the α-β C-terminal domain interacting with ATP, molecular weight of 44 kDa for the monomeric form and about 90 kDa for the dimeric form

Conserved amino acids important for enzymatic activity or protein stability have been identified around the active site of the enzyme. In particular, E225, E314, and C271 of AK make ionic contacts with the guanidium group of the arginine substrate, the backbone of G64, V65, and G66 forms hydrogen bonds with the carboxyl group of the arginine substrate and Y68 interacts by the OH group of its lateral chain with the amino group of the argi nine substrate by forming an hydrogen bond. In addition four conserved arginine residues (R126,

R229, R280, and R309) located in the catalytic site stabilize the phosphate groups of ADP by hydrogen bonds (Delfino et al., 2024).

From the work of (Delfino et al., 2024), four AK isoforms has been identified named X1-X4 (XP_037904709, XP_037904710 XP_037904711, XP_037904712). AK X1-X4 shows variable lengths between 356 and 604 amino acids. The AK enzymes recognized as allergens in crustacean and mites contain 356 or 357 amino acids and, accordingly, the most similar sequence from BSF is the isoform X4 (XP_037904712) which is composed by 356 amino acids and shares sequence identities of 77–88% with homologous allergens of Arthropoda .

2.3.1 *AK allergen from Hermetia Illucens*

Invertebrate AK is recognized as a major allergen responsible for food allergy to crustacean and for airway allergy to dust mites together with tropomyosin.

There are specific regions of AK from different invertebrate organisms as potential IgE linear and conformational epitopes, Figure 4 and Figure 5.

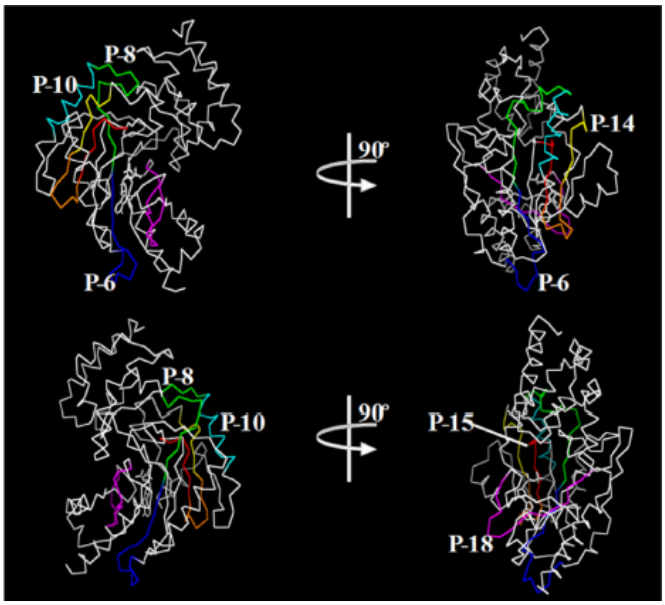

Figure 4: Mapping the IgE-specific peptides onto a molecular model of Scylla paramamosain AK. P, peptide *(Yang et al., 2015)*

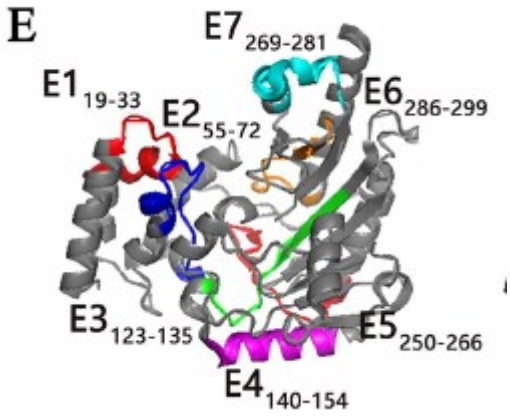

Figure 5: Position of allergenic epitopes determined on Penaeus chinensis AK protein 3D structure model *(Fu et al., 2018)*

In addition, cross-reactivity among invertebrate AK has been demonstrated highlighting the relevance of this enzyme as a cross-reactive pan-allergen (Delfino et al., 2024).

Several studies, combining bioinformatics analysis and peptide immunoassays, have identified specific regions of AK from different invertebrate organisms as potential IgE linear and conformational epitopes. In particular, six linear epitopes and four conformational epitopes were predicted in *Scylla paramamosain* AK and seven linear epitopes were predicted in *Penaeus chinensis* AK. (Delfino et al., 2024).

CHAPTER 3
MATERIALS AND METHODS

3.1 **Production of Arginine kinase**

Protein Production was performed in the lab of Prof. Claudia Folli (Department of Food and Drug, University of Parma, IT). The sequence encoding for *H. illucens* AK isoform X4 (NCBI protein id XP_037904712), was purchased from Eurofins (Milan, Italy) in the pEX A128 vector and optimized for the expression in E. coli. The AK cDNA was sub-cloned into the NdeI and BamHI restriction sites of the expression vector pET28b (Merck). The resulting plasmid, named pET28b-AKX4, was used for E. coli BL21 (DE3) cells (Merck) transformation by electroporation. For the expression of AK X4, cells were grown in Luria-Bertani (LB) medium at 37°C, supplemented with 50µg mL−1 of kanamycin under gentle shaking until OD 600 reached 0.6–0.8. Protein expression was induced with 1 mM isopropyl β-D-thiogalactoside (IPTG) at 20 °Co/n. Cells were collected by centrifugation and the pellet from 1 L of culture was resuspended in 18 mL of lysis buffer (50 mM sodium phosphate, 300 mM NaCl, and 10% glycerol, pH 7.5) containing lysozyme 1mg mL−1. After incubation for 45 min at 4 °C, cells were sonicated on ice for 30 min(1 s pulse on and 3 s pulse off) and then centrifuged. The recombinant AK X4 was purified from soluble fraction by affinity chromatography by using a HiTrap Q HP column (GE Healthcare, Chicago, IL, USA) equilibrated with 50 mM sodium phosphate, 300 mM NaCl, containing 15 mM imidazole pH 7.5 and connected to an ÄKTA Pure FPLC System (GE Healthcare). Protein elution was performed by a linear gradient of imidazole (0–500 mM), monitored by absorbance at 280nm,and checked on 12 % SDS-PAGE. Fractions containing the purified protein were pooled and the protein concentration was estimated by absorbance at 280 nm using the extinction coefficient (ϵ280 = 30745 M−1 cm−1) generated by ProtParam (https://web.expasy.org/protparam).

To characterize AK from BSF, recombinant AK X4 was produced in *Escherichia coli* by using the pET expression system and purified by affinity chromatography thanks to a six-

histidine tag linked to its N-terminus obtaining a yield of about 120 mg of purified protein per liter of cell culture

The most similar sequence from BSF is the isoform X4 (XP_037904712) which is composed by 356 amino acids and shares sequence identities of 77–88% with homologous allergens of Arthropoda (Delfino et al., 2024).

3.2 **Electrophoresis LAEMMLI gel**

The electrophoresis was performed to check the purity of the AK.

The higher the purity of the AK, the higher the tendency of the protein to crystallize.

For the preparation of the Electrophoresis LAEMMLI gel, composition Table 1, the gel supporter was mounted. To avoid the dryness of the gel a little amount of water was added on the top. Subsequently, the stacking gel was prepared, Table 2, and the comb of 1mm x 1 mm was put and waited the solidification of the gel.

Table 1: 10% Acrylamide Running gel

Water	2,35 mL
Running gel	1 mL
Acrylamide (30% stock)	1,65 mL
APS	100 µL
Temed	2,5 µL

Table 2: Stacking gel

Water	1,5 mL
Stacking buffer	0,5 mL
Acrylamide (30% stock)	0,5 mL
APS	50 µL
Temed	3 µL

Starting from the initial concentration of the Arginine kinase of 7,7 mg/mL, were made some dilution of 3 µL, 2 µL, 1 µL. From each of them has been taken 1 µL of protein, added 5 µL of sample buffer 4x to denaturate the protein and added 14 µL of water to reach for a total of 20 µL each.

The gel was placed in a holder and to balance it was placed another slide. In the gel uploaded respectively: “Pierce ™ Unstained protein MW Marker” 26610 Table 3, 20 µL of 3 µL, 20 µL of 2 µL, 20 µL of 1 µL.

Table 3: Thermo Scientific™ Pierce™ Unstained Protein MW Marker

Tris-H3PO4 (pH 7.5 at 25°C)	62.5 mM
EDTA	1 mM
SDS	2 %
DTT	50 mM
NaCl	30 mM
Bromophenol blue	0,01 %
Glycerol	50%

Between the two gels was added the Electrode running buffer, Table 4, outside the gel was used the running buffer used 1 time.
For the running the Amper was constant, for the stacking it was at 10 mA, for the running for each gel it was at 16 mA for a total of 2 hours.

Table 4: Electrode running buffer

Glycin	0,192 M
SDS	0,1 %
	25mM

After the 2 hours of running it was added the staining buffer for 20 minutes, Table 5, then the strong destaining buffer for 20 minutes, Table 6, and at the end the weak destaining buffer overnight, Table 7.

Table 5: Staining buffer

Brillant blue comassie R 250	0,1 %
Acetic acid	10%
Methanol	50%

Table 6: Strong destaining buffer

Acetic acid	10%
Methanol	50%

Table 7: Destaining buffer

Acetic acid	7%

Methanol	5%

After defrosting the protein, there was the presence of particulate matter, so the solution was spinned. The surnatant was taken and loaded into the gel at 3 different concentrations: 1 µg, 2 µg, 3 µg.

For the visualization of the gel was used the "Calibrated densitometer GS-800".

The protein ladder used was the Pierce™ Unstained Protein MW Marker (Thermo Fisher Scientific®), Figure 6.

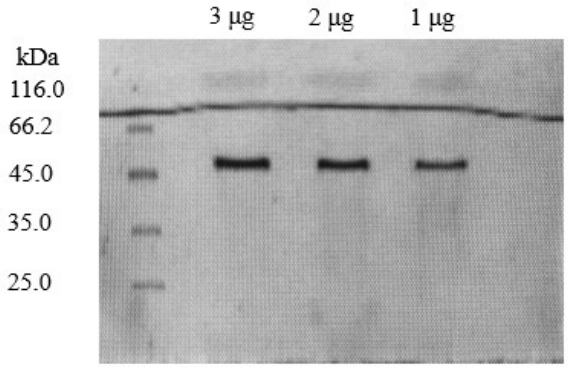

Figure 6: Gel obtained from electrophoresis at different concentration

3.3 **Bradford Protein Assay**

To check the protein concentration, was performed the Bradford protein assay.

In case of denatured condition, the protein cannot crystallize.

The Bradford used was B6916 from Sigma-Aldrich® (2024 Merck KGaA, Darmstadt, Germania).

The preparation of the standard curve was performed by using different concentration of BSA from 1 µg to 10 µg.

The absorbance was 0,5282, the spectrum of absorbance was 595 nm.

3.4 **Crystallization**

Crystallization involves taking randomly oriented molecules in solution and ordering them into regular positions in the solid state called crystals.

The crystallization occurs by using a range of salt and polyethylene glycol (PEG) solution at different concentration.

The crystallization has been performed by using the Crystal Screening kit (Jena Bioscience, Germany):

- JBScreen Classic HTS I CS-201L (PEG based);
- JBScreen Classic HTS II CS-202L, (Ammonium Sulphate, MPD, Alcohol and Salt based);
- JBScreen Basic HTS CS-203L;
- JBScreen JCSG++ HTS CS-206L;
- JBScreen PACT++ HTS CS-207L;

Since it was the first time of using them, before the opening of the kits they were centrifuged at 300 g x 60 seconds by using the "Eppendorf centrifuge 5804".

The procedure used was: 100 µL taken from the cockpit of the kit and put in the reservoir of the plate corresponding; take 1 µL from the reservoir to put in the cockpit of the plate; take 1 µL of Arginine kinase and put in the cockpit of the plate containing the 1 µL of the kit. Repeat the procedure for all the kits, later the plates were put in the fridge at 18°C.

An example of plate is the Figure 7.

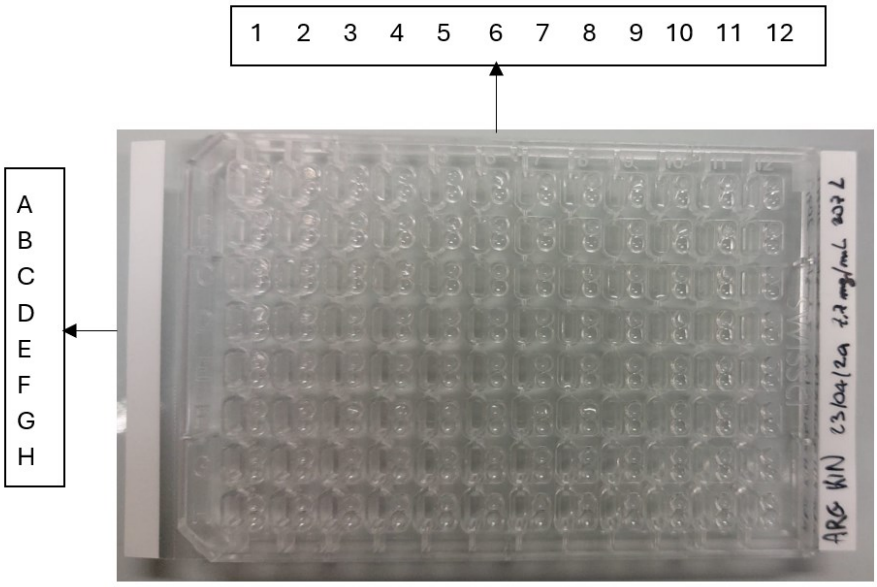

Figure 7: Plate used for the crystallization. On top there are the numbers, on the left there are the letters. Plate CS-207 L, 7,7 mg/mL of 23/04/2024

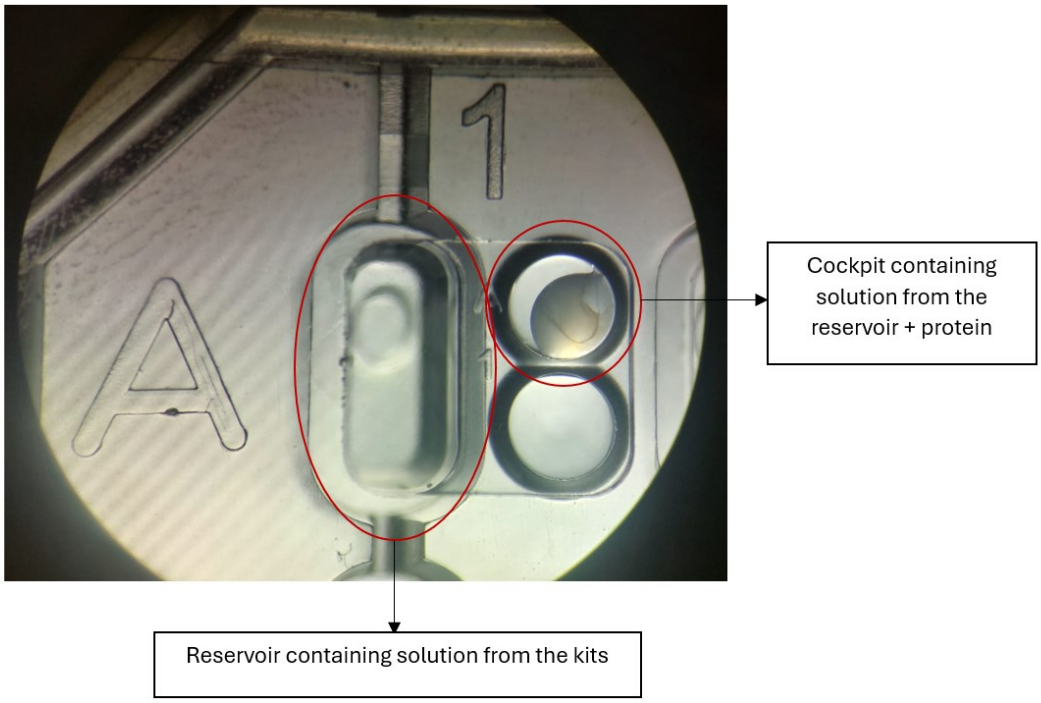

Figure 8: Visualization of the plate from the microscope, cockpit A1

The initial concentration of the AK was 7,7 mg/mL but for the plate CS 203-L and CS 206-L was used a concentration of 7 mg/mL because it was another batch not from the original.
The kit CS 201-L was carry out putting 100 µL in the reservoir, 1 µL of AK 7,7 mg/mL
The kit CS 202-L was carry out putting 100 µL in the reservoir, 1 µL of AK 7,7 mg/mL.
The kit CS 203-L was carry out putting 100 µL in the reservoir, 1 µL of AK 7 mg/mL.
The kit CS 206-L was carry out putting 100 µL in the reservoir, 1 µL of AK 7 mg/mL.
The kit CS 207-L was carry out putting 100 µL in the reservoir, 1 µL of AK 7,7 mg/mL.

Table 8: quantity used in the plate

	CS 201-L	**CS 202-L**	**CS 203-L**	**CS 206-L**	**CS 207-L**
Reservoir	100 µL	100 µL	100 µL	100 µL	100 µL
AK 7,7 mg/mL	1 µL	1 µL			1 µL
AK 7 mg/mL			1 µL	1 µL	
Reservoir in the cockpit	1 µL	1 µL	1 µL	1 µL	1 µL

From the kit CS 201-L the crystals were obtained from the condition C4 and F8, Figure 9, Figure 10.
The condition used are listed in Table 9.

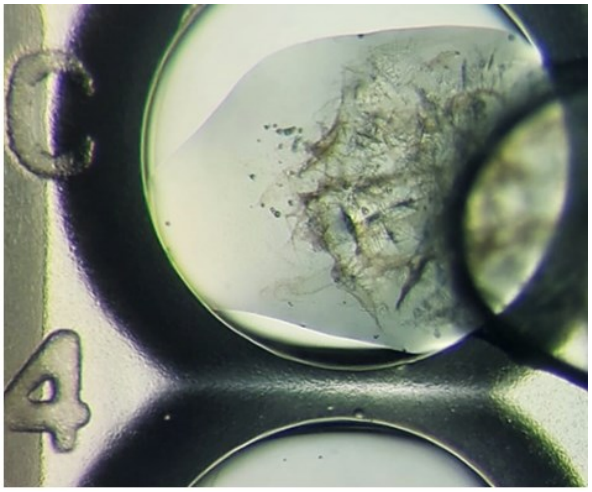

Figure 9: crystal formed from the condition C4 CS 201-L

Figure 10: crystal formed from the condition F8 CS 201-L

From the other kits the crystals formed were not in perfect condition, so just the conditions C4 and F8 from the kit CS 201-L were analyzed.

Table 9: crystal conditions from CS 201-L

	C4	**F8**
Precipitant 1	20% W/v PEG 4000	30% W/V PEG 6000
Precipitant 2	none	1,00 M Lithium chloride
Buffer	100 mM TRIS, pH 8,5	none
Additive	200mM Lithium sulphate	100 mM Sodium acetate

The condition C4 was performed under pH 7,4 and pH 8,5, PEG 4000 from 16% W/v to 27% W/v, Lithium sulphate from 180 mM to 210 mM, the concentration of TRIS was 100 mM, Figure 11. The plate contained 100 µL of reservoir and 1 µL of 3,5 mg/mL of AK.
Unfortunately the plate did not produce any crystal.

Vf (ul) in eppendorf 1000				PEG 4000																			
				1		2		3		4		5		6		7		8		9		10		11		12	
Ci [mM]	Compound			CifmM]	Vi [ul]	CifmM]	Vi [ul]	CifmM]	Vi [ul]	CifmM]	Vi [ul]	CifmM]	Vi [ul]	CifmM]	Vi [ul]	CifmM]	Vi [ul]	CifmM]	Vi [ul]	CifmM]	Vi [ul]	CifmM]	Vi [ul]	CifmM]	Vi [ul]
50	EG 4000 (wiv)	pH 7,4	A	16,00	320,00	17,00	340,00	18,00	360,00	19,00	380,00	20,00	400,00	21,00	420,00	22,00	440,00	23,00	460,00	24,00	480,00	25,00	500,00	26,00	520,00	27,00	540,00
2400	thium sulpha			180,00	75,00	180,00	75,00	180,00	75,00	180,00	75,00	180,00	75,00	180,00	75,00	180,00	75,00	180,00	75,00	180,00	75,00	180,00	75,00	180,00	75,00	180,00	75,00
1000	TRIS			100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	H2O				505,00		485,00		465,00		445,00		425,00		405,00		385,00		365,00		345,00		325,00		305,00		285,00
50	EG 4000 (wiv)		B	16,00	320,00	17,00	340,00	18,00	360,00	19,00	380,00	20,00	400,00	21,00	420,00	22,00	440,00	23,00	460,00	24,00	480,00	25,00	500,00	26,00	520,00	27,00	540,00
2400	thium sulpha			190,00	79,17	190,00	79,17	190,00	79,17	190,00	79,17	190,00	79,17	190,00	79,17	190,00	79,17	190,00	79,17	190,00	79,17	190,00	79,17	190,00	79,17	190,00	79,17
1000	TRIS			100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	H2O				500,83		480,83		460,83		440,83		420,83		400,83		380,83		360,83		340,83		420,83		400,83		380,83
50	EG 4000 (wiv)		C	16,00	320,00	17,00	340,00	18,00	360,00	19,00	380,00	20,00	400,00	21,00	420,00	22,00	440,00	23,00	460,00	24,00	480,00	25,00	500,00	26,00	520,00	27,00	540,00
2400	thium sulpha			200,00	83,33	200,00	83,33	200,00	83,33	200,00	83,33	200,00	83,33	200,00	83,33	200,00	83,33	200,00	83,33	200,00	83,33	200,00	83,33	200,00	83,33	200,00	83,33
1000	TRIS			100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	H2O				496,67		476,67		456,67		436,67		416,67		396,67		476,67		456,67		436,67		416,67		396,67		276,67
50	EG 4000 (wiv)		D	16,00	320,00	17,00	340,00	18,00	360,00	19,00	380,00	20,00	400,00	21,00	420,00	22,00	440,00	23,00	460,00	24,00	480,00	25,00	500,00	26,00	520,00	27,00	540,00
2400	thium sulpha			210,00	87,50	210,00	87,50	210,00	87,50	210,00	87,50	210,00	87,50	210,00	87,50	210,00	87,50	210,00	87,50	210,00	87,50	210,00	87,50	210,00	87,50	210,00	87,50
1000	TRIS			100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	H2O				492,50		472,50		552,50		532,50		512,50		492,50		472,50		452,50		432,50		412,50		392,50		272,50
50	EG 4000 (wiv)		E	16,00	320,00	17,00	340,00	18,00	360,00	19,00	380,00	20,00	400,00	21,00	420,00	22,00	440,00	23,00	460,00	24,00	480,00	25,00	500,00	26,00	520,00	27,00	540,00
2400	thium sulpha			180,00	75,00	180,00	75,00	180,00	75,00	180,00	75,00	180,00	75,00	180,00	75,00	180,00	75,00	180,00	75,00	180,00	75,00	180,00	75,00	180,00	75,00	180,00	75,00
1000	TRIS			100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	H2O				505,00		485,00		465,00		545,00		525,00		505,00		485,00		465,00		445,00		425,00		405,00		285,00
50	EG 4000 (wiv)		F	16,00	320,00	17,00	340,00	18,00	360,00	19,00	380,00	20,00	400,00	21,00	420,00	22,00	440,00	23,00	460,00	24,00	480,00	25,00	500,00	26,00	520,00	27,00	540,00
2400	thium sulpha			190,00	79,17	190,00	79,17	190,00	79,17	190,00	79,17	190,00	79,17	190,00	79,17	190,00	79,17	190,00	79,17	190,00	79,17	190,00	79,17	190,00	79,17	190,00	79,17
1000	TRIS			100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	H2O				500,83		480,83		560,83		540,83		520,83		500,83		480,83		460,83		440,83		420,83		400,83		280,83
50	EG 4000 (wiv)	pH 8,5	G	16,00	320,00	17,00	340,00	18,00	360,00	19,00	380,00	20,00	400,00	21,00	420,00	22,00	440,00	23,00	460,00	24,00	480,00	25,00	500,00	26,00	520,00	27,00	540,00
2400	thium sulpha			200,00	83,33	200,00	83,33	200,00	83,33	200,00	83,33	200,00	83,33	200,00	83,33	200,00	83,33	200,00	83,33	200,00	83,33	200,00	83,33	200,00	83,33	200,00	83,33
1000	TRIS			100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	H2O				496,67		476,67		556,67		536,67		516,67		396,67		476,67		456,67		336,67		416,67		396,67		276,67
50	EG 4000 (wiv)		H	16,00	320,00	17,00	340,00	18,00	360,00	19,00	380,00	20,00	400,00	21,00	420,00	22,00	440,00	23,00	460,00	24,00	480,00	25,00	500,00	26,00	520,00	27,00	540,00
2400	thium sulpha			210,00	87,50	210,00	87,50	210,00	87,50	210,00	87,50	210,00	87,50	210,00	87,50	210,00	87,50	210,00	87,50	210,00	87,50	210,00	87,50	210,00	87,50	210,00	87,50
1000	TRIS			100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00
	H2O				492,50		472,50		452,50		532,50		512,50		492,50		472,50		452,50		432,50		412,50		292,50		272,50

Figure 11: optimized condition C4 from the plate CS 201-L

The condition F8 was optimized by using PEG 6000 from 20% to 35%, lithium chloride from 600 mM to 1300 mM, the concentration of sodium acetate was kept at 100 mM.
The plate contained 75 µL of reservoir and 0,75 µL of 3,5 mg/mL of AK.
Several crystals developed, the condition considered for the data collection was the A6, Table 10, Figure 12.

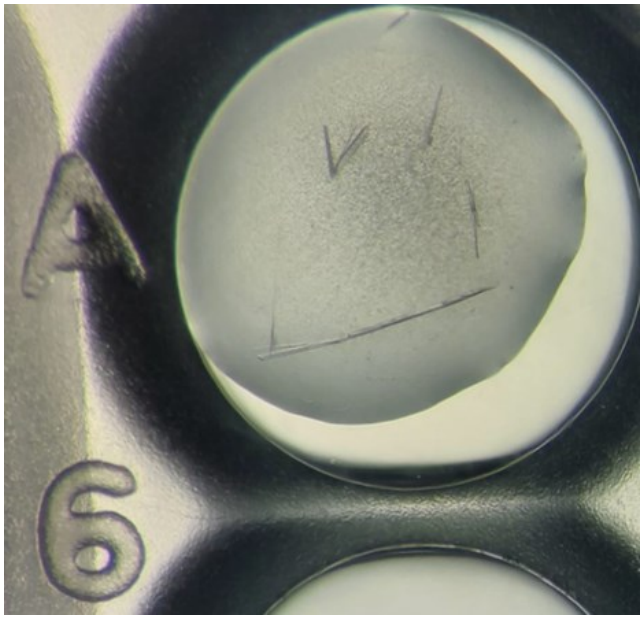

Figure 12: Crystal formed in A6 from optimized condition F8

Table 10: A6 condition from the new plate of F8

	A6
Precipitant 1	29% W/V PEG 6000
Precipitant 2	600 mM Lithium chloride
Buffer	none
Additive	100 mM Sodium acetate

		Vi(ul) in eppendorf 1000																																			
				1		2		3		4		5		6		7		8		9		10		11		12	
Ci[mM]	Compound			Ci[mM]	Vi[uL]	Ci[mM]	Vi[uL]	Ci[mM]	Vi[uL]	Ci[mM]	Vi[uL]	Ci[mM]	Vi[uL]	Ci[mM]	Vi[uL]	Ci[mM]	Vi[uL]	Ci[mM]	Vi[uL]	Ci[mM]	Vi[uL]	Ci[mM]	Vi[uL]	Ci[mM]	Vi[uL]	Ci[mM]	Vi[uL]
50	PEG 6000 (w/v)%			20	400	22	440	24	480	26	520	28	560	29	580	30	600	31	620	32	640	33	660	34	680	35	700
5000	Lituim chloride			600	120	600	120	600	120	600	120	600	120	600	120	600	120	600	120	600	120	600	120	600	120	600	120
4000	Sodium acetate	A		100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25
					455		415		375		335		295		275		255		235		215		195		175		155
50	PEG 6000 (w/v)%			20	400	22	440	24	480	26	520	28	560	29	580	30	600	31	620	32	640	33	660	34	680	35	700
5000	Lituim chloride			700	140	700	140	700	140	700	140	700	140	700	140	700	140	700	140	700	140	700	140	700	140	700	140
4000	Sodium acetate	B		100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25
					435		395		355		315		275		255		235		215		195		175		155		135
50	PEG 6000 (w/v)%			20	400	22	440	24	480	26	520	28	560	29	580	30	600	31	620	32	640	33	660	34	680	35	700
5000	Lituim chloride			800	160	800	160	800	160	800	160	800	160	800	160	800	160	800	160	800	160	800	160	800	160	800	160
4000	Sodium acetate	C		100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25
					415		375		335		295		255		235		215		195		175		155		135		115
50	PEG 6000 (w/v)%			20	400	22	440	24	480	26	520	28	560	29	580	30	600	31	620	32	640	33	660	34	680	35	700
5000	Lituim chloride			900	180	900	180	900	180	900	180	900	180	900	180	900	180	900	180	900	180	900	180	900	180	900	180
4000	Sodium acetate	D		100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25
					395		355		315		275		235		215		195		175		155		135		115		95
50	PEG 6000 (w/v)%			20	400	22	440	24	480	26	520	28	560	29	580	30	600	31	620	32	640	33	660	34	680	35	700
5000	Lituim chloride			1000	200	1000	200	1000	200	1000	200	1000	200	1000	200	1000	200	1000	200	1000	200	1000	200	1000	200	1000	200
4000	Sodium acetate	E		100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25
					375		335		295		255		215		195		175		155		135		115		95		75
50	PEG 6000 (w/v)%			20	400	22	440	24	480	26	520	28	560	29	580	30	600	31	620	32	640	33	660	34	680	35	700
5000	Lituim chloride			1100	220	1100	220	1100	220	1100	220	1100	220	1100	220	1100	220	1100	220	1100	220	1100	220	1100	220	1100	220
4000	Sodium acetate	F		100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25
					355		315		275		235		195		175		155		135		115		95		75		55
50	PEG 6000 (w/v)%			20	400	22	440	24	480	26	520	28	560	29	580	30	600	31	620	32	640	33	660	34	680	35	700
5000	Lituim chloride			1200	240	1200	240	1200	240	1200	240	1200	240	1200	240	1200	240	1200	240	1200	240	1200	240	1200	240	1200	240
4000	Sodium acetate	G		100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25
					335		295		255		215		175		155		135		115		95		75		55		35
50	PEG 6000 (w/v)%			20	400	22	440	24	480	26	520	28	560	29	580	30	600	31	620	32	640	33	660	34	680	35	700
5000	Lituim chloride			1300	260	1300	260	1300	260	1300	260	1300	260	1300	260	1300	260	1300	260	1300	260	1300	260	1300	260	1300	260
4000	Sodium acetate	H		100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25	100	25
					315		275		235		195		155		135		115		95		75		55		35		15
				20		22		24		26		28		29		30		31		32		33		34		35	

Figure 13: optimized condition F8 from the plate CS 201-L

3.5 **Data collection**

Crystals of native AK were transported to Elettra synchrotron in plates, mounted in nylon loops, flash-frozen in liquid nitrogen and transferred to a Unipuck for diffraction measurements. Given the crystallization conditions no further cryoprotection was needed. Diffraction data were collected at beamline XRD2, Elettra Synchrotron Trieste, Italy (Lausi et al., 2015). Using X-rays of wavelength 1 Å and size 100µm x 100µm, 360° data were collected with an oscillation of 0.5° on a Dectris Pilatus 6M detector, for a total time of 360s.

Data were processed using XDS (Kabsch, 2010), yielding a dataset up to 1.82 Å.

Data collection statistics are summarized in Table 11.

The diffraction pattern is highlighted in Figure 14.

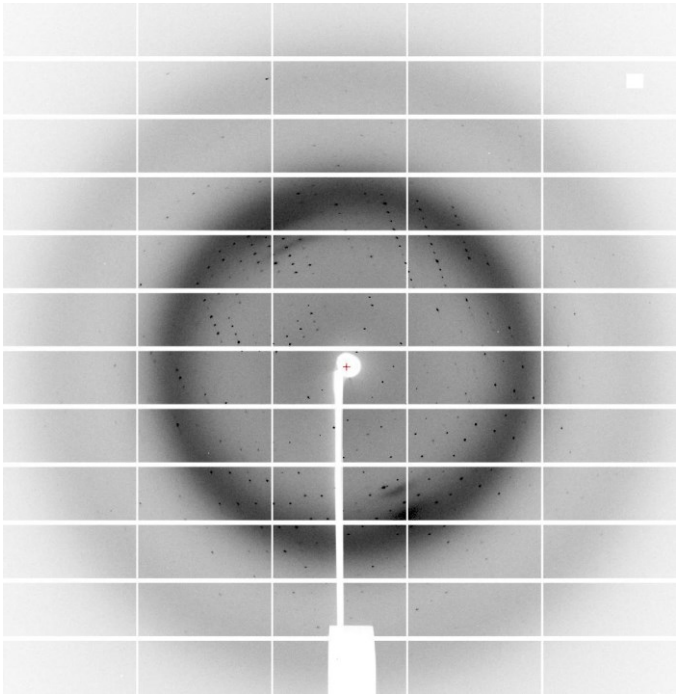

Figure 14: Diffraction pattern from the data collection of AK

Table 11: Crystallographic data collection and refinement statistics for Arginine kinase from Hermetia illucens. Statistics for the highest-resolution shell are given in parentheses

	Arginine kinase
Data collection	
Beam line	Elettra XRD2
Wavelength (Å)	1

Space group	C 1 2 1
Cell parameters (a, b, c, Å)	130.41 38.85 72.96 90 113.91 90
Resolution range (Å)	37.53 - 1.82 (1.885 - 1.82)
Total reflections	189139 (27388)
Unique reflections	30332 (4852)
Redundancy	6,2(5.6)
Completeness (%)	99.78 (99.93)
Mean I/sigma(I)	12.64 (1.03)
Rmerge [b]	7.9 (148.3)
CC1/2	99.9 (52.9)
Refinement	
Reflections used in refinement	30324 (2998)
Reflections used for R-free	1517 (150)
Wilson B-factor (Å2)	38.21
Rwork [d]	0.1924 (0.4235)
Rfree [d]	0.2440 (0.4343)
Total n. of atoms [d]	3066
macromolecules	2865
ligands	5
waters	196
Protein residues	356
Matthews Coeff *(% solvent content)*	43.76
r.m.s.d. [d]	
bond length (Å)	0.010
angles (°)	1.23
Clashscore	4.21
Ramachandran [d]	

favored (%)	96.23
allowed (%)	3.77
outliers (%)	0.0

[a] Values in the highest resolution shell.
[b] $R_{merge} = \sum_{hkl} \sum_j |I_j - \langle I \rangle| / \sum_{hkl} \sum_j I_j$, where I is the intensity of a reflection, and $\langle I \rangle$ is the mean intensity of all symmetry related reflections j.
[c] $R_{p.i.m.} = \sum_{hkl} \{[1/(N-1)]^{1/2} \sum_j |I_j - \langle I \rangle|\} / \sum_{hkl} \sum_j I_j$, where I is the intensity of a reflection, and $\langle I \rangle$ is the mean intensity of all symmetry related reflections j, and N is the multiplicity. (Weiss, 2001)
[d] Calculated with PHENIX suite (Paul D. Adams et al., 2002), R_{free} is calculated using 5% of the total reflections that were randomly selected and excluded from refinement.

3.6 **Resolution of the structure**

One monomer from the crystal structure of Scylla paramamosain arginine kinase (PDB entry 5ZHQ) was prepared as starting model for molecular replacement, using CHAINSAW (N. Stein, 2008).

CHAINSAW: a program for mutating PDB files used as templates in molecular replacement from the CCP4 suite (Collaborative Computational Project, 1994).

The AK structure was solved by Molecular Replacement using MOLREP (A. Vagin and A. Teplyakov, 1997), using the above mentioned starting model, followed by rigid body and restrained refinement using Refmac5 (Murshudov et al., 2011) of the CCP4 package. Automated model building was accomplished by PHENIX suite (Paul D. Adams et al., 2002) which allowed the partial building of the main chains with one molecule in the asymmetric unit. The manual fitting of the side chains and solvent molecules into electron density maps were performed using COOT (Paul Emsleya & Kevin Cowtan, 2004) PHENIX suite (Paul D. Adams et al., 2002) and Refmac5 (Murshudov et al., 2011), while monitoring R_{work}, R_{free}, Ramachandran plot (R. A. Laskowski et al., 1993) and related geometrical parameters.

3.7 **Structure refinement protocol**

At the beginning of the structure refinement was performed an autobuild refinement, the R_{free} went from 0.2838 to 0.2762.

At first the refinement setting were: XYZ (reciprocal space); XYZ (real space); rigid body, individual B-factors; target function: optimize X-ray/stereochemistry weight, optimize X-ray/ADP weight; other option: automatically correct N/Q/h errors.

The statistic of each refinement are reported in Table 12.

Refinement 19 a sequence from aa 312-320 was eliminated because there was no map available in this region, probably because it's a very flexible sequence; the R_{free} was 0,2762 and was constant even if the structure was continuously modified by using the real space refinement zone.

For this reason in the refinement 22 the sequence 312-320 was added again while trying to find a right alignment. This stretch of amino acid 312-320 was not present in other AK from different species, while in *Scylla paramosain* the position of that stretch in AK was visible but different from the one of H. illucens and was starting to align from the amino acid 319.

The setting parameters were change by removing the optimize X-ray/stereochemistry weight, optimize X- ray/ADP weight.

In the refinement 23, the sequence of AK of BSF was checked and has been added a Methionine as 1° amino acid.

In *Scylla paramosain* the first amino acid was Alanine.

Each amino acid was fixed by checking the map, improving the angles, this allowed to reduce the R_{free}, improve the clashscore, and the Ramachandran score

From the refinement 24 to refinement 29 the structure was continuously modified to improve.

In the refinement 30 the setting were changed by adding:

- the optimize X-ray/stereochemistry weight,
- optimize X- ray/ADP weight,
- simulated annealing (cartesian).

This allowed the reduction of R_{free} to 0,262.

Refinement 31 "simulated annealing" was removed as setting

Refinement 33 the setting parameter "rigid body" was added but removed in the refinement 35 improving the Ramachandran outliers

In the refinement 38 the R_{free} decreased to 0.2628 but the Ramachandran outliers was worse and the rotamers outliers were improved.

After the further changes to the structure, could not improve anymore the R_{free}, the waters were added in the refinement 39, allowing the R_{free} to decrease from 0.2628 to 0.2353,

From the refinement 42 was removed the setting " automatically correct N/Q/H errors

In the refinement 51 was removed the setting “XYZ real-space”
In the refinement 52, 2 molecules of Chlorine (Cl) were added. There was a “green zone” that was too big to fit a water molecule. After adding the chlorine molecule the map was more confined.
The chlorine molecule was added because in the crystallization condition there was lithium chlorine as precipitant 2
Also in this refinement, 3 molecules of sodium (Na) were added, from the interaction with the water molecules nearby we saw that there were sodium molecule.
In the refinement 55 there was a problem with the CYS 271 because there was a very big map in that site. We tried to add a Cl near the cysteine to improve it but without any success due to the clashes with the cysteine
We give it two alternate conformation to the residue 21 (ASP), 124 (ARG), 270 (PHE) and 271 (CYS) because there was still a big density in this region.
For the amino acids 21, 124 270 it works, but there was still some issue with the amino acid 271.

In the refinement 58 was removed the setting “ the optimize X-ray/stereochemistry weight and optimize X- ray/ADP weight” causing an increase of the several values and R_{free}; also was moved the position of the second Chlorine ion which was near the CYS 271.
In the refinement 59 the Chlorine ion and the alternative conformation of CYS 271 was removed.
In the refinement 61 the alternative conformation of amino acid 124 (ARG) and 233 (MET) were removed.
In the refinement 64 the alternative conformation of amino acid 271 was removed.
In the total of 6 cysteine, there were 2 of the CYS in oxidized form (CSX) the amino acids 201 and 271 the catalytic site. The reason why was the “green zone” , it was too big to fit a non-oxidized cysteine, and after adding the oxidazed one the map was perfectly fitting it.
The oxidazed cysteines were added in the refinement 65.
The amino acid 271 in *Scylla paramosain* was also a cysteine but not oxidized.
In the refinement 69 was added an alternative conformation to the 21 ASP, 124 ARG, 233 MET, 271 CYS.

Table 12: Refinement protocol

	R free	Ramachandran Outliers (%)	Rotamers Outliers (%)	Clashscore	Ramachandran favored (%)
Autobuild 14	0,2762				
Refinement 19	0.2762	0.29	0.99	4.48	97.14
Refinement 22	0.2762	0.28	1.32	4.64	96.89
Refinement 23	0.2761	0.28	0.98	4.45	96.61
Refinement 29	0.265	0.29	1.64	4.28	97.46
Refinement 30	0.2632	0.28	1.31	4.28	96.89
Refinement 31	0.262	0.56	0.98	4.10	97.18
Refinement 33	0.2614	0.56	0.66	5.7	96.89
Refinement 35	0.2636	0	1.64	4.81	96.89
Refinement 38	0.2628	0.56	0.65	4.07	97
Refinement 39	0.2353	0	0.65	4.62	96.9
Refinement 42	0.2387	0	0	3.56	96.3
Refinement 50	0.2427	0	0	4	95.4
Refinement 51	0.2408	0	0.64	3.47	95.7
Refinement 52	0.2411	0	1.27	3.29	95.4
Refinement 55	0.2389	0	0.32	4	95.4
Refinement 58	0.2427	0.28	1.28	4.52	95.73
Refinement 59	0.239	0	0.64	4.54	95.44
Refinement 61	0.2421	0	0	4.79	95.76
Refinement 65	0.2417	0	0.97	3.87	95.4
Refinement 68	0.2413	0.29	1.30	4.21	95.36
Refinement 69	0.2446	0	0.32	4.21	96.23

CHAPTER 4
RESULTS

4.1 **Description of the structure**

There are 356 residues, 1 chloride ion (Cl), 4 sodium ions (Na), 196 water atoms.

The AK crystal structure shows the classic AK fold with an α-helical N-terminal domain and an α-β C-terminal domain (Yang et al., 2019).

The N-terminal domain contains five α-helixes, and the C-terminal domain contains seven α-helixes and eight β-sheets.

Molecular graphics and analyses was performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, https://www.rbvi.ucsf.edu/chimerax , (Meng et al., 2023).

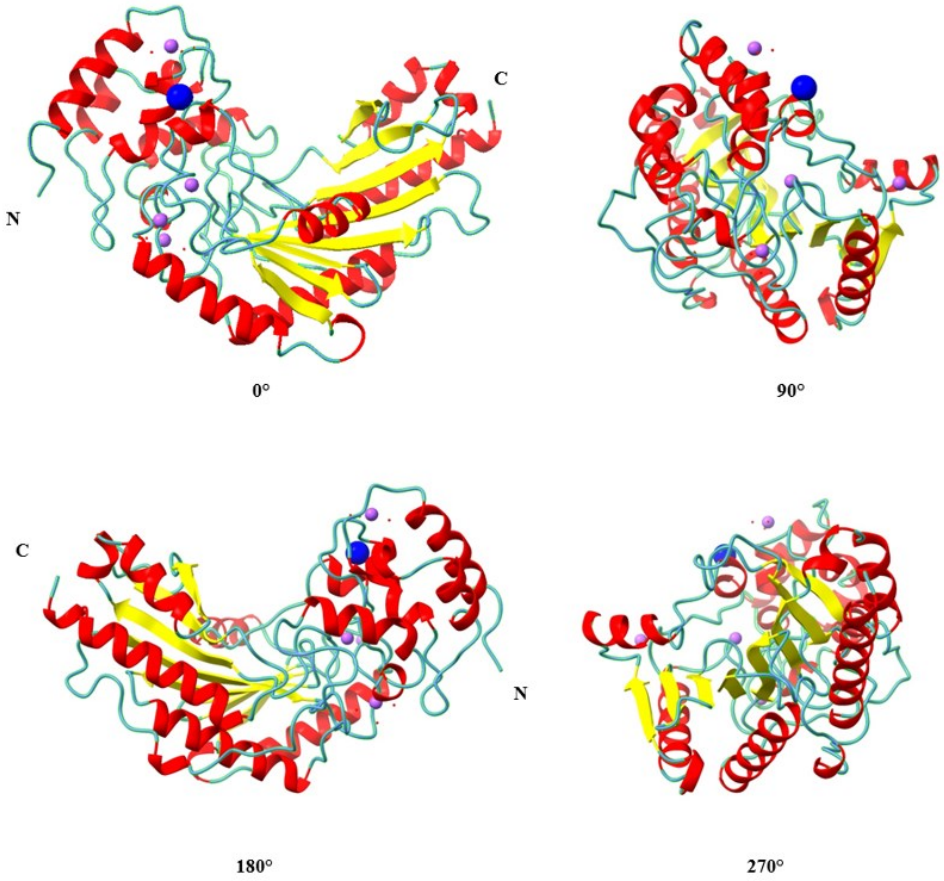

Figure 15: Structure of AK at 0, 90, 180 and 270 degrees with the α-helixes colored red, ß-sheets colored yellow, random coils colored light-blue. Chlorine ion is blue and sodium ion is purple.

From the predicted structure of AK, Figure 16, there are a lot of similarities with AK from BSF, Figure 17.

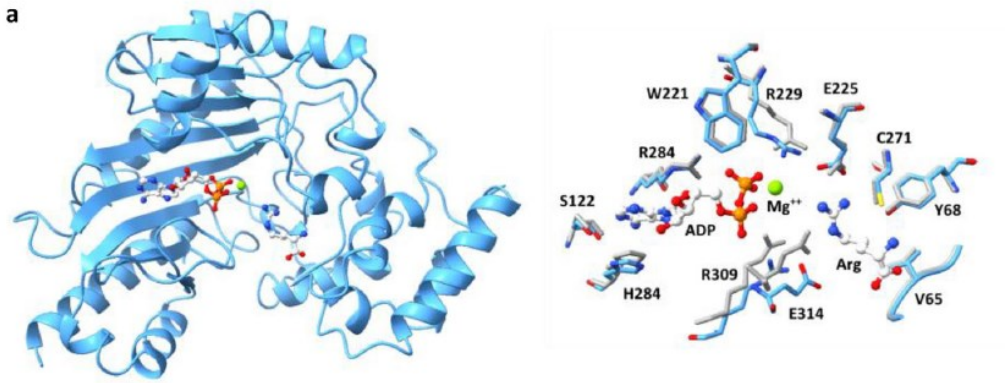

Figure 16: Structure prediction and enzymatic activity of AK X4 from BSF from AK from Litopenaeus vannamei, PDB: 4BG4 *(Delfino et al., 2024)*

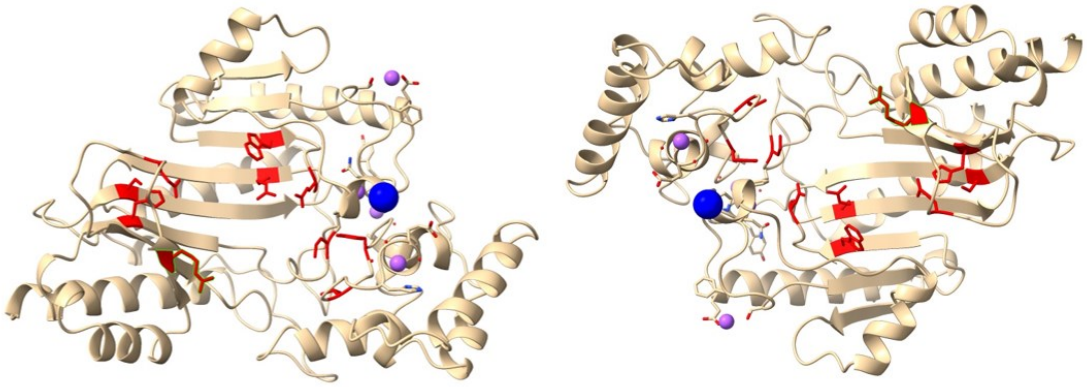

Figure 17: details of the AK X4 residues in the catalytic site (red) upside down (left) and at 0 degrees (right) from H. illucens

4.2 **Model quality**

The structure model quality can be summarized in Figure 18.

The validation report was made from **www.pdb.org** (Berman et al., 2003).

The report resolution is 1.82 Å. The R-free is 0.2446, considering other x-ray structure is good but as compared to X-ray structure at similar resolution it can still be improved.

Also the clashscore of 4.21 some enhancement can still be done in the future.

The Ramachandran outliers and sidechain outliers are perfect.

The RSRZ (Real-space R-value) has a great score.

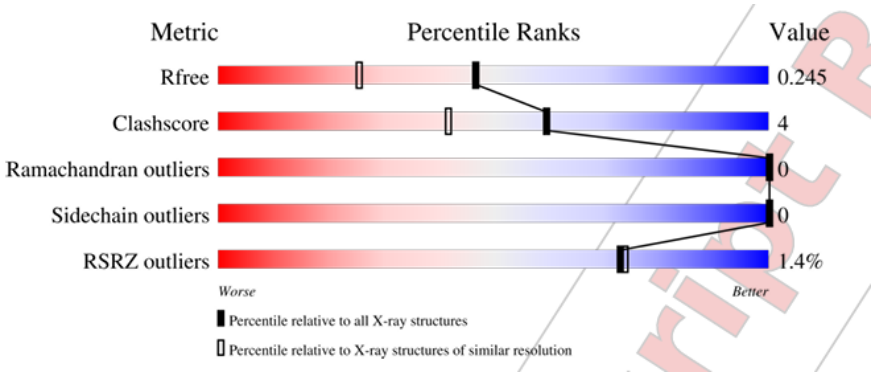

Figure 18: percentile score for global validation

R_{free} is a measure of the fit of the model to a small subset of the experimental data, which was not used in model refinement.

Clashscore derived from the number of pairs of atoms in the model that are unusually close to each other, it is expressed as the number or such clashes per thousand atoms.

A residue is considered to be a Ramachandran plot outlier if the combination of its φ and ψ torsion angles is unusual. The Ramachandran outlier score for an entry is calculated as the percentage of Ramachandran outliers with respect to the total number of residues in the entry for which the outlier assessment is available.

Protein sidechain has a preferred torsion angle value, called rotamers or rotameric conformers. Sidechain outliers is calculated as the percentage of residues with an unusual sidechain conformation with respect to the total number of residues for which the assessment is available.

Real-space R-value (RSR) is a measure of the quality of fit between a part of an atomic model (in this case, one residue) and the data in real space. The RSR Z-score (RSRZ) is a normalisation of RSR specific to a residue type and a resolution bin.

RSRZ is calculated only for standard amino acids and nucleotides in protein, DNA and RNA chains. A residue is considered an RSRZ outlier if its RSRZ value is greater than 2. The RSRZ outlier score, is calculated as the percentage RSRZ outliers with respect to the total number of residues for which RSRZ was computed.

This information is from https://www.wwpdb.org/ (Berman et al., 2003)

The quality of the chain is represented by the Figure 19. Overall is a great structure by being 88% indicating the fraction of residue that contain 0 outliers for model-only validation criteria, while the yellow region content is 1.

The Figure 20, indicates the fit to the electron density.

Mol	Chain	Length
1	A	356

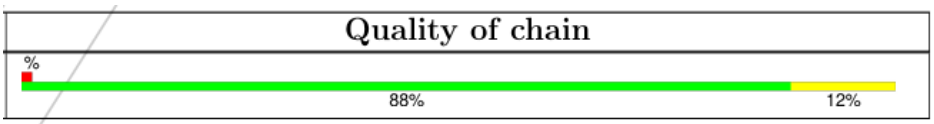

Figure 19: graphical summary of the quality of all polymeric chains

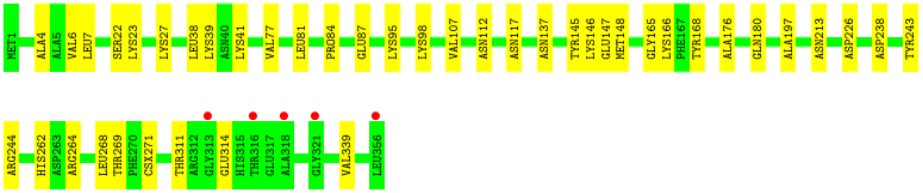

Figure 20: Residue property plots. A red dot above a residue indicates a poor fit to the electron density.

CHAPTER 5
DISCUSSION

5.1 **Comparison of epitopes from *Hermetia illucens* with other epitopes from other species**

By comparing the linear epitopes of AK from BSF with AK of other organisms, eighteen epitopes has been identified in AK of BSF, suggesting that this protein could trigger immunoreactivity, Table 13.

Only one epitope from *Scylla paramamosain,* 204-218, has been detected after digestion (Delfino et al., 2024).

Table 13: Linear epitopes from IEDB identified in AK X4 from BSF and their fate after simulated gastrointestinal digestion *(Delfino et al., 2024)*

Epitopes	Sequence	Color	Already detected in another organism	Detection after digestion	Grouping
16-18	KLQ	Black	*Crassostrea gigas*	No	1
24-26	SLL	Dim gray	*C. gigas*	No	2
127-141	CGRSMEGYPFNPCLT	Lime	*Scylla paramamosain*	No	3
130-142	SMEGYPFNPCLTE	Navy	*Oratosquilla oratoria*	No	
134-148	YPFNPCLTEAQYKEM	Pale green	*S. paramamosain*	No	
162-164	ELK	Magenta	*C. gigas*	No	4
181-195	LIDDHFLFKEGDRFL	Yellow	*P. vannamei*	No	5
184-198	DHFLFKEGDRFLQAA	Red	*P. vannamei*	No	
188-202	FKEGDRFLQAANACR	Purple	*Dermatophagoides pteronyssinus*	No	
193-205	RFLQAANACRFWP	Brown saddle	*O. oratoria*	No	
204-218	**WPTGRGIYHNDNKTF**	Cyan	***S. paramamosain***	**Yes**	
211-225	YHNDNKTFLVWCNEE	Forest green	*S. paramamosain*	No	
216-230	KTFLVWCNEEDHLRI	Blue	*Blattella germanica*	No	

263-277	DRLGFLTFCPTNLGT	Cornflower blue	*D. pteronyssinus*	No	6
310-320	VRGTRGEHTEAEG	Dark olive green	*O. oratoria*	No	7
314-316	EHT	Hot pink	*C. gigas*	No	
319-333	EGGIYDISNKRRMGL	Orange	*Penaeus vannamei*	No	
321-335	GIYDISNKRRMGLTE	Pale pink	*O. oratoria*	No	

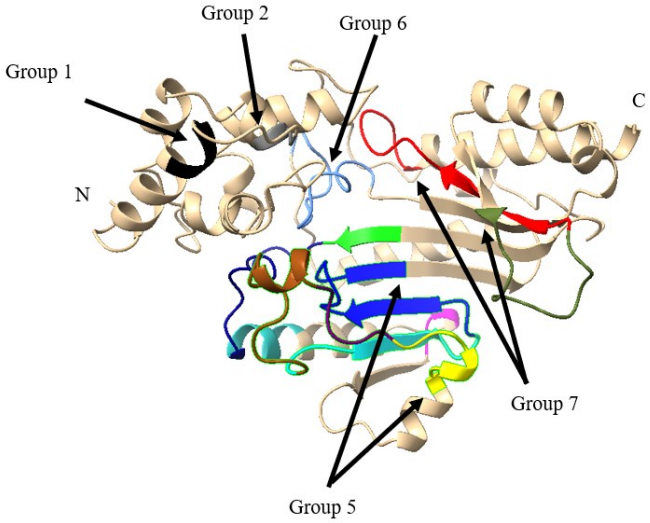

Figure 21: Grouping of the overlapping epitopes in AK from Hermetia illucens

Group 1, 16-18; Group 2, 24-26; Group 5, 181-230; Group 6, 263-277; Group 7, 310-355

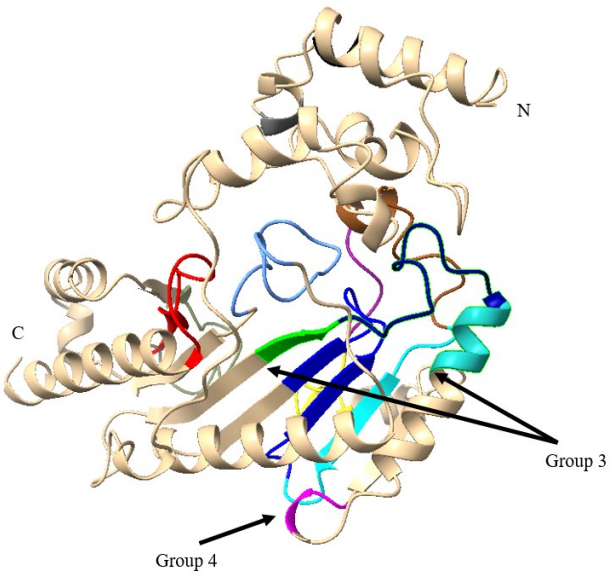

Figure 22: Grouping of the overlapping epitopes in AK from H. illucens

Group 3, 127-148; Group 4, 162-164

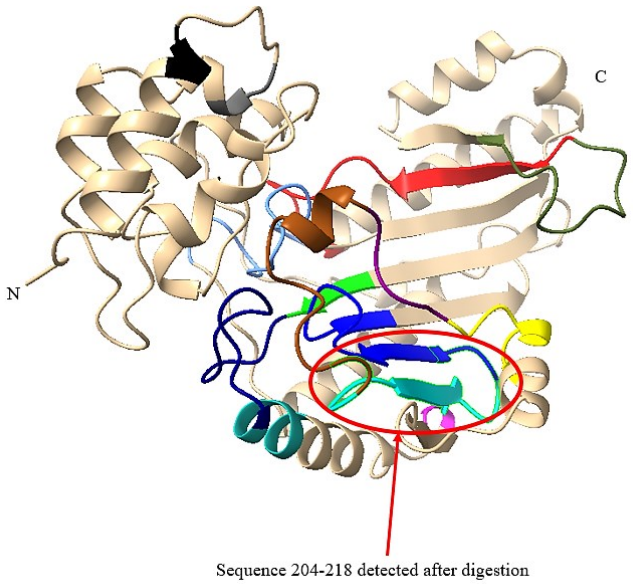

Figure 23: Highlighting the sequence 204-205 detected after digestion in AK from H. illucens

5.2 **Location of epitopes on the AK structure from *Hermetia illucens***

The epitopes, their sequence and their colors are summarized in Table 13.

Some of the epitopes are inside the structure so less likely to induce the allergenic reaction, like the lime, cyan and blue region, respectively the sequence: 127-141, group 3; 204-218, group 5 Figure 23; 216-230, group 5.

There are other epitopes on the outside of the structure more prone to induce the allergenic reaction like black, dim gray, navy, magenta, purple, brown saddle, dark olive green, hot pink, respectively the sequence: 16-18, group 1; 24-26 group 2; 130-142, group 3; 162-164, group 4; 188-202, 193-205, 310-320, group 5; 314-316, group 7.

The groups are highlighted in Figure 21 and Figure 22.

CONCLUSIONS AND FUTURE PERSPECTIVES

The AK protein, expressed via recombinant techniques, is fully folded.

The protein could be crystallized and the structure was revealed by X-ray diffraction to a resolution of 1.82 Å.

The fold of AK is conserved with the fold of analog proteins from different species like *Scylla paramamosain, Penaeus vannamei* and *Blattella germanica* as predicted by *Delfino et al., 2024*

Today the AK allergy is registered with the World Health Organization and International Union of Immunological Societies (WHO/IUIS). AK is one of the five allergens present in *Penaeus monodon*, one of the most widely farmed and consumed shrimps worldwide. Among all registered allergens, tropomyosin is the major allergen of shrimp-sensitized patients; however, there is an increase of the sensitization rate of AK allergy (Wai et al., 2022). In this study, the structure of AK from BSL highlighted the presence of several epitopes on the surface of the protein, and one of them still survive to the digestion, representing an issue for people with sensitivity to crustaceans or mites.

In fact, similarities in tridimensional structure and repeated positioning of epitopes among allergens correlate with its allergenic activity. In case of AK proteins from H. illucens and *Penaeus vannamei* which do share structural similarities, the cross-reactivity will be likely.

Hermetia illucens may be a source of allergenic AK for individuals with sensitivity to crustaceans or mites, but further studies are needed to evaluate the stability of AK in food matrices under different processing conditions typical of the food industry, by using high temperature.

At 60°C, pH 7 there is a complete unfolding of the protein and a complete degradation of the conformational epitopes; by coming back at 20°C the unfolding will be 90% (Delfino et al., 2024).

However, the linear epitope, 204-218, survived to the digestion, could still survive at higher temperatures, so the risk of allergenicity would still be presence.

To unsure the total degradation of the linear epitope further food preparation studies must be evaluated according to the residual allergenic activity. For instance enzymatic degradation

could be explored.

Today H. illucens has not been approved as food, only as feed for aquaculture, poultry and swine animals, but several studies are still in progress and understanding the allergenicity is an important step to allow a safe commercialization as food.

At the moment the only safe measure to apply is avoid any contact with BSFL preparations in case of a certified allergy to crustaceans and mites by the consumer.

BIBLIOGRAPHY

A. Vagin and A. Teplyakov. (1997). *MOLREP: an Automated Program for Molecular Replacement*. https://doi.org/10.1107/S0021889897006766

Berman, H., Henrick, K., & Nakamura, H. (2003). *Announcing the worldwide Protein Data Bank*. *10*. https://doi.org/10.1038/nsb1203-980

Bessa, L. W., Pieterse, E., Marais, J., Dhanani, K., & Hoffman, L. C. (2021). Food safety of consuming black soldier fly (Hermetia illucens) larvae: Microbial, heavy metal and cross-reactive allergen risks. *Foods*, *10*(8). https://doi.org/10.3390/foods10081934

Broekman, H. C. H. P., Knulst, A. C., de Jong, G., Gaspari, M., den Hartog Jager, C. F., Houben, G. F., & Verhoeckx, K. C. M. (2017). Is mealworm or shrimp allergy indicative for food allergy to insects? *Molecular Nutrition and Food Research*, *61*(9). https://doi.org/10.1002/mnfr.201601061

Collaborative Computational Project, N. 4. (1994). *The CCP4 suite: programs for protein crystallography*. https://doi.org/10.1107/S0907444994003112

De Smet, J., Wynants, E., Cos, P., Leen, C., & Campenhout, V. (2018). *Microbial Community Dynamics during Rearing of Black Soldier Fly Larvae (Hermetia illucens) and Impact on Exploitation Potential*. https://doi.org/10.1128/AEM

Delfino, D., Prandi, B., Calcinai, L., Ridolo, E., Dellafiora, L., Pedroni, L., Nicoletta, F., Cavazzini, D., Tedeschi, T., & Folli, C. (2024). Molecular Characterization of the Allergenic Arginine Kinase from the Edible Insect Hermetia illucens (Black Soldier Fly). *Molecular Nutrition and Food Research*, *68*(9). https://doi.org/10.1002/mnfr.202300911

Diclaro, J. W., & Kaufman, P. E. (2009). *Black soldier fly Hermetia illucens Linnaeus (Insecta: Diptera: Stratiomyidae) 1*. http://creatures.ifas.ufl.edu.

Fao. (2013a). *Edible Insects - Future prospects for food and feed security*.

Fao. (2013b). *The conTribuTion of insecTs To food securiTy, livelihoods and The environmenT 1 Why insecTs?* www.fao.org/forestry/edibleinsects/en/

Fu, L., Wang, J., Ni, S., Wang, C., & Wang, Y. (2018). Identification of Allergenic Epitopes and Critical Amino Acids of Major Allergens in Chinese Shrimp (Penaeus chinensis) by Immunoinformatics Coupled with Competitive-Binding Strategy. *Journal of*

Agricultural and Food Chemistry, *66*(11), 2944–2953. https://doi.org/10.1021/acs.jafc.7b06042

Gligorescu, A., Toft, S., Hauggaard-Nielsen, H., Axelsen, J. A., & Nielsen, S. A. (2019). Development, growth and metabolic rate of Hermetia illucens larvae. *Journal of Applied Entomology*, *143*(8), 875–881. https://doi.org/10.1111/jen.12653

Kabsch, W. (2010). *XDS*. *Acta Crystallographica Section D Biological Crystallography*, *66*(2), 125–132. https://doi.org/10.1107/S0907444909047337

Lausi, A., Polentarutti, M., Onesti, S., Plaisier, J. R., Busetto, E., Bais, G., Barba, L., Cassetta, A., Campi, G., Lamba, D., Pifferi, A., Mande, S. C., Sarma, D. D., Sharma, S. M., & Paolucci, G. (2015). Status of the crystallography beamlines at Elettra. *European Physical Journal Plus*, *130*(3). https://doi.org/10.1140/epjp/i2015-15043-3

Leni, G., Tedeschi, T., Faccini, A., Pratesi, F., Folli, C., Puxeddu, I., Migliorini, P., Gianotten, N., Jacobs, J., Depraetere, S., Caligiani, A., & Sforza, S. (2020). Shotgun proteomics, in-silico evaluation and immunoblotting assays for allergenicity assessment of lesser mealworm, black soldier fly and their protein hydrolysates. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-57863-5

Leoni, C., Volpicella, M., Dileo, M. C. G., Gattulli, B. A. R., & Ceci, L. R. (2019). Chitinases as food allergens. *Molecules*, *24*(11). https://doi.org/10.3390/molecules24112087

Marzouk, S. (2016). *Physiological, ethological and ecological features of Hermetia illucen*. https://doi.org/10.7287/peerj.preprints.2436v1

Meng, E. C., Goddard, T. D., Pettersen, E. F., Couch, G. S., Pearson, Z. J., Morris, J. H., & Ferrin, T. E. (2023). UCSF ChimeraX: Tools for structure building and analysis. *Protein Science*, *32*(11). https://doi.org/10.1002/pro.4792

Mézes, M. (2018). Food safety aspect of insects: A review. In *Acta Alimentaria* (Vol. 47, Issue 4, pp. 513–522). Akademiai Kiado Rt. https://doi.org/10.1556/066.2018.47.4.15

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., & Vagin, A. A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, *67*(4), 355–367. https://doi.org/10.1107/S0907444911001314

N. Stein. (2008). *CHAINSAW: a program for mutating pdb files used as templates in molecular replacement*. https://doi.org/10.1107/S0021889808006985

Oliveira, F., Doelle, K., List, R., & O'reilly, J. R. (2015). Assessment of Diptera: Stratiomyidae, genus Hermetia illucens (L., 1758) using electron microscopy. *~ 147 ~ Journal of Entomology and Zoology Studies*, *3*(5), 147–152.

Pan, J., Xu, H., Cheng, Y., Mintah, B. K., Dabbour, M., Yang, F., Chen, W., Zhang, Z., Dai, C., He, R., & Ma, H. (2022). Recent Insight on Edible Insect Protein: Extraction, Functional Properties, Allergenicity, Bioactivity, and Applications. In *Foods* (Vol. 11, Issue 19). MDPI. https://doi.org/10.3390/foods11192931

Paul D. Adams, Ralf W. Grosse-Kunstleve, Li-Wei Hung, Thomas R. Ioerger, Airlie J. McCoy Nigel, W. Moriarty, Randy J. Read, James C. Sacchettini, Nicholas K. Sautera, & Thomas C. Terwilligerf. (2002). *PHENIX: building new software for automated crystallographic structure determination*. *58*, 1948–1954. https://doi.org/10.1107/S0907444902016657

Paul Emsleya, & Kevin Cowtan. (2004). *Coot: model-building tools for molecular graphics*. *60*, 2126–2132. https://doi.org/10.1107/S0907444904019158

R. A. Laskowski, M. W. MacArthur, D. S. Moss, & J. M. Thornton. (1993). *PROCHECK: a program to check the stereochemical quality of protein structures*. https://doi.org/10.1107/S0021889892009944

R-biopharm. (2022, January 4). *Allergeni alimentari*. https://food.r-biopharm.com/it/news/insetti-commestibili-prelibatezza-o-rischio-di-allergia/

Shen, Y., Cao, M. J., Cai, Q. F., Su, W. J., Yu, H. L., Ruan, W. W., & Liu, G. M. (2011). Purification, cloning, expression and immunological analysis of Scylla serrata arginine kinase, the crab allergen. *Journal of the Science of Food and Agriculture*, *91*(7), 1326–1335. https://doi.org/10.1002/jsfa.4322

Wai, C. Y. Y., Leung, N. Y. H., Leung, A. S. Y., Ngai, S. M., Pacharn, P., Yau, Y. S., Rosa Duque, J. S. Da, Kwan, M. Y. W., Jirapongsananuruk, O., Chan, W. H., Chua, G. T., Lee, Q. U., Piboonpocanun, S., Ho, P. K., Wong, J. S. C., Li, S., Xu, K. J. Y., Wong, G. W. K., Chu, K. H., … Leung, T. F. (2022). Comprehending the allergen repertoire of shrimp for precision molecular diagnosis of shrimp allergy. *Allergy: European Journal of Allergy and Clinical Immunology*, *77*(10), 3041–3051. https://doi.org/10.1111/all.15370

Wang, Y. S., & Shelomi, M. (2017). Review of black soldier fly (Hermetia illucens) as animal feed and human food. In *Foods* (Vol. 6, Issue 10). MDPI AG. https://doi.org/10.3390/foods6100091

Yang, Y., Cao, M. J., Alcocer, M., Liu, Q. M., Fei, D. X., Mao, H. Y., & Liu, G. M. (2015). Mapping and characterization of antigenic epitopes of arginine kinase of Scylla paramamosain. *Molecular Immunology*, *65*(2), 310–320. https://doi.org/10.1016/j.molimm.2015.02.010

Yang, Y., Liu, G. Y., Yang, H., Hu, M. J., Cao, M. J., Su, W. J., Jin, T., & Liu, G. M. (2019). Crystal structure determination of Scylla paramamosain arginine kinase, an allergen that may cause cross-reactivity among invertebrates. *Food Chemistry*, *271*, 597–605. https://doi.org/10.1016/j.foodchem.2018.08.003

Zhang, Y. X., Chen, H. L., Maleki, S. J., Cao, M. J., Zhang, L. J., Su, W. J., & Liu, G. M. (2015). Purification, Characterization, and Analysis of the Allergenic Properties of Myosin Light Chain in Procambarus clarkii. *Journal of Agricultural and Food Chemistry*, *63*(27), 6271–6282. https://doi.org/10.1021/acs.jafc.5b01318

ATTACHMENTS

Regulation EU 2011/1169

Regulation EU 2015/2283

PDB validation report

ACKNOWLEDGEMENTS

This master's degree course has finally come to an end, with relieve and sadness for all that has been in this two years.

First of all I would like to thanks my supervisor Prof. Michele Cianci, with his support and knowledge during the work in this thesis.

X-ray diffraction data were collected at the Elettra Synchrotron Trieste storage ring (Basovizza, Italy) under the beam time award number 20240567.

In particular, I would like to thank the staff of the XRD2 beamline and Dr. Raghurama P Hegde for his assistance and support.

I would like to thank Prof.ssa Claudia Folli (Department of Food and Drug, University of Parma, IT) for the provision of recombinant AK protein from H. illucens used for the research reported in this thesis.

I would like to thank all the people of the Biochemistry Laboratory of the D3A, in particular Maria Claudia Cedri for being by my side in the laboratory or outside for a break.

I would like to thanks my parents for letting me do this journey and for their help of everyday.

A thanks goes to my boyfriend Carlo, for his encouragement and support to keep me going every day, for being there every happy and unhappy moment, I couldn't have wished for anyone better.

A thanks goes to my colleagues, Erica, Noemi, Angelica and Carla for making my days more brightful and cheerful in these years.

I would like to thanks all my friends outside the university who shared joys, sacrifices and successes with me every day without ever turning their backs on me. Thank you for never leaving me alone.