UNIVERSITY OF CRETE

DEPARTMENT OF BIOLOGY AND DEPARTMENT OF CHEMISTRY


POSTGRADUATE PROGRAM “PROTEIN BIOTECHNOLOGY”

MSc Thesis

# *In silico* and experimental study for the identification of new sweet proteins as additives

Author: Lazou Thomai (AM 944)
Supervisor: Assistant Professor Ioannis Pavlidis

Heraklion, 2022

## Examination Committee

Pavlidis Ioannis (supervisor)

Assistant Professor, Department of Chemistry, University of Crete

Tsiotis Georgios
Professor, Department of Chemistry, University of Crete

Velonia Kelly
Assistant Professor, Department of Materials Science and Technology, University of Crete

# Table of contents


Acknowledgments....6

Abstract....7

Περίληψη....8

1.Introduction....12

1.1 Sugar and human health....12

1.2 Requirements for the ideal sweet-tasting compound....13

1.3 Sweeteners: Natural carbohydrates, artificial sweeteners and natural sweeteners....14

1.5 Sweet taste receptor and gustatory function....21

1.6 Sweet taste receptor anatomy....22

1.7 Wedge model: a mechanism for sweet taste sensation....23

1.8 Sweet taste in mammals....24

1.9 *Drosophila melanogaster*: a model organism to study gustation....25

1.10 *Drosophila* life cycle....26

1.11 Aim of the thesis....27

2. Materials....28

2.1 Scientific literature for meta-analysis....28

2.2 Bioinformatic tools and other software....28

2.2.1 Protein Data Bank....28

2.2.2 Blastp in NCBI database....28

2.2.3 Geneious....29

2.2.4 YASARA....29

2.2.5 MolProbity....29

2.2.6 PyMol....29

2.2.7 ClusPro....30

2.3 Genes of interest....30

2.4 Proteins....30

2.5 Rearing conditions of flies....31

2.6 Flies diet....31

3. Methods....32

3.1 Bioinformatic protocols....32

3.1.1 Homology modeling....32

3.1.2 Structure evaluation....33

3.1.3 Refinement....33

3.1.4 Energy minimization....34

3.1.5 Docking Analysis....34

3.1.6 Binding Energy ..... 35
3.2 Heterologous expression ..... 35
3.2.1 Growth medium preparation ..... 36
3.2.2 Chemocompetent cells *E. coli* ..... 36
3.2.3 Plasmid isolation ..... 37
3.2.4 Transformation in chemically competent *E. coli* cells ..... 37
3.2.5 Glycerol stocks ..... 38
3.2.6 Recombinant protein expression ..... 38
3.2.7 Harvest ..... 38
3.2.8 Co-expression of chaperones ..... 38
3.2.9 Cell lysis ..... 39
3.2.10 Protein purification ..... 39
3.2.11 SDS-Polyacrylamide gel electrophoresis ..... 40
3.2.12 Protein concentration determination ..... 41
3.3 Sweet taste assay with *Drosophila melanogaster* ..... 41
3.3.1 Taste preference assay for adult *Drosophila* ..... 41
3.3.2 Taste preference assay for larva *Drosophila* ..... 42
4. Results and Discussion ..... 43
4.1 Meta-analysis ..... 43
4.2 Selection of genes of known sweetener proteins ..... 48
4.3 Selection of genes of putative sweetener proteins ..... 48
4.4 Bioinformatic analysis ..... 49
4.4.1 Ligand preparation ..... 50
4.4.1.1 Homology modeling ..... 50
4.4.1.2 Evaluation and optimization of hybrid models and Crystal structures ..... 53
4.4.2 Receptor preparation ..... 60
4.4.3 Docking simulations ..... 64
4.5 Heterologous expression in *E.coli* ..... 70
4.5.1 Small scale expression test ..... 70
4.5.2 Solubility teste: Co-expression of target genes with chaperones ..... 75
4.5.3 Solubility teste: Cell lysis ..... 76
4.5.4 Large scale production and purification via metal affinity chromatography ..... 80
4.6 Sweet Taste Assay ..... 82
4.6.1 First Behavioral assay ..... 82
4.6.2 Second Behavioral assay ..... 83
5. Conclusions and future perspectives ..... 89

6. Appendix.....90
6. References.....97

# Acknowledgments

First of all, I would like to express my sincere gratitude to my supervisor, Assist. Prof. Ioannis Pavlidis for giving me this exciting and challenging project. I am thankful for his mentoring, scientific guidance, usefully advising, and support during my master thesis. Moreover, I would like to thank Prof. Tsiotis and Assist. Prof. Velonia agreed to evaluate my Master's thesis. They were kind enough to share their advice and lab equipment. Also, I would like to thank all the members of their groups for support during the accomplishment of my Master's thesis.

Next, I would express my thanks to ELKE for funding my research program (KA10712), contributing significantly to its logistical assistance.

Furthermore, my grateful gratitude are extended to Prof. M. Pavlidis from the department of Biology, and especially the Ph.D. candidate Antonia Theodoridi for her valuable guidance and assistance with the sweet taste assay. Additionally, I would like to thank Maria Belenioti, who shared advice and equipment for our fly stock.

My honest thanks to all the members of the Enzyme Technology laboratory (new and old), and especially Chara Aggeli, Nikolaos Kaloudis, Elina Konia, Vasileios Tsopanakis for their enormous support and contributions to the projects. Last but not least, I would like to thank my family, George, Stella, and Alexandra, for their support and patience all these years and my friends, Konstantina, Naouma, Kelly, Eleni, Dimitri, and John, who are always there for me.

# Abstract

The food industry is in constant search of alternative sweeteners to cover the large-scale demands for sugar. However, this extensive use of sugars in daily diet is correlated to health issues, such as obesity, type II diabetes, etc. In an effort to replace carbohydrate sweeteners, the manufactories oriented toward the use of artificial or natural nonnutritive sweeteners like aspartame and steviol glucosides. Nowadays, these sweeteners have already proved to have adverse side effects on health, such as dizziness, headaches, psychological problems, so the interest shifted to natural non-carbohydrate sweeteners (terpenoids, flavonoids, and proteins).

Most proteins are tasteless and flavorless; however, some proteins elicit a sweet taste or have taste modifying properties response on the human palate. Six proteins, thaumatin, brazzein, mabinlin, monellin, pentadin, and egg white lysozyme, were identified to elicit a sweet taste. Furthermore, there are two more proteins Miraculin, and Neoculin, which are taste-modifying proteins, converting sourness into sweetness. In general, their sweetening effect is hundreds or thousands of times higher compared to sucrose. However, almost 50 years later, only thaumatin has an industrial application of these proteins, reasoning to the difficult availability and the expensive production.

In the framework of this project, we focused on the comprehension of the chemical and structural determinants that provide the sweetening effect on sweet taste proteins. We perform a meta-analysis to identify amino acids and motifs critical for the elucidation of the sweet taste. Our interest concentrated on the two sweet proteins, MNE and Mabinlin-II, and along with the results from the structure-function analysis, we searched for putative sweetener proteins from the sequence space. Subsequently, we found five putative sequences namely according to their organism: BrasCret1, BrasCret2, BrasCret3, ArabisAlp, and Cryza.M.

The first goal of the framework was the elucidation of the interaction of the targeted proteins with the sweet taste receptor (T1R2-T1R3) through docking simulations. For the receptor was used a model from the literature, as the protein was not crystallized, for the known sweet proteins their crystal structures, and for the putative proteins' hybrid models were constructed. According to docking simulations, the proteins interacted with the receptor adopting a wedge motif in their structure. Our results establish five interactions, namely Arg177 of T1R3 and Asn152, Glu170, Asp173, and Asp 218 of T1R2, that could have a critical role in the interaction, both with the characterized sweet proteins and also with the putative ones.

In a step further, we express these seven proteins in the *E.coli* expression host. Contrasting the results from the literature, we did not succeed express these proteins in high yields (2,1mg/ L cultivation for Mabinlin-II and 0,72mg/ L for MNEI).

In the third part of this project, we implement a sweet taste assay using *Drosophila melanogaster. Drosophila* can detect basic tastes comprising of sweet, bitter, and salty. According to this assay, the third instar larvae indicate a preference for the sweet proteins Thaumatin, MNEI, and Mabinlin-II under specific conditions.



Keywords: Sweeteners, sweet proteins, MNEI, Mabinlin-II, sweet taste receptor (T1R2-T1R3) *Drosophila melanogaster*

# Περίληψη

Η βιομηχανία τροφίμων βρίσκεται σε μια συνεχή αναζήτηση εναλλακτικών γλυκαντικών ώστε να καταφέρει να καλύψει τις μεγάλες απαιτήσεις ζάχαρης. Ωστόσο, η εκτεταμένη χρήση σακχάρων στην καθημερινή διατροφή του ανθρώπου, έχει συσχετισθεί με πληθώρα θέματα υγείας, όπως η παχυσαρκία, ο διαβήτης τύπου ΙΙ κ.λπ. Σε μια προσπάθεια να αντικατασταθούν αυτές οι πηγές ζάχαρης, τα εργοστάσια προσανατολίστηκαν στη χρήση τεχνητών ή φυσικών μη θρεπτικών γλυκαντικών όπως η ασπαρτάμη και οι γλυκοζίτες στεβιόλης. Σήμερα, αυτά τα γλυκαντικά έχουν ήδη αποδειχθεί ότι έχουν αρνητικές παρενέργειες στην υγεία του ανθρώπου, όπως ζάλη, πονοκεφάλους και ψυχολογικά προβλήματα, έτσι το ενδιαφέρον στράφηκε στα φυσικά μη υδατάνθρακά γλυκαντικά (τερπενοειδή, φλαβονοειδή και πρωτεΐνες).

Οι περισσότερες πρωτεΐνες είναι άγευστες και άοσμες. Ωστόσο, ορισμένες πρωτεΐνες έχουν από μόνες τους γλυκιά γεύση ή έχουν τροποποιητικές ιδιότητες όταν βρεθούν στον ανθρώπινο ουρανίσκο. Έχει αποδειχθεί ότι έξι πρωτεΐνες, η θαυματίνη, η μπραζεΐνη, η μαμπινλίνη, η μονελίνη, η πενταντίνη και η λυσοζύμη του λευκού αυγού, έχουν γλυκιά γεύση. Επιπλέον, υπάρχουν δύο ακόμη πρωτεΐνες, μιρακουλίνη και νεοκουλίνη, οι οποίες μπορούν να μετατρέψουν την ξινή γεύση σε γλυκιά. Η γλυκαντική δράσει αυτών των πρωτεϊνών είναι εκατοντάδες ή χιλιάδες φορές μεγαλύτερη συγκριτικά με την επιτραπέζια ζάχαρη. Παρόλα αυτά τα τελευταία 50 χρόνια, λόγω της δύσκολής διαθεσιμότητας και του ακριβού κόστους παραγωγής, μόνο η θαυματίνη έχει βιομηχανική εφαρμογή.

Στο πλαίσιο της παρούσας ερευνητικής εργασία, εστιάσαμε στην κατανόηση των χημικών και δομικών χαρακτηριστικών που είναι καθοριστικά για την εκδήλωση της γλυκιάς γεύσης. Πραγματοποιήσαμε, λοιπόν, μια μετα-ανάλυση για να εντοπίσουμε αμινοξέα και μοτίβα κρίσιμα για την αποσαφήνιση της εκδήλωσης της γλυκύτητας. Το ενδιαφέρον μας επικεντρώθηκε στις δύο γλυκές πρωτεΐνες, μονελίνη (MNEI) και μαμπινλινης-2, και μαζί με τα αποτελέσματα από την ανάλυση δομής-λειτουργίας, αναζητήσαμε σε βάσεις δεδομένων αλληλουχιών, πιθανές γλυκές πρωτεΐνες. Στη συνέχεια, βρήκαμε πέντε αλληλουχίες που έχουν υποθετικά γλυκιά γεύση, BrasCret1, BrasCret2, BrasCret3, ArabisAlp και Cryza.M. Η ονοματολογία τους βασίστηκε στο όνομα του εκάστοτε οργανισμού.

Ο πρώτος στόχος μας ήταν να αποσαφηνίσουμε της αλληλεπίδρασης των πρωτεϊνών με τον υποδοχέα της γλυκιάς γεύσης (T1R2-T1R3,) μέσω πειραμάτων μοριακής προσομοίωσης. Για τον υποδοχέα χρησιμοποιήσαμε ένα μοντέλο από τη βιβλιογραφία, καθώς η πρωτεΐνη δεν έχει κρυσταλλώθηκε ακόμα, για τις γνωστές γλυκές πρωτεΐνες τις κρυσταλλικές τους δομές και για τις υποτιθέμενες πρωτεΐνες κατασκευάσαμε υβριδικά μοντέλα ομολογίας. Σύμφωνα με τα αποτελέσματα, οι πρωτεΐνες αλληλεπιδρούν με τον υποδοχέα υιοθετώντας ένα μοτίβο σφήνας στη δομή τους. Βρήκαμε πέντε αλληλεπιδράσεις, συγκεκριμένα Arg177 του T1R3 και Asn152, Glu170, Asp173 και Asp 218 του T1R2, που θα μπορούσαν να έχουν κρίσιμο ρόλο στην αλληλεπίδραση, τόσο με τις χαρακτηρισμένες γλυκές πρωτεΐνες όσο και με τις υποτιθέμενες.

Στα επόμενα βήματα, δοκιμάσαμε την έκφραση και των επτά πρωτεΐνες σε βακτηριακά κύτταρα E.coli. Σε αντίθεση με τα αποτελέσματα της βιβλιογραφία, δεν καταφέραμε να εκφράσουμε αυτές τις πρωτεΐνες σε υψηλές αποδόσεις (2,1 mg/L για το μαμπινλινη-2 και 0,72 mg/L για την MNEI).

Στο τρίτο μέρος της ερευνάς μας, φτιάξαμε ένα assay για την ανίχνευση της γλυκιάς γεύσης χρησιμοποιώντας την *Drosophila melanogaster*. Η *Drosophila* μπορεί να ανιχνεύσει βασικές γεύσεις που αποτελούνται από γλυκό, πικρό και αλμυρό. Σύμφωνα με τα αποτελέσματα μας, οι προνύμφες τρίτου σταδίου δείχνουν μια προτίμηση στις γλυκές πρωτεΐνες θαυματινη, MNEI και μαμπινλίνη-2 κάτω υπό συγκεκριμένες συνθήκες.

Λέξεις κλειδιά: Γλυκαντικά, γλυκιές πρωτεΐνες, MNEI, μαμπιλίνη-2, γλυκός υποδοχέας (T1R2-T1R3), *Drosophila melanogaster*:

# Abbreviations

| | |
|---|---|
| °C | Degrees Celsius |
| 7TMD | Seventh transmembrane domain |
| AB mix | Acrylamide/bis-acrylamide |
| ADI | Acceptable daily intake |
| Amp | Ampicillin |
| APS | Ammonium persulfate |
| ATP | Adenosine triphosphate |
| CPU | Central processing unit |
| CRD | Cysteine-rich domain |
| FAO | Food and Agriculture Organization |
| FDA | Food and Drug Administration |
| g | Gram(s) |
| GPCRs | G-protein coupled receptors |
| GRAS | Generally Recognized as Safe |
| GRs | Gustatory receptors |
| h | Hour(s) |
| IBs | Inclusion bodies |
| IPTG | Isopropyl-β-D-1-thiogalacto-pyranoside |
| JECFA | Joint Food and Agriculture Organization Expert Committee on Food Additives |
| KPi | Sodium phosphate |
| L | Liter |
| LB | Lysogenic Broth |
| M | Molarity |
| Mab-II | Mabinlin |
| Min | Minute(s) |
| mL | Milliliter |
| MNEI | Single-chain monellin linked by a dipeptide (Gly-Phe) |
| NMR | Nuclear magnetic resonance |
| OD | Optical Density |

| | |
|---|---|
| PAGE | Polyacrylamide Gel Electrophoresis |
| PDB | Protein Data Bank |
| pI | Isoelectric point |
| PREF | Gustatory preference index |
| PSI-BLAST | Position-Specific Iterative-Basic Local Alignment Search Tool |
| SCF | Scientific Committee on Food |
| SCM | Single-chain monellin |
| SDS | Sodium dodecyl sulfate) |
| SSBs | Sugar-sweetened beverages |
| T1R2 | taste receptor type 1, member 2 |
| T1R3 | taste receptor type 1, member 3 |
| VFD | Venus flytrap domain |
| WHO | World Health Organization |
| YASARA | Yet Another Scientific Artificial Reality Application |
| µL | Microliter |

# 1.Introduction

In recent years, the correlation of the extensive use of sugar to diseases, such as type II diabetes, obesity, and poor oral health, have led to a worldwide trend to reduce sugar consumption. Food manufacturers are in constant research for alternative sweeteners (food additives that can mimic the sweetness of sugar). So far, this trend is based on the use of artificial or natural non-nutritive sweeteners (Haque et al., 2020). However, artificial sweeteners, like aspartame, have already been proved to have adverse side effects on health, such as dizziness, headaches, gastrointestinal issues, mood changes, psychological problems, and bladder cancer, and thus public acceptance is decreased in the last few years (Belloir et al., 2017). One of the solutions lies in the commercialization of sweet proteins, which are low-calory sweeteners and do not associate with adverse health effects.

## 1.1Sugar and human health

Sugar is one of the fundamental supplements in food manufacturers, with the total production reaching 178,9 million tons in the years 2018-2019, and is processed to expand to 198 million metric tons in 2027 (Heilmann et al., 2013). India produces 24 % of the total sugar, followed by Brazil (16 %) and the EU (10 %) (Heilmann et al., 2013). During the past 40 years, the sugar intake has tripled. The reason for that is the vast availability and the easier accessibility in fast food, processed food, and sugar-sweetened beverages (SSBs). The sugar industry categorized sugar into ten types, namely: glucose and dextrose, fructose, household sugar (sucrose), maltose, and also sugars that are found in honey, syrups, fruit juices, and fruit juice concentrates. Excess sugar consumption leads to various long-and short-term health issues, such as for overweight and obesity, type 2 diabetes, and cardiovascular disease (WHO, 2021; Yach et al., 2016; Heilmann et al., 2013; Haque et al., 2020).

Overweight and obesity are increasingly prevalent conditions all over the world. The recent World Health Organization (WHO) report: “Obesity and overweight” illustrates this critical condition, revealing that global obesity has approximately tripled since 1975, and nearly 13 % of the world’s adult population (11 % of men and 15 % of women) were obese in 2016 (WHO, 2021b). Furthermore, the numbers about the young ages remain inauspicious. The prevalence of overweight and obese children aged 5-19 has 340 million in 2016, while in 1975 they were more than four-fold less. This tremendous increase transpired likewise among both boys and girls: in 2016, 18% of girls and 19% of boys were overweight. (Malik, et al.,2010a; Lobstein & Uauy, 2004).Furthermore, diabetes has arisen as a dominant global health problem. Diabetes is a chronic, metabolic disease characterized by the inability of the body to regulate and control blood sugar (glucose) levels Persistent and long-term raised blood sugar is associated with damage to the body's systems, especially in blood vessels, nerves, and several organs (Bloomgraden, 2004; Roglic, 2014). The majority of the patients have type 2 diabetes, which occurs when the body becomes resistant to insulin or does not make sufficient amounts. The number of deaths has reached 1.5 million per year worldwide, and the number of cases of diabetes in developing countries is likely to increase more than two-fold in the next 30 years, from 115 million in 2000 to 353 million in 2030 (Figure 1) (WHO, 2021a).

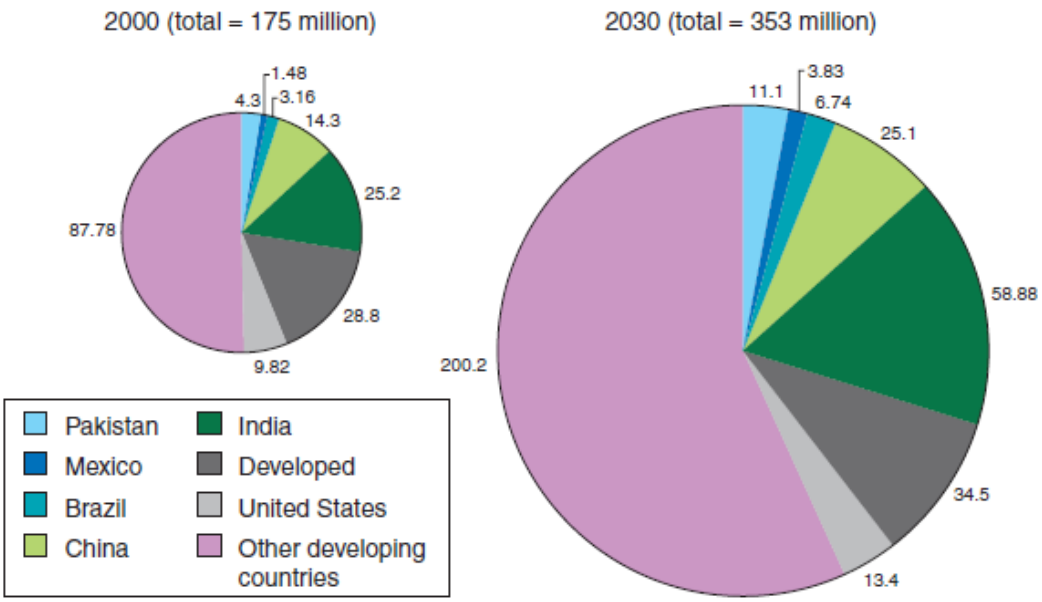



Figure 1: The figure illustrates the number of people with diabetes categorized by country in 2000 and the projection of the same number in 2030. Reprinted with permission from: Wild, S., Roglic, G., Green, A., Sicree, R. & King, H. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. *Diabetes Care* 27, 1047–1053 (2004).

The modern lifestyle that includes the large consumption of junk and processed food, and SSBs, play a critical role in the expansion of related chronic metabolic diseases, such as metabolic syndrome and type 2 diabetes. SSBs, which include soft drinks, energy drinks, iced tea, juice, have spread across the globe in the last decades. Malik and co-workers perform a meta-analysis using the MEDLINE database and reveal that individuals who consume SSBs 1-2 times per week have 26 % higher risk of developing type 2 diabetes and 20 % higher of developing metabolic syndrome (Malik, et al., 2010b).

## 1.2 Requirements for the ideal sweet-tasting compound

The demand has never been more urgent in the promotion of a new sweetener that enables more cost-effective food and beverage products with zero calories and safety demands. The research for an ideal sweetener is very challenging and, mostly, impasse because they fail to meet one or more of the specific requirements for commercial success. In general, the criteria include: good taste, similar to sucrose, clean flavor profile, safety, solubility, stability, cost-effectiveness, and patentability (Fry, 2012).

The taste profile is significant for food manufacturers. Surprisingly, it is too hard for two sweet compounds to have similar taste. There are many variations in quality, including saturation, side-tastes or after-tastes, such as bitter, metallic, and licorice, lead to the rejection of molecules candidate for commercialization. Requirements for the assurance of food component safety differ from country to country. For instance, in the US, the Food and Drug Administration (FDA) regulates the usage of sweeteners and sweet additives, and declares

them as Generally Recognized as Safe (GRAS) (DuBois, 2008). In the European Union (EU) the safety has been evaluated by the EU's Scientific Committee on Food (SCF) and by the Joint Food and Agriculture Organization (FAO)/ (WHO) Expert Committee on Food Additives (JECFA) (Mortensen, 2006). Mortensen clarifies the requirements for safety evaluation, highlighting the significance of the in vitro and in vivo toxicological testing (Mortensen, 2006). Furthermore, hydrolytic and photochemical stability are crucial factors for food and beverage compounds. Sweeteners must be stable at a range of ambient temperatures and a long duration of transporting and storage. The manufacturer needs to be able to process the sweetener at high temperature, in pasteurization and sterilization procedures, without loss of the properties. Solubility is one of the dominant aspects of an ideal sweetener. The sweetener needs to be able to dissolve in sufficient amounts in aqueous systems, to provide the desired level of sweetness to the final product. Finally, cost-effectiveness is meaningful for the industry together with the consumers. Sweeteners, artificial or natural, are always compared to sucrose, so the product must be cost-competitive with sucrose, something very challenging for biotechnological research (DuBois, et al., 2008; Mortensen, 2006).

## 1.3 Sweeteners: Natural carbohydrates, artificial sweeteners and natural sweeteners

The sweeteners are categorized according to their origin (natural or artificial), their special characteristics (sweetness and aftertaste), their nutritional value (caloric or noncaloric and digestibility), and their sweetness potency (Das & Chakraborty, 2016). The sweeteners are highly diverse, from proteins to halogenated sugars and the sweetness of these compounds ranges from 0,6 (Erythritol) to 20.000 (Advatame) times than sucrose on a weight basis. Sucrose is the usual standard in these comparison tests and thus is assigned a sweetness factor of 1.

The most ubiquitous category is natural carbohydrates, such as sugars, which are naturally occurring organic substances with varying sweetness and caloric contributions (Belloir et al., 2017). Sucrose, commonly mentioned as table sugar, is a disaccharide composed of glucose and fructose, and is one of the main ingredients of the human diet worldwide. In addition to sucrose, there are hundreds of other compounds commonly named sugars, such as maltose, lactose, trehalose, the monosaccharide fructose, and some keto-disaccharides, like maltulose, isomaltulose, lactulose, cellobiulose, and melibiulose (Tappy, 2018). Furthermore, sugar alcohols (polyols) such as xylitol, mannitol, glycerol, sorbitol, lactitol, and erythritol, are used as sweeteners. Polyols naturally exist in fruit and vegetables, have a few calories, and possibly contribute to health conditions on high consumption (Grembecka, 2018). According to Temussi, erythritol and fructose are two of the most well-known sweeteners in food manufacture (Temussi,2006). Fructose is naturally found in honey, fruits, and several root vegetables and is one of the three most essential sugars found in the blood. Also, fructose is often used by people with diabetes but, its benefits are tempered by concern that it may have an unfavorable influence on plasma lipids. Erythritol occurs naturally in some fruit and

fermented foods, is around 60% as sweet as sucrose, and is implicitly non-caloric. The WHO and FDA approved erythritol as a sugar substitute in 1999. It is generally free of side effects but inclines to cause laxative results when consumed in large amounts (Belloir et al., 2017; Joseph et al., 2019; Temussi, 2006).

The increasing demand for low-calorie or no-calorie sweeteners led to the identification of artificial sweeteners as sugar substitutes. Artificial sweeteners have been used in considerable and growing amounts in food and beverages, especially for those who are diabetic and obese, are also used in cosmetic products and medications. This group of sweeteners consists of substances with a very intense sweet taste, much higher sweetness than sucrose, and is usually used in tiny amounts to replacing the sweetness of a much higher amount of sugar. In 1879, Fahlberg working at Johns Hopkins University in the laboratory of I. Remsen serendipitously discovered a sweet substance and named it saccharin (Arnold, 1986). Saccharin is the first commercially available artificial sweetener (300 times sweeter than sugar) with FDA approval and has no calories (Belloir, et al., 2017). Saccharin is also known as E954 and is approved in a wide range of food and beverages. The acceptable levels of use range from 100 to 500 mg/kg, depending on the product (Mortensen, 2006). Since then, semisynthetic peptides (aspartame and neotame) and synthetic chemicals (cyclamate, acesulfame-K, and sucralose) have been identified and approved by the FDA and the EFSA (Das & Chakraborty, 2016; Mortensen, 2006). Aspartame is one of the most rigorously tested food supplements. It is a methyl ester of aspartic acid and phenylalanine and was discovered in 1965 by James Schlatter (Schlatter, et al.,1970). It is 200 times sweeter than sucrose, but its sweet profile varies from sucrose. Currently, a study based on the understating of the said effects of aspartame showed that the sweetener was a carcinogenic agent inducing a considerable dose-related raised incidence of several types of malignant tumors, such as hematological neoplasia, in rats and mice and both genders (Moreno, e tal.,2018). In 2014 FDA approved advantame as a food and flavor enhancer (Das & Chakraborty, 2016; Mortensen, 2006). Advantame is an aspartame analog developed by Ajinomoto Japanese food and biotechnology corporation. It is 20.000 times sweeter than sucrose, with an exceptional sugar-like taste, heat stability, and low-cost production (Otebe, et al., 2011). Notwithstanding the advantage of using artificial sweeteners as an alternative source to sugars, there has been a controversy over their safety. Numerous investigations reveal that those artificial sweeteners, especially aspartame, are related to health issues such as dizziness, headaches, gastrointestinal difficulties, and mood changes (Joseph, et al., 2019; Otebe, et al., 2011). Ferland et al, working on the effects of aspartame on glucose and insulin levels in blood plasma reinforce these opinions and reveal that aspartame led to a rise in glucose and insulin levels comparable to that of sucrose (Ferland et al., 2007). Another investigation performed by Pearlman et al. highlights that artificial sweetener affect the host microbiome, gut-brain axis, glucose homeostasis, weight gain and enhance the danger of glucose intolerance (Pearlman et al., 2017).

The research and industry spotlight has been recently shifted towards natural and healthy non-nutritive sweeteners. Plants have been explored extensively for their sweetener metabolites. In general, natural sweeteners can be classified into three groups: flavonoids,

terpenoids, and proteins (Joseph, et al., 2019). For the terpenoids, there are several subclasses of diterpenoids and triterpenoids known for their sweet sensation. Among them, stevia (*Stevia rebaudiana)* is one of the most popular plant-based sweeteners. It is abundant in Paraguay, in South America and its leaves contain two sweet compounds, stevioside, and rebaudioside A, with the first one being the amplest sweetened compound in this plant part (Angelini, et al.,2018). Stevioside was found in 1905 by the chemist Rebaudi, while rebaudioside A was isolated and structurally defined in 1976 by Tanaka and co-workers (Prakash & Chaturvedula, 2018; Word, 1976). The sweetness intensity of stevioside and rebaudioside A is estimated to be around 200 times sweeter than sucrose (Kinghorn et al., 2010; Mortensen, 2006). Nevertheless, rebaudioside A is remarkably more pleasant-tasting and more water-soluble than stevioside (Fry, 2012). In 2008 the FDA gave rebaudioside A, a GRAS status, and three years later the European Commission has also accepted it as a tabletop sweetener. In the same year, the JECFA establish an acceptable daily intake (ADI) for adults at 0–4 mg/kg body weight per day (Fry, 2012). Stevia leaf and raw extracts are not be reported as GRAS in the US and are not allowed for usage as sweeteners. Furthermore, the fruit Luohan Guo (*Siraitia grosvenorii)*, popularly known as Lo Han Kuo or Monk fruit, native to southern China and northern Thailand, contains a mixture of curcurbitane-type triterpenoid saponins identified as mogrosides, which are around 300 times sweeter than sucrose (Chiu, et al., 2020). It is approved as a food additive in Japan, China, Taiwan, Singapore, and Hong Kong, and as a dietary supplement in Australia, New Zealand, and in the US. Until 2020, *Siraitia grosvenorii* was not recorded among approved Novel Foods in the EU (Fitch & Keim, 2012). Following stevia and monk fruit, glycyrrhizin is the last commercial high-potency sweetener from terpenoids. It is a triterpene glycoside isolated from the roots of a licorice plant, *Glycyrrhiza glabra* L, in several places throughout the world (Graebin, 2018; Sharma, et al., 2018). Glycyrrhizin, also namely as glycyrrhizic acid, is almost 100 times sweeter than sucrose and has a rich licorice flavor. The ammonium salt of glycyrrhizin is approved with GRAS status in the US and is used principally as a flavor enhancer in Europe, Australia, China, India, and numerous other countries worldwide (Fry, 2012). After various toxicological analyses, the EU Scientific Committee for Food (SCF) define the ADI for glycyrrhizin to 100 mg/day while the Japanese, Netherlands, and Dutch governments to 200 mg/day. On the other hand, flavonoids have two significant sweeteners subgroups, the dihydrochalcones, and the dihydroflavonols, although, are not any commercial high-potency sweeteners available (Kinghorn et al., 2010).

## 1.4 Sweet Proteins

The proteins are the least explored category of natural sweeteners, even though the first sweetener protein, miraculin, was isolated in 1968 (Brouwer et al., 1968). Their sweetening effect is hundreds or thousands of times higher compared to sucrose. Until now, eight sweet and taste-modifying proteins have been identified, namely, thaumatin, monellin, mabinlin, lysozyme, pentadin, brazzein, curculin (neoculin), and miraculin (Faus, 2000; Kinghorn et al.; 2010, Picone & Temussi, 2012; Temussi, 2006). These proteins are isolated mostly from

tropical plants except for the lysozyme, which naturally occurs in egg whites. Plant origin proteins are divided into two categories, the sweet-tasting proteins, and the taste-modifying proteins. These proteins are extremely sweet and low-caloric.

Thaumatin is an intensely sweet-tasting protein isolated from the katemfe fruit of *Thaumatococcus daniellii* Benth (Marantaceae), a plant in West Africa (Izawa, et al.,2010; Mesuda, 2018). Thaumatin elicits a sweet taste sensation at approximately 50 nM, with a licorice aftertaste, and is nearly 100.000-fold sweeter than sucrose on a molar basis, or 3.000 on a weight basis (Izawa et al., 2010). It consists of a single chain of 207 amino acid residues with eight intramolecular disulfide bonds (Kant, 2005). The 3D structure of thaumatin has been solved by X-ray studies and reveals that the structure contains three domains: I) an 11-stranded β-sandwich folded into two Greek key motifs in which all beta-strands are antiparallel besides the N- and C-terminal. II) a large disulfide-rich region, forming an α-helix and three short helical fragments. III) a β-ribbon containing two β-strands (Figure 2) (Masuda et al., 2011A; Masuda et al., 2011B; Ogata et al., 1992). This sweet protein contains at least five different variants: two major (thaumatin I and II) and three minors (thaumatin a, b, and c), with molecular weight around 22 kDa and isoelectric point about 12 (Table 1) (Masuda & Kitabatake, 2006).

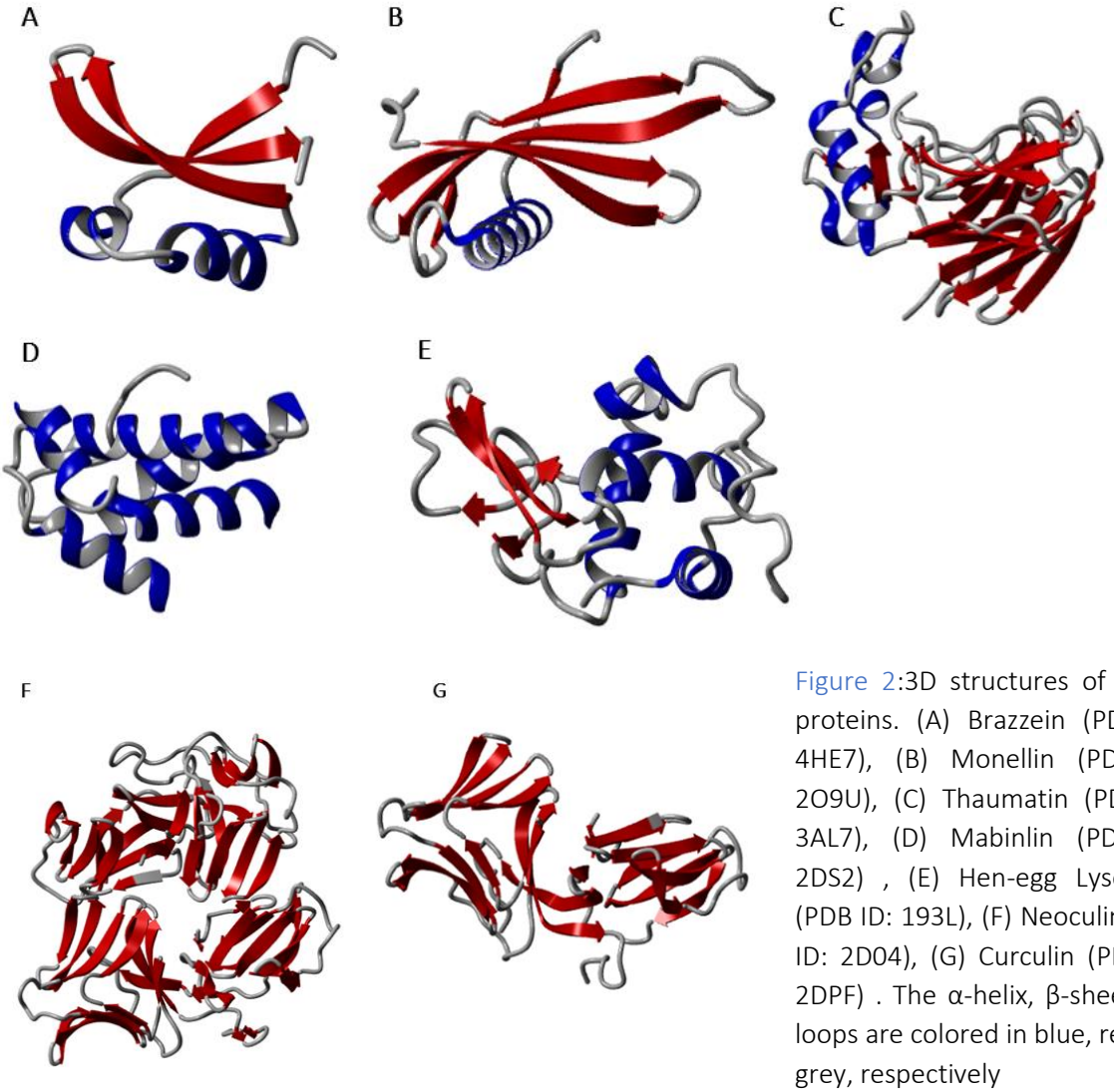



Figure 2:3D structures of sweet proteins. (A) Brazzein (PDB ID: 4HE7), (B) Monellin (PDB ID: 2O9U), (C) Thaumatin (PDB ID: 3AL7), (D) Mabinlin (PDB ID: 2DS2) , (E) Hen-egg Lysozyme (PDB ID: 193L), (F) Neoculin (PDB ID: 2D04), (G) Curculin (PDB ID: 2DPF) . The α-helix, β-sheet and loops are colored in blue, red and grey, respectively

The results from ion-exchange column chromatography and native-PAGE reveals that the plant sample consists of 60% thaumatin I and 20% thaumatin II. These two main forms differ in four amino acid residues, all located on the molecular surface (thaumatin I: Asn46, Ser63, Lys67, and Arg76, thaumatin II: Lys46, Arg63, Arg67, and Gln76) (Izawa et al., 2010; Masuda et al., 2011A). Furthermore, it is the only sweet protein available on a commercial scale with the name Talin. Talin is an approved sweetener in more than 20 countries. The FDA approved thaumatin as GRAS for about 30 applications. In the EU it is approved as both flavor and enhancer sweetener with the number E957 (Mortensen, 2006).

Monellin was isolated from the fruit, known as the serendipity berry, of the tropical plant *Dioscoreophyllum cumminsii* Diels in West Africa (Hung, et al., 1999). Monellin is likewise a basic protein with an isoelectric point nearly to 9,3 and has no disulfide bonds (Masuda & Kitabatake, 2006). In its natural form, monellin is a 10,7 kDa and consists of two non-covalent polypeptide chains, A with 44 amino acids and B with 50 (Table 1) (Aghera & Udgaonkar, 2011). The chains are not sweet separately, but in the native form, the sweetness value is estimated to be around 3.000 times sweeter than sucrose on a weight basis and 100.000 times on a molar basis, and leaves an aftertaste (Izawa et al., 2010). The sweet taste emerges after several seconds, but the perception lasts for over an hour. The 3D structure of monellin shows that the two polypeptide chains interact through intermolecular hydrogen bonds between the amino-terminal of the A chain with the carboxyl terminus of the B chain and several hydrophobic interactions (Figure 2) (Lee, et al., 1999). To enhance the thermal stability of monellin two single-chain constructs have been engineered, either directly by combining the C-terminus of the B -chain (Glu50) with the N-terminus of the A-chain (Arg1), single-chain monellin (SCM), or linked by a dipeptide (Gly-Phe), MNEI (Kant, 2005; Masuda & Kitabatake, 2006). Sweetness evaluation reveals that both SCM and MNEI have a similar taste as the native monellin, with a recognition threshold around 127 nM (Masuda & Kitabatake, 2006).

Brazzein, a 6,5 kDa protein, is the smallest sweet-tasting protein with a high pH and thermal stability. This sweet protein consisting of 54 amino acids and has four intramolecular disulfide bonds. Brazzein and Pentadin were isolated from the African plant Oubli, *Pentadiplandra brazzeana* Bailon (Nookaraju &Upadyaya, 2010; Neiers, et al., 2018), and it is 500-2.000 times sweeter than sucrose on a weight basis and 9.500 times on a molar basis with the threshold value around 500 nM (Table 1) (Masuda & Kitabatake, 2006). Brazzein has a pure sweet sensation, without a sour, salty, or bitter, but a lingering aftertaste thus, it is culturally considered to be safe to consume although it has not been formally recognized with GRAS status by the FDA (Ming & Hellekant, 1994). The structural characterization reveals two forms of brazzein; the major, which contains a pyroglutamate residue (pGlu-brazzein) at its N-terminal, and the minor form (des-pGlu1-brazzein) without this residue (Figure 2) (Poirier, et al., 2012). NMR spectroscopy reveals that the full-length brazzein consists of one short α-helix (residues 21-29) and three β-strands [residues (5-7), (34-39), (44-50)] which form a triple-stranded antiparallel β-sheet. Brazzein also maintains three loops (loop I, residues 9–19; loop II, 30–33; loop III, residues 38–45) (Tancredi et al., 1992; Gao et al., 1999; Lee et al., 2013). Another unique tertiary characteristic of brazzein is the “cysteine-stabilized alpha-

beta" motif (CSαβ), which is involved in the stabilization of an α-helix with the closest β-strand by forming two disulfide bonds (Temussi, 2006). Sweetness characterization indicates that the latter has twice the sweetness of the former. Pentadin was a molecular mass of 12 kDa, and its sweetness is estimated to be about 500 times sweeter than sucrose on a weight basis (Nookaraju & Upadyaya, 2010). So far, no further details are accessible, but it was assumed that pentadin is a dimeric of brazzein, consisting of two subunits coupled by disulfide bonds.

Mabinlin is isolated from the mature seeds of Mabinlang (*Capparis masaaki* Levl.) and grows in the subtropical region of Yunnan, China (Nirasawa, et al., 1994). These seeds elicit a sweet perception with a long aftertaste. There are four known isoforms, called mabinlin I, mabinlin II, mabinlin III, and mabinlin IV, with molecular mass about 12 kDa (Table 1) (Lui, et al., 1993). Mabinlin II is the most abundant isoform in nature, and it is a basic protein with an isoelectric point of 11,3 (Figure 2). Also, it is 400 times sweeter than sucrose on a molar basis and ten times on a weight basis, and its activity is retained after 48 h by incubation at boiling temperature (Kant, 2005; Nirasawa, et al., 1994). Mabinlin II is a heterodimer consisting of an A-chain with 33 amino acid residues and a B-chain with 72 residues, connected by two intermolecular disulfide bridges (Faus, 2000). Besides, B-chain has two intramolecular disulfide bonds (Masuda & Kitabatake, 2006). It has been elucidated that the encoded precursor of mabinlin II is a single-chain of 155 amino acids, comprising a signal peptide of 20 residues, an N-terminal extension peptide of 15 residues, a linker peptide of 14 residues, and a single-residue C-terminal extension (Sun, et al., 1996).

Table 1: Comparison on the characteristics of sweet proteins

| Sweet proteins | Geographic distribution[a] | Organism (Plant name)[a] | Threshold value (Weight basis)[b] | Amino acids[a,b] | Active form [a,b] |
|---|---|---|---|---|---|
| Thaumatin | West Africa | *Thaumatococcus danielli benth.* | 3.000x | 207 (22kDa) | Monomer |
| Monellin | West and Cental Africa | *Dioscoreophyllum Cumminsii diels* | 3.000x | 45-A chain 50-B chain (10.7kDa) | Heterodimer |
| Brazzein | Africa | *Pentadiplandra brazzeana baillon* | 2.000x | 54 (6.5kDa) | Monomer |
| Pentadin | Africa | *Pentadiplandra brazzeana baillon* | 500x | - | - |
| Mabinlin | China | *Capparis masakai levl* | 400x | 33-A chain 72-B chain (12.4kDa) | Heterodimer |
| Neoculin | Malaysia | *Curculigo latifolia* | 550x | 114 | Heterodimer |
| Miraculin | West Africa | *Richadella dulcifica* | | 191(49.2 kDa) | Homodimer |

a: Izawa, et al.,2010
b: Neiers, et al., 2018

Miraculin is a basic glycoprotein isolated from red and yellow berries (*Richadella dulcifica*), known as "Miracle Fruit", in West Africa (Moore & Buffington,1968). It is a taste-modifying protein, transforming a sour taste to a sweet at acidic pH, by an unknown molecular mechanism (Nookaraju & Upadyaya, 2010). Miraculin consists of 191 amino acids with an N-linked oligosaccharide (Theerasilp & Kurinhara, 1988). Furthermore, it occurs either as a homodimer (49,2 kDa) or as a tetramer (98,4 kDa) (Table 1). The tetramer form is a combination of four monomers grouped into dimers linked by intramolecular disulfide bridges. Miraculin is expected to be 400.000 times sweeter than sucrose on a molar basis, although its sweetness depending on pH (Ezura & Hiwasa-Tanase, 2018).

Curculin is isolated from the plant *Curculigo Latifolia* Dryand in Western Malaysia (Yamashita, et al., 1990). It is the only known protein that elicits a sweet taste 500 times sweeter than sucrose on a weight basis, and also has taste-modifying properties, changing the sour taste into sweet (Suzuki, et al., 2004). Moreover, curculin is a homodimer comprised of two identical subunits of 114 residues, linked through two disulfide bridges (Nookaraju &Upadyaya, 2010). Neoculin (Figure 2), which is isolated from the same plant, is an acidic glycoprotein with taste-modifying activity. Its sensation is estimated to be over 500 times sweeter than sucrose on a weight basis, and its sweetness depends on pH (Kurimoto et al., 2007; Shimizu-Ibuka et al., 2006). Also, as a taste modifier, Neoculin converts sour taste to sweet only for 30-60 minutes. Neoculin is a heterodimer form composed of an acidic, N-glycosylated subunit (NAS) with 113 amino acids and a basic, curculin subunit (NBS) (Masuda & Kitabatake, 2006; Nookaraju & Upadyaya, 2010) (Table 1).

Lysozyme is one of the best-characterized enzymes, catalyzing the hydrolysis of β-1,4 glycosidic bonds in bacterial cell walls, and is the only known sweet protein of animal origin (Masuda et al., 2001). Egg white lysozyme is divided into three groups, chicken type (c-type), goose type (g-type), and viral type (v-type), and are all reported to be sweet. Hen egg lysozyme (c-type) has 129 residues with a molecular weight of about 14,5 kDa (Figure 2), while goose and ostrich egg lysozyme (g-type) consist of 185 residues and a molecular weight of 20,5 kDa (Masuda et al., 2005). G-type and c-type have a high percentage of identity, although the N-terminus of g-type contains four half-cystine residues that do not appear in the c-type. However, these proteins are low in potency, the threshold value of g-type is around 6 μM, and c-type is 10 μM (Maehashi & Udaka, 1998; Masuda & Kitabatake, 2006). Egg white lysozyme has significantly reduced sweet properties in comparison with the other sweet proteins. As mentioned above, hen egg lysozyme possesses a clear sweet taste, although lysozymes from other origins have a different sweet profile. (Table 2) (Maehashi & Udaka, 1998). Nonetheless, these proteins have no commercial application as sweeteners, but are of interest, as naturally sweet molecules.

Table 2: Sweetness profile comparison of various lysozymes

| SWEETNESS PROFILE [a] | SOURCE ORIGIN |
|---|---|
| Heavy sweetness, strong, long-lasting aftertaste with weak astringency | Hen-egg<br>Turkey (94% homologous to hen)<br>Quail (97% homologous to hen) |
| Light sweetness, persistent licorice-like taste with weak astringency | Guinea fowl (92% homologous to hen)<br>Soft-shelled turtle (68% homologous to hen) |
| Tasteless | Human (60% homologous to hen)<br>Bovine milk ( a-lactalbumin) (38% homologous to hen) |

a: Maehashi & Udaka, 1998

## 1.5 Sweet taste receptor and gustatory function

Humans' taste system identifies compounds that elicit at least five tastes: sweet, umami, bitter, sour and salty. Recently, pioneering research identified lipid sensors on the tongue, suggesting that fat can exist as the sixth taste (Lee and Owyang, 2017). The first step of taste processing starts at taste receptor cells (TRCs) in taste buds on the tongue. A tastant enters the oral cavity, activates the TRCs, and then TRCs transmits the information via sensory afferent fibers to specific areas in the brain. There are four morphologic subtypes of TRCs. Sweet taste receptors are located in Type II taste cells. Also, Type II cells express G-protein coupled receptors (GPCRs) to detect umami and bitter tastes, while Type IV cells likely represent stem or progenitor taste cells. Type I, namely glial-like cells, recognize the salty taste, and Type III cells the sour (Neiers, et al., 2016).

More accurately, the sweet taste signaling pathway starts binding a specific ligand to the sweet taste receptor, activated the heterotrimeric G-protein α-gustducin (Figure 3). Afterward, the phospholipase $C\beta_2$ is stimulated, attending to the release of intracellular $Ca^{2+}$ and then the transient receptor potential cation channel M5 (TRPM5) is also activated. This cascade of activations leads to the release of ATP. Finally, ATP activates the purinergic receptors on presynaptic cells and afferent sensory fibers (Depoortere, 2014; Heidari-Beni & Kelishadi, 2018).

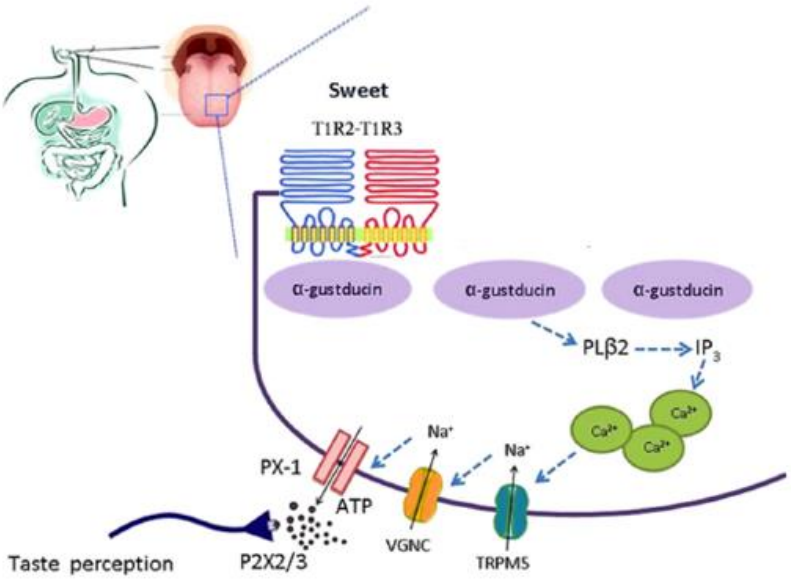



Figure 3: Simplified model of the taste GPCR signaling pathways. T1R2-T1R3 heterodimeric sweet taste receptor; ATP, adenosine triphosphate; $Ca^{2+}$ Calcium ions; $pL\beta_2$, phospholipase C β2; $IP_3$, Inositol 1,4,5-triphosphate receptor; $Na^+$, Sodium ion; PX-1, pannexin 1-hemichannel; P2X2/3, purinergic receptor-channels; TRPM5, transient receptor potential cation channel M5; VGNC, voltage-gated $Na^+$ channel. Reproduced with free license from: Lee, A. A., & Owyang, C. (2017). Sugars, sweet taste receptors, and brain responses. *Nutrients*, *9*(7), 1–13.

## 1.6 Sweet taste receptor anatomy

Existing knowledge illustrates that the mammalian sweet receptor is a heterodimer of two subunits belonging to the class C G-protein coupled receptor (GPCR) family, namely, T1R2 (taste receptor type 1, member 2) and T1R3 (taste receptor type 1, member 3) (Mace, et al.,2007). Members of class C GPCRs share distinct structural characteristics comprising a large extracellular N-terminal domain (NTD) linked to the seventh transmembrane domain (7TMD) by a well-conserved cysteine-rich domain (CRD). Additionally, the NTD domain of this class adopts a unique conformation named the Venus flytrap domain (VFD). The VFD is comprised of the two lobes and is responsible for ligand activation of the receptor. The CRD is responsible for transmitting conformational changes from the extracellular domain to 7TMD. The 7TMD transfers the signal to the intracellular signaling elements (Li, et al.,2002; Kniazeff, et al.,2011)(Figure 4).

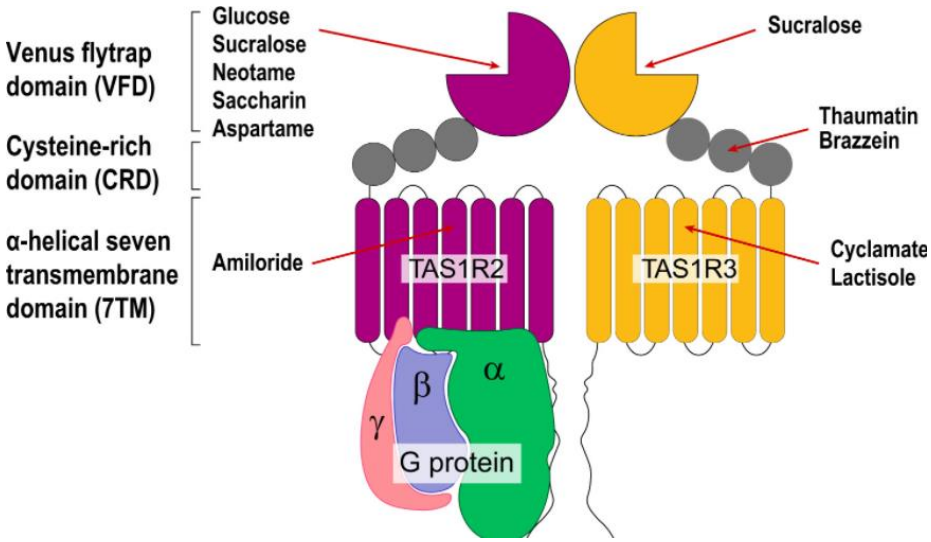



Figure 4: The picture represents the topology of the sweet taste receptor and its ligand-binding sites. The T1R2 subunit showed with magenta and the T1R3 with orange. The domains T1Rs are also presented. Ligand-binding sites exist in different locations along both subunits of the STR, with the orthosteric-binding site being found in the VFD of TAS1R2. Reprinted with permission from: DuBois, G. E. (2016). Molecular mechanism of sweetness sensation. *Physiology and Behavior*, *164*, 453–463.

## 1.7 Wedge model: a mechanism for sweet taste sensation

The sweet-taste receptor can identify different classes of sweet molecules ranging from sugar such as sucrose, amino acids to proteins. Site-directed mutagenesis studies (Cheron et al., 2017; Masuda et al., 2012) and molecular modeling (Perez-Aguilar et al., 2019) for the sweet receptor show the existence of an orthosteric binding site in the cavity of the two lobes of the VFD. The orthosteric binding pockets recognize the binding of small sweeteners such as sucrose, saccharin, glucose, and neotame. Liu and co-workers identified four essential residues in the T1R2 orthosteric binding site (Ser40, Val66, Ile67 and Asp142) important for the interaction of neotame and aspartame (Liu et al., 2011). Similarly, Zhang and his group found seven critical residues in T1R2 for sucrose and sucralose, Ser40, Tyr103, Asp142, Asp278, Glu302, Pro277, and Arg383 (Zhang et al., 2010). However, many sweeteners cannot fit in the cavity of the orthosteric binding sites, suggesting that the sweet-taste receptor has multiple binding sites. Actually, further investigations reveal at least three allosteric binding sites (Cheron et al., 2017). One of these is in the 7TMD of T1R3 and seems to interact with cyclamate (Jiang et al., 2005) and neohesperidin dihydrochalcone (Winnig et al., 2007). Although, sweet proteins appear to bind in different sites of the sweet receptor (Figure 4). Sensory analysis for thaumatin (Acevedo et al., 2018) and brazzein (Walters & Hellekant, 2006) show an interaction in the cysteine-rich domain of the human T1R3 and in the VFD of T1R2, taste-modifying protein neoculin (Koizumi et al.,2007) and monellin to VFD of T1R2 (Acevedo et al., 2018).

However, structural analysis and multiple alignments among the sweet proteins show that they do not have any sequence homology or any obvious similarity among their folds. In

addition, there is no mechanism for the interaction between the sweet proteins and the sweet receptor, and any effort to identify ‘‘sweet fingers’’ failed (Tancredi et al., 2004). The only mechanism that explains the interaction of small molecules and the larger sweet proteins is the ‘‘wedge model’’ (Figure 5) (Morine et al., 2005). In this model, it is assumed that the sweet receptor exists in an equilibrium between two conformations, one inactive, with the VFD from both subunits being in an open conformation (Roo, open–open), and the active, with the one VFD open and the other closed (Aoc, open-closed). According to the wedge model, both small molecular weight sweeteners and the sweet proteins shift the equilibrium between the two conformations in favor of the active form.

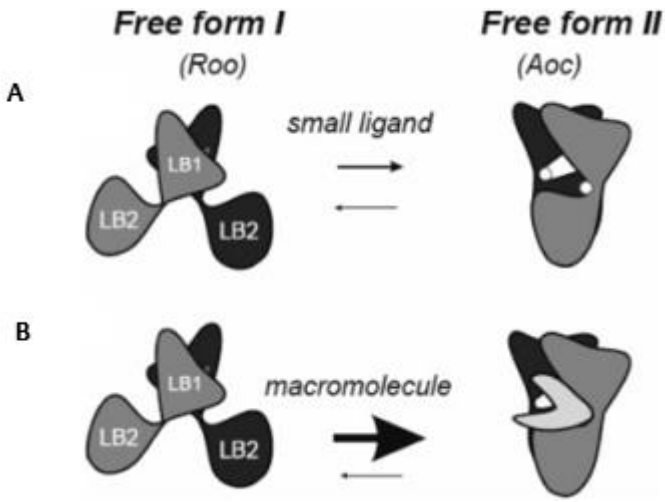



Figure 5: Cartoon representation of the Wedge model. LB1 and LB2 represent the extracellular domain of T1R2/T1R3. On the right side, the angle between LB1 and LB2 is smaller in the closed protomer (black). A) Binding a small ligand converts inactive free form I into the complexed free form II. (B) The binding of a macromolecule stabilizes the formation of free form II. Small ligands are shown as white balls, the wedge protein is colored gray. Reprinted with permission from: Morini, G., Bassoli, A., & Temussi, P. A. (2005). From small sweeteners to sweet proteins: Anatomy of the binding sites of the human T1R2_T1R3 receptor. *Journal of Medicinal Chemistry*, *48*(17), 5520–5529.

The difference is that the macromolecule binds on a large external site of the active conformation. Docking experiments, performed on monellin, thaumatin, and brazzein, showed that all three proteins bind to the active conformation of the receptor, confirming the wedge model (Morine et al., 2005; Temussi P. A. 2011).

## 1.8 Sweet taste in mammals

As mentioned above, humans can recognize a variety of sweet compounds via the heterodimeric T1R2-T1R3. However, other mammals do not respond to all sweeteners as humans. For them, the ability to distinguish compounds by taste is critical to their nutrition and to avoid toxic chemicals (Kaushik et al.,2020). Naturally, mammals respond to nutritive sweet compounds, including carbohydrate sugars. In addition, most mammals can recognize a range of high-potency sweeteners that stimulate sweetness, although the reason for

sensitivity to these non-nutritive compounds remains mostly unexplained. As an illustration, mice and rats stimulate acesulfame-K, nasaccharin, sucralose, and sorbitol but do not identify aspartame, Na cyclamate, glycyrrhizin, neohesperidin dihydrochalcone, or thaumatin (Tordoff et al., 2001; Danilova et al., 1998). Lemurs respond appetitively to stevioside but do not prefer aspartame, monellin, saccharin, sucralose, and thaumatin (Sclafani et al., 2004). In addition, there are several references for the relation between sweet proteins and mammals. Experiments about the electrophysiological behavior of thaumatin and monellin have been performed on dogs, domestic and guinea pigs, rats, rabbits, and rodents, revealing any responses to these proteins (Brouwer et al., 1973; Hellekant et al., 1976; Wel et al., 1992). Also, rodents cannot taste brazzein. Nevertheless, electrophysiological experiments on old-world and new-world primates display a distinct response in brazzein, while monellin and thaumatin did not elicit any or little response (Danilova et al., 2003; Danilova et al., 1999). Decidedly, these examples illustrate the difference between sweet perception humans and other animals. A good reason for these behaviors among mammals could be the variations in the sequence's homology of their respective sweet-taste receptors. For instance, mouse and human sweet taste receptors exhibit a mere 71 % sequence identity, while all T1Rs of old-world primates have more than 90 % sequence identity (Wintjens et al., 2011). Although investigating the gustatory systems of insects found that *Drosophila melanogaster* can detect many sweeteners, qualitatively and quantitatively similar to humans' abilities (Bantel & Tessier, 2016).

## 1.9 *Drosophila melanogaster*: a model organism to study gustation

The fruit fly *Drosophila melanogaster* is widely used as a model organism for behavioral assays, and over the last few years, larval *Drosophila* gained a more dominant position. *Drosophila* can detect basic tastes comprising of sweet, bitter, and salty. In flies, taste cells are distributed over the whole body like the labellum, legs, wings, and female genitalia (Figure 6A). The labellum is the central taste organ in *Drosophila* located at the end of the proboscis, equivalent to the human tongue (Kaushik et al., 2020). In larvae, the major chemosensory organs are located on the tip of the head, the dorsal (DO), the terminal (TO), and the ventral (VO) (Figure 6B) (König et al., 2014; Apostolopoulou et al., 2015). Like mammals, the *Drosophila* gustatory receptor (GRs) genes detect taste compounds. Although, there are not any significant sequence similarities between them. *Drosophila* taste receptors have seven transmembrane domains instead of three of GPCRs, and they are distantly associated with *Drosophila* olfactory receptors (ORs). Until now, 68 GRs genes are found in *Drosophila*. Furthermore, in adult flies, there are four significant receptors responsible for the sweet sensation the Gr5a, Gr43a, Gr64a, and Gr64f (Dahanukar et al., 2007). The Gr64a detects sucrose and maltose, and the Gr5a the trehalose and melezitose. The role of Gr64f is cooperative for the responses for all sugars tested with Gr5a and Gr64a (Dahanukar et al.,2007; Jiao et al.,2007). Gr43a is the only receptor known to detect fructose. In larvae, only Gr43a seems to have a critical role in taste recognition (König et al., 2014; Apostolopoulou, et al.,2015).

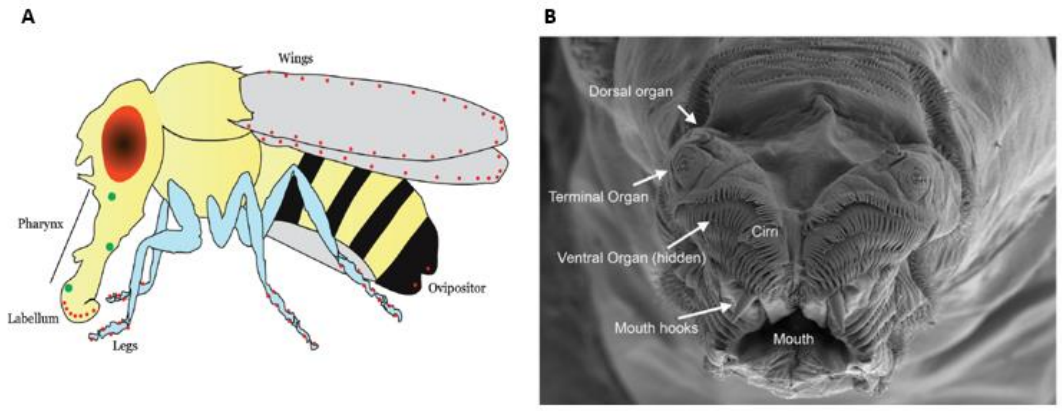



Figure 6: Gustatory system of adult *Drosophila.* (A) Labellum, pharynx, legs, wings and ovipositor as shown with red and green dots. (B) Fine structure of the larval cephalic organs: frontal view of the larval cephalon with the external sensory organs on each cephalic lobe: the dorsal organ (DO), the terminal organ (TO), the ventral organ (VO), which is hidden behind a row of cirri. All organs are located as paired structures dorsally to the mouth and mouth hooks. Reprinted with permission from: Apostolopoulou, A. A., Rist, A., & Thum, A. S. (2015). Taste processing in *Drosophila* larvae. *Frontiers in Integrative Neuroscience*, *9*(OCT), 1–9.

## 1.10 *Drosophila* life cycle

*Drosophila melanogaster* is a model organism for understanding how insects perceive smell and taste, gaining popularity from other model organisms used in the laboratory. Fruit flies offer tremendous advantages in the dissection of behavioral, cellular, and molecular pathways. Another factor that forced fruit fly in the leading position as a model organism is the short life cycle. The life cycle lasts about 12 days (Figure 7), and therefore it is easy to raise the desired population (Va & Paul, 2009; Spremulli, 2014).

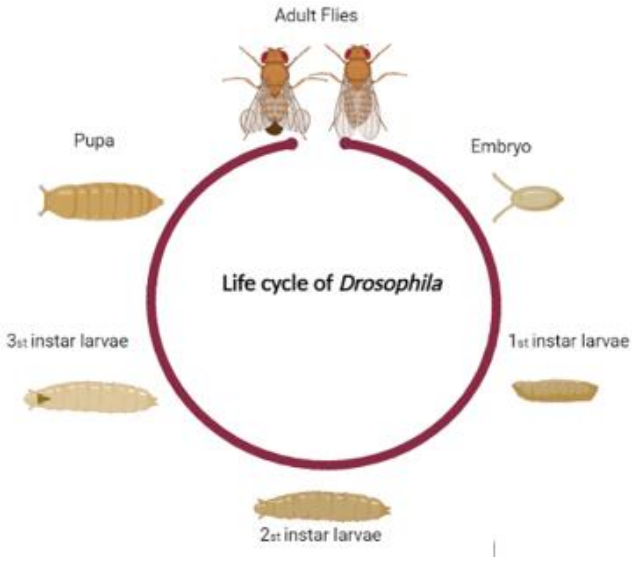



Figure 7: The life cycle of *Drosophila melanogaster*. Modified figure from: Va, D., Sa, A., & Paul, S. (2009). Wonder animal model for genetic studies Drosophila melanogaster-its life cycle and breeding methods. A review. *Sri Ramachandra Journal of Medicine, II(2),*33

Also, *Drosophila* is a holometabolous insect, and its life cycle can be divided into four stages: embryogenesis, larva, pupa, and adult life. The embryonic stage is a fast process, which follows fertilization and the formation of the zygote occurs within the egg membrane. Females lay approximately 100 embryos per day. Next, there are three larval stages (3 instars) and last around 5 days (Affleck & Walker, 2019). The larva is a white, segmented, worm-shaped burrower with black mouthparts in the narrower head region. The first instar larvae starts to feed directly on the surface of the medium. Second instar larvae stay into the medium, and the third instar larvae is mature, leaves the culture medium, and wanders up the walls of the flask, searching for a place to pupariate for 24–48 h. Following encapsulation of the 3 d instar larva, the pupal stage begins and continues around 4 days (Va & Paul, 2009; Spremulli, 2014). During pupariation, the most drastic changes arise, known as "metamorphosis". Most larval tissues are degraded, and adult organs (head, legs, thorax, and wings) develop from the imaginal disks. Finally, when the metamorphosis is completed, the adult flies emerge (Robertson, 1936). Adult flies have the typical insect anatomy, they are fragile, light in color, and their wings are not fully expanded. A few hours later the color changes to darker, and the flies take the normal appearance of the adult fly.

## 1.11 Aim of the thesis

The first goal of the thesis is to highlight the potentials of sweet proteins as low-calorie sweeteners. More than that, the project aspires to determine structural elements and motifs that are crucial for the sweet taste of the proteins, as well as for the interaction with the sweet taste receptor. This information are intended to guided us to the identification of novel sweet proteins. To achieve our purpose, a meta-analysis was performed and combining the outcomes, five putative sweet proteins were selected from the sequence space. Bioinformatic analysis was performed to analyze the interaction with the sweet taste receptor by docking analysis, and efforts on the recombinant expression in *E.coli* were performed. Finally, in this framework, a sweet taste assay was established to avoid larger scale experiments with taste panels and to analyze in high throughput several variants, employing *Drosophila melanogaster* as a model organism to determine the sweetness of the proteins of interest.

# 2. Materials

## 2.1 Scientific literature for meta-analysis

PubMed (https://pubmed.ncbi.nlm.nih.gov/), Scopus (https://www.scopus.com/home.uri) and Google Scholar (https://scholar.google.com/) databases were used as the principal data source for scientific literature on sweet proteins. Searches using combinations of the term "sweet proteins" and term "critical amino acids" were performed to determine publications related to sweetening effects. Also, combinational searches with the terms "sweet proteins" and "motif" were performed to identify structural elements responsible for sweetness. In further steps, after the recognition of a specific motif of the sweet proteins, we continue searches using the term "wedge motif". The data (authors, publications, identity numbers, and papers) from searches were saved for review.

## 2.2 Bioinformatic tools and other software

### 2.2.1 Protein Data Bank

The Protein Data Bank (PDB- https://www.rcsb.org/), Berman et al., 2000) is a no-cost database that organizes 3D structures of biological molecules such as proteins and nucleic acids of all organisms including bacteria, yeast, plants, flies, other animals, and humans. The 3D structures are determined in their vast majority by X-ray crystallography, Nuclear magnetic resonance (NMR) spectroscopy and electron microscopy. The structures are chosen based on bibliographic data.

### 2.2.2 Blastp in NCBI database

The Basic Local Alignment Search Tool (Blast) (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was also used to recognize proteins with high identity matches with the sweet proteins. The criteria used for BLASTP[1] were:

• Choose search set: database: non-redundant protein sequences (nr), Organism: plant

• Algorithm parameters: BLOSUM62 scoring matrix, Word size 6, Expect threshold 10, Max target sequences 250, Gap costs (Existence:11, and Extension: 1), window size 40, threshold 21.

[1] BLAST or *Basic Local Alignment Search Tool* is an algorithm/program which finds regions of local similarity between sequences.

### 2.2.3 Geneious

Geneious (https://www.geneious.com/) is a DNA, RNA, and protein sequence alignment, assembly, and analysis software platform integrating bioinformatics and molecular biology tools. In this study, Geneious was used for Sanger Sequencing analysis, and multiple alignments to find a correlation between sweet proteins. The version of the software used in this study was 9.1.8.

### 2.2.4 YASARA

YASARA (http://www.yasara.org/) (Yet Another Scientific Artificial Reality Application) is a molecular graphics, modeling, and a dynamics computer program that started being developed by Elmar Krieger in 1993. YASARA was written using the C, assembly, and Python programming languages. Also, YASARA can be used to perform energy minimization, structure refinement, homology modeling, and docking experiments under a license fee. In this study, it was used for homology modeling, refinement, and also finding the binding energy of docking complexes. The version of YASARA used in this study was YASARA Structure v. 20.4.24.

### 2.2.5 MolProbity

MolProbity http://molprobity.biochem.duke.edu/index.php?MolProbSID=8je92crfdp6tjejfkvkslchcg0&eventID=24) (Chen et al., 2010) is a structure-validation web service that provides a broad-spectrum solidly based evaluation of model quality at both global and local levels. MolProbity relies heavily on the power and sensitivity provided by optimized hydrogen placement and all-atom contact analysis. In this study, it was used for protein structure validation. The version of MolProbity used in this study was version 4.2.

### 2.2.6 PyMol

PyMol (https://pymol.org/2/) is a 3D molecular visualization system on an open-source foundation. The program was used to generate all protein figures. Also, was used to calculate atom distance, and evaluate the results of molecular docking. The version of PyMol used in this study was version 0.99.

### 2.2.7 ClusPro

The ClusPro server(https://cluspro.org/home.php) is an openly used tool for protein-protein docking. It is free software that provides a simple home page for basic use. For this project, ClusPro (v.2.0) was employed for docking experiments.

## 2.3 Genes of interest

At the present Master thesis, two already known sweet proteins (MNEI and Mab-II) and five putative sweet proteins (ArabisAlp, BrasCret1-3, and Cryza.M) are tested. The gene is presented in Table 3, and their sequences are in Appendix 1. The gene synthesis and subcloning were performed from GenScript. The synthesized genes were subcloned in pET-21a(+), using Nde1 and XhoI as restriction enzymes. Once the constructs were delivered, *E. coli* BL21(DE3) were transformed (see Paragraph 3.2.4) and plasmid isolation in mini scale was performed. The isolated plasmid was sent for sequencing (Genewiz, Germany) to prove the correct constructs.

Table 3: Proteins of interest of this thesis.

| Name | Coded sequence | Original organism | Vector | Selection | Inducer |
|---|---|---|---|---|---|
| MNEI | Sweet protein | *Dioscoreophyllum cumminsii diels* | pET-21a(+) | Amp | IPTG |
| Cryza.M | Hypothetical protein F2Q69_00052222 | *Cryza meyeriana var. granulata* | pET-21a(+) | Amp | IPTG |
| Mabinlin-II | Sweet protein | *Capparis masakai levl* | pET-21a(+) | Amp | IPTG |
| BrasCret1 | Hypothetical protein E2562_004804 | *Brassica cretica* | pET-21a(+) | Amp | IPTG |
| BrasCret2 | Hypothetical protein F2Q68_00034154 | *Brassica cretica* | pET-21a(+) | Amp | IPTG |
| BrasCret3 | Hypothetical protein F2Q68_00033573 | *Brassica cretica* | pET-21a(+) | Amp | IPTG |
| ArabisAlp | Hypothetical protein AALP_AA1G345700 | *Arabis alpina* | pET-21a(+) | Amp | IPTG |

## 2.4 Proteins

In the present thesis, was used the following commercially available protein:

- Thaumatin, from *Thaumatococcus daniellii* Benth organism, in purity (Tokyo Chemical Industry, TCI)

- Lysozyme from chicken egg white, in activity 15.000U/mg (Serva)
- Bovine Serum Albumin (BSA), Albumin bovine Fraction V powder (Sigma)

## 2.5 Rearing conditions of flies

The strain of *Origari G.* used in this study was originally obtained from the Group of Prof. Ladoukakis, Laboratory of Entomology, Department of Biology, University of Crete. Flies were housed and bred in the quarantine facility at the laboratory of Prof. Nikolaos Chaniotakis, Department of Chemistry. The population was reared under a 12 h light : 12 h dark photoperiod at 21 ± 2°C temperature and 65 ± 5 % relative humidity. Insects were kept in sterile vials which contain a cornmeal diet (Belenioti & Chaniotakis, 2020).

## 2.6 Flies diet

Flies were kept in sterile vials that contain a cornmeal diet that consisted of agar (5 g), sugar (11.6 g), fresh yeast (28.33 g), and cornmeal (33.33 g) in nanopure water (1 L), and heated to 60°C for 5 h under stirring. Nipagin (0.83 g) dissolved in ethanol (8.33 mL) was subsequently added to sterilize the diet before its distribution into the vials.

## 2.7 Molecular chaperons

Plasmids pG-Tf2 and pKJE7 (Figure 8) were used in this work. The plasmids belong to Chaperone Plasmid Set (TAKARA Bio). The plasmids carry an origin of replication (ori) derived from pACYC, and chloramphenicol resistance gene (Cmr). The Cmr gene permits the plasmids to be used with *E. coli* expression systems that operate ColE1-type plasmids containing the ampicillin resistance gene. The chaperone genes are downstream of either the araB or Pzt-1 (tet) promoter. Thus, the expression of target proteins and chaperones can be induced individually if the target gene is placed under the control of another promoter (e.g., lac). These plasmids also contain either araC or tet-R for each promoter.

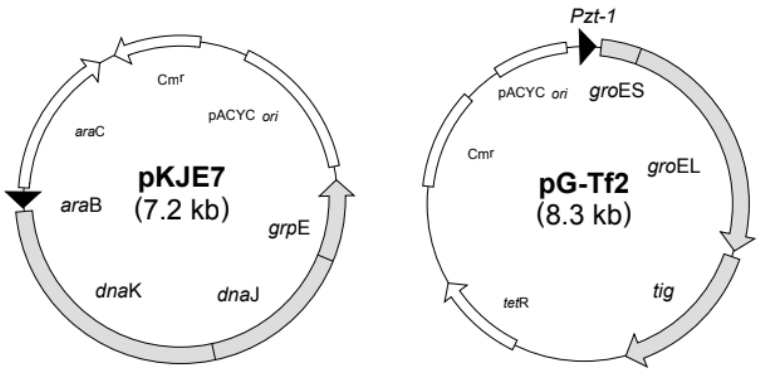



Figure 8: Representation of plasmid maps for pG-Tf2 and pKJE7. Takara Bio Inc. (n.d.). Chaperone Plasmid Set For. *Product Manual*. Retrieved from http://www.takara-bio.com

# 3. Methods

## 3.1Bioinformatic protocols

The bioinformatic analysis started with the construction of the homology models for genes of interest. Also, the crystal structures for the already known sweet proteins were downloaded for PDB (Paragraph 3.1.1). Then, the models were analyzed by the MolProbity platform (Paragraph 3.1.2). Afterwards, the models were refined by the YASARA (Paragraph 3.1.3), and then energy minimization was performed by the same software (Paragraph 3.1.4). The refined structure was evaluated again in MolProbity. Afterward, in silico experiment took place in the software ClusPro, where the proteins of interest were put in interaction with the sweet taste receptor, a process called docking (Paragraph 3.1.5). The favorable complexes obtained from the dockings are depicted in PyMOL in better resolution , and lastly, the energy of binding of the final complexes was calculated by YASARA (Paragraph 3.1.6).

### 3.1.1 Homology modeling

Homology modeling is considered to be the most accurate computational structure prediction method. It is used to determine the 3D structure of a protein from its amino acid sequence. Five putative sweet proteins are studied in the current thesis. So far, there are no crystal structures for these proteins.

YASARA structure software was used to create a homology model for the putative sweet proteins. The procedure commences with defining the working directory[2] (Options>Working directory). Then, the computer CPU threads and the graphics card have to be set accordingly (Options>Processors>Set CPU threads / Options>Processors>Set GPU). Subsequently, the homology modeling experiment is ready to be performed (Options>Choose experiment>Homology modeling). The program then requires the amino acid sequence of the target in FASTA format[3]. Also, the program provides the option to use any templates to build the homology model. Consequently, suitable parameters for the run were defined and pressing ok to star homology modeling. For this experiment the parameters were:

- Number of PSI-BLAST iterations in template search (PsiBLASTs): 3
- Maximum number of templates to be used (Templates Total): 5
- Maximum number of templates with same sequence (Templates SameSeq): 1
- Maximum number of alignment variations per template: (Alignments): 5
- Maximum allowed (PSI-)BLAST E-value to consider template (E-Value Max): 0.1
- Maximum oligomerization state (OligoState): 1

[2] A working directory is a folder in which all the resulting files are saved.
[3] FASTA format is a text-based format for representing nucleotide or amino acid sequence, always starting with a “>” symbol that indicates the single-line description of the sequence.

- Maximum number of conformations tried per loop (LoopSamples): 50
- Maximum number of residues added to the termini (TermExtension): 10
- Modeling speed (slow = best): Slow

After the experiment completed running, several files were saved in our working directory, including a *.html* file, containing the report of the experiment. The *.html* file includes information about: (1) the homology modeling target, (2) the homology modeling parameters, (3) the homology modeling templates, (4) the secondary structure prediction, (5) the target sequence profile, (6) the initial homology models, (7) the model ranking, and (8) the hybrid model.

### 3.1.2 Structure evaluation

MolProbity is a valuable collection of structural analysis tools for the validation of protein structures and homology models. In the current project, MolProbity was used for structure evaluation after the creation of the homology modeling and after the refinement. The software accepts only *.pdb* format. After uploading the structure, "add hydrogens" was selected and afterward "analyze sterics and geometry". Adding hydrogens is not required, but improves the analysis reliability. Most homology modeling programs do not include hydrogens because hydrogens are not observed in X-ray crystallography experiments. Regenerate H, applying only selected flips, was be selected. Generally, flips are applied for Asn, Gln, and His residues. There is a variety of parameters that can be set. Finally, a "summary statistics" table appears with information regarding the "quality" of the structure, including the existence of bad angles, bad bonds, Ramachandran outliers, etc. Ramachandran plot pinpoints the amino acids with poor or favored phi/psi angles. Phi/Psi angles are the dihedral angles in the protein backbone. Only certain angles are typically found in proteins They were categorized as an outlier, allowed (outer limit), and favored (completely allowed) regions inside the Ramachandran plot. Moreover, in the summary statistics table, a clash score is presented. Clash score is the number of unfavorable all-atom steric overlaps ≥ 0.4 Å per 1000 atoms, and the closer its value is to zero, the better. Also, the clash score is defined as a color code red/yellow/green for absolute quality, highlighting bad, moderate, and good values, respectively.

### 3.1.3 Refinement

To further advance the quality of homology models, especially in the media of interest, as well as the quality of the two crystal structures, performed a protein refinement by YASARA software. After launching YASARA, the first steps were to define the working directory and

the CPU threads. These steps are displayed in detail in Paragraph 3.1.1. Then the force field[4] was set to YAMBER3[5] (Simulation>Force field). Next, the conditions of the refinement (temperature, density, and pH) are determined by a ".mcr" (macro) file titled "md_refine.mcr" (Local datadase>YASARA>mcr>md_refine). Copy the file md_refine in the working directory folder. To "play" the macro first had to set the protein as the "target" (Options>Macro&Movie>Set target). Remove the solvent and water molecules (Edit>Delete>Waters). Afterward, "play" the macro (Options>Macro&Movie>Play macro), and the refinement is started. At the end of the refinement command, 20 snapshots (pdb files) were created in the working directory file, along with a "*.tab*"[6] file that displayed all the information about the snapshots (formation energy, Dihedrals, Packing1D, Packing3D, and Average). The snapshot with the minimum formation energy is the one with the maximum quality. This structure was further evaluated using the MolProbity web service. If any of the 20 snapshots had the minimum energy a second refinement should be performed to bring the quality of the structure to satisfactory levels.

### 3.1.4 Energy minimization

Energy minimization is performed to find the structure with the most stable conformation. To achieve this conformation; the structures must go through many processes of shifting arrangements in space until the net of interatomic force on each atom is acceptably near zero. To run an energy minimization stimulation, firstly, the force field was defined (Simulation>Force field>YAMBER3). Then, an "Energy minimization" experiment was performed (Options>Choose experiment>Energy minimization). We repeated the energy minimization experiment several times until the energy of the structure remained unchanging, meaning that it had reached its most stable conformation, as calculated by the program. Finally, the file (File>Save as) of the 3D structure was saved as a *.yob* file.

### 3.1.5 Docking Analysis

Docking is a computational method that can be used to predict the behavior of a ligand in the binding site of a receptor. In this thesis, docking is performed by ClusPro online software. In order to run a docking experiment, the job title was initially defined, and then 'cpu' was selected, the docking was performed at the Massachusetts Green High-Performance Computing Center. At the next step, the structures for the receptor as well as ligand, were uploaded. It is possible to define additional advanced options depending on the

[4] A force field in molecular dynamics refers to the combination of mathematical equations and parameters that are used to explain the potential energy of the protein as a function of its atomic coordinates.
[5] YAMBER 3 stands for Yet Another Model Building and Energy Refinement force field three and is pre-defined from YASARA.
[6] TAB files are essentially tab-delimited text or data files. It can be opened using a text editor (Notepad, Notepad++, Sublime text, etc.)

requirements. In the current work, only the selection of "attraction and repulsion" was applied. This option allows the user to select critical amino acids for the interaction. The analysis took a few hours and the results were sent at the defined e-mail account. ClusPro main page shows small pictures representing the top 10 clusters. For each cluster, the table of the results showed the size (the number of docked structures), the PIPER[7] energy of the cluster center (the structure that has the highest number of neighbor structures in the cluster), and the energy of the lowest-energy structure in the cluster (Kozakov, et al., 2017). The energy-like scoring was defined by the cluster of inter-and intra- molecular contributions. Although, the energy calculated by the PIPER algorithm does not immediately correlate to binding affinity. It is not recommended to evaluate the results based on these scores.

### 3.1.6 Binding Energy

The binding energy of the protein complexes was calculated by YASARA Structure software. All structures were energy minimized before calculating the binding energy. In this project, this step was accomplished in ClusPro docking simulations. Then the solvent (water) was removed, to give space to side-chains to adopt new conformations. The optimization was performed by selecting "Edit>Optimize>Molecule". To calculate the energy of the ligand, it must be split by the receptor (Edit>Split>Molecule>Ligand). The force field was set to AMBER14 (Simulation> Force Field> AMBER14 [8] . To run the calculation, "Analyze> Energy>Binding Energy>Molecule" was selected. According to YASARA manual: *the more positive the binding energy, the more favorable the interaction in the context of the chosen force field.*

## 3.2 Heterologous expression

According to this chapter, we study the heterologous expression of the seven synthetic genes (Table 3). The genes, already subcloned into expression plasmids, were transformed in *E. coli* BL21 (DE3) and SHuffle T7 chemoreceptor cells (Paragraph 3.2.4). Then, small-scale expression tests were performed to verify the suitable conditions for the expression (Paragraph 3.2.6). After the solubility tests (Paragraph 3.2.8, Paragraph 3.2.9) and the validation of the proper conditions, we expressed the genes on a large scale (Paragraph 3.2.6). Subsequently, the cells were harvested by centrifugation (Paragraph 3.2.7) and were prepared for cell disruption, using ultrasound to release the cytoplasmic proteins (Paragraph 3.2.9). Finally, the proteins were purified by Ni-NTA affinity chromatography (Paragraph

[7] Piper is the FFT-based rigid docking program developed in Vajda Lab and ABC Group Boston University and Stony Brook University.(https://www.vajdalab.org/ )

[8] The AMBER force fields have been developed for general use in all kinds of applications. The force field is meant for simulations in explicit solvent. When used *in vacuo* for quick modeling tasks, electrostatic interactions (like formation of salt-bridges) will be more pronounced than in reality. The AMBER force fields are numbered by the year of publication, AMBER14 is the most recent.

3.2.10), and the results were analyzed by SDS-PAGE (Paragraph 3.2.11). Purified proteins were used for the sweet taste assay (Paragraph 3.3).

The media and the equipment autoclaved at 120°C for 20-30 min. The expression protocols were performed in a laminar flow hood after sterilizing with an ultraviolet lamp and ethanol. The experiments took place in the presence of flame.

### 3.2.1 Growth medium preparation

The Luria-Bertani LB medium is a common growth medium for bacterial cultivation. It contains 0.5% yeast extract, 1.0% peptone, and 1.0% NaCl (all in w/w), dissolved in 1 L $dH_2O$. The solution was autoclaved for 20-25 min at 120°C. For the preparation of LB-agar plates, 1.5% (w/w) agar was added before the sterilization.

Another growth medium was used in the current thesis was PYA8 (phosphate/yeast extract/acetate pH 8.0 medium). PYA8 contains 16.1 g/L $Na_2HPO_4 \cdot 7H_2O$ , 1.36 g/L $KH_2PO_4$, 0.5 g/L NaCl, and 5.0 g/L yeast extract. The sodium acetate was added from a sterile 10× stock solution at 0.4% concentration. pH adjustment was performed with 0.1 M NaOH or 5% $H_2SO_4$.The solution was autoclaved for 20-25 min at 120°C.

### 3.2.2 Chemocompetent cells *E. coli*

The production of chemically competent *E. coli* cells was carried out according to the Rubidium chloride method (Hanahan, 1983). For the process, RF1 (Table 4) and RF2 (Table 5) solutions are required.

The production of chemically competent *E. coli* BL21 (DH3) and SHuffle T7 was divided into two days. On the first day, 4 mL of LB medium were inoculated with 50 µL of the respective strain and incubated at 30°C, 120 rpm overnight (16 h). The next day, 100 mL of LB medium were inoculated with 1 mL from the preculture and incubated at 30°C, 120 rpm, until the $OD_{600}$ is 0.3-0.5. The culture was resting at ice for 15 min and then centrifuged for 20 min at 3112 x *g*, at 4°C. The supernatant was removed, and the pellet was resuspended in 20 mL RF1-solution.

Table 4 Recipe of RF1 solution.

| Chemicals | Concentration (mM) |
|---|---|
| Rubidium Chloride ( RbCl) | 100 |
| Manganese(II) chloride ($MnCl_2$) | 50 |
| Potassium acetate ($CH_3COOK$) | 30 |
| Calcium chloride ($CaCl_2$) | 10 |
| Glycerol | 15% (v/v) |

The cells remained in the ice bath for 15 min and are then were centrifuged again. The supernatant was discarded and the pellet was resuspended in 4 mL of RF2.The cell solution is incubated for 15 min on ice and then was separated into portions of 50 µL. The cells were instantly freezed in liquid nitrogen and were finally stored at 80 ºC.

Table 5. Recipe of RF2 solution.

| Chemicals | Concentration (mM) |
|---|---|
| Rubidium Chloride ( RbCl) | 10 |
| MOPS (3-morpholinopropane- 1-sulfonic acid) | 10 |
| Calcium chloride ($CaCl_2$) | 75 |
| Glycerol | 15% (v/v) |

### 3.2.3 Plasmid isolation

The plasmid isolation was performed with the GeneJET Plasmid Miniprep. An overnight culture (Paragraph 2.3.4) of 5 mL volume was prepared with the *E. coli* strain, containing the plasmid of interest. The biomass was harvested and resuspended in 250 µL of resuspension buffer. The solution was transferred in a 1.5 mL tube, gently shook and 350 µL of the equilibration buffer was added. After a 5 min centrifugation at 13000 rpm, the supernatant was transferred by pipetting in the GeneJET column and washed with 1 mL of the washing buffer. The spin GeneJET column was transferred in a 1.5 mL tube, and 50 µL of the elution buffer was added, incubated at room temperature for 2 min, and centrifuged for 2 min, at 13000 rpm. The column was discarded afterward, and the plasmid concentration was measured at 260 nm using 2 µL of the plasmid solution on the µDrop plate, using the photometer Multiskan Sky (Thermo Scientific). The extracted plasmid was stored at 20°C.

### 3.2.4 Transformation in chemically competent *E. coli* cells

For the transformation, 2-3 µL of plasmid DNA (50–100 ng/µL) were added to an aliquot and the cells were incubated on ice for 30 min. The heat shock followed, at 42°C for 30 s, and then the cells were incubated on ice for 2 min. Afterward, 1 mL of LB medium was added, and the eppendorf with the cells were incubated at 37°C for 1 h. For this incubation, a horizontal position of the eppendorf is preferred. Finally, from the resulting cell culture, 100 µL were spread on an LB agar plate with the corresponding antibiotic. The agar plates were incubated at 37°C for 16-24 h.

### 3.2.5 Glycerol stocks

After the transformations, overnight cultures (Paragraph 3.2.4) were prepared in LB medium from a single colony and glycerol stocks were prepared the next day by mixing 1 mL of the cell culture and 1 mL of glycerol 60% [glycerol in water (v/v)] under sterile conditions. The glycerol stocks were stored at-80°C.

### 3.2.6 Recombinant protein expression

The *E. coli* growth and induction protocol require two days. On the first day, after bacterial transformation, a single colony was selected or 10 µL from the glycerol stock (Paragraph 2.3.5) were cultured in a 5 mL overnight culture in LB using the corresponded antibiotic at final concentrations of 100 µg/mL ampicillin (amp) at 30°C. After 16-18 h of cultivation, 50 mL cultures were prepared in LB medium at flasks with a ratio liquid: air, 1:4, with the corresponding antibiotic and 0,5mL form the preculture, at 120 rpm, 37°C, until $OD_{600}$ = 0.6-0.8 was reached. Formerly, the cultures were cooled down to 20°C, and to validate the induction a volume equal to $10/OD_{600}$ was collected for analysis by SDS-PAGE. Then, 0.2 mM IPTG (final concentration) was added as an inducer. The cultures were incubated at 20°C, 120 rpm, for 16-18 h. For large scale expressions, the amount of the pre-culture is 10 mL, and the corresponding volume of the culture is 1 L.

### 3.2.7 Harvest

After the overnight culture of paragraph 3.2.6, the cultures were harvested and a volume equal to $10/OD_{600}$ was collected to verify the induction for analysis by SDS-PAGE. The rest culture is centrifuged at 9000 x *g*, for 15 min, at 4°C. The supernatant is rejected, and the pellet is resuspended in sodium phosphate buffer (50 mM, pH 7.4), and the centrifugation was repeated under the same conditions. The cell pellet is collected and stored at 20°C.

### 3.2.8 Co-expression of chaperones

Co-expression of chaperones was incorporated to facilitate proper folding and to improve the stability of the heterologous proteins. The protocol is divided into three steps: co-transformation, co-expression, and harvesting. Chemo-competent cells (*E. coli* BL21 DE3) were transformed with 1 µL of plasmid DNA and 1 µL of chaperone plasmid. Carefully flick the tube 4-5 times to mix cells and DNA, and the mixture stays on ice for 30 min. Then heat shock was performed at exactly 42°C for exactly 30 s, and the mixture was placed on the ice for 2 min. Afterward, 1 mL of LB medium was added, and the eppendorf with the cells were incubated at 37°C for 1 h. Finally, from the resulting cell culture, 0.1 mL were spread on an

LB agar plate with the corresponding antibiotics. The agar plates were incubated at 37°C for 16-24 h. To perform co-expression, the transformants were inoculated into a 5 mL LB medium containing 20 μg/mL chloramphenicol and 50 μg/mL amp. DNA plasmids provide resistance in ampicillin, and chaperone plasmids provide resistance in chloramphenicol. The preculture was incubated at 30°C and 220 rpm. After 16-18 h, a 250 mL flask was inoculated with 50 mL LB, 20 μg/mL chloramphenicol, and 50 μg/mL ampicillin with 500 μL from the preculture and 0.5 mg/mL of L-arabinose and/or 5 ng/mL tetracycline for induction of chaperone expression. Incubate at 37°C, 120 rpm, until the $OD_{600}$ of the culture reaches a value of 0.4-0.6, then add IPTG to a final concentration of 0.2 mM, and culture at 37°C for $1-4$ h. To validate the expression of the protein collected volume equal to 10/ $OD_{600}$ before and after induction for analysis by SDS-PAGE.

### 3.2.9 Cell lysis

The pellet from the culture (Paragraph 2.3.6) was resuspended in sodium phosphate buffer-Lysis Buffers (Table 6). Subsequently, cell lysis was accomplished by sonication on ice, 40 % power, 5 min of lysis (30 s lysis followed by 30 s rest on ice, Sonotrode Bandelein sonoplus). The sample was centrifuged for 20 min at 6000 x *g*, 4°C. The soluble fraction was separated from cell debris. A 50 μL sample was collected from the supernatant for analysis by SDS-PAGE (Paragraph 3.2.11). The rest of the supernatant is collected for purification using a Ni-NTA column. In addition, we performed solubility tests to increase the solubility of the expressed proteins. Therefore, were used five Lysis buffers as described in Table 6.

Table 6 : In the current thesis was used the following cell lysis buffer

| Lysis Buffer 1 | Lysis Buffer 2 | Lysis Buffer 3 | Lysis Buffer 4 | Lysis Buffer 5 | Lysis Buffer 6 |
|---|---|---|---|---|---|
| KPi 50mM pH=8 | KPi 50mM pH=8 | KPi 50mM pH=8 | KPi 50mM pH=8 | KPi 50mM pH=8 | KPi 50mM pH=8 |
| 100mM NaCl | 200mM NaCl | 100mM NaCl | 100mM NaCl | 100mM NaCl | |
| 20% Glycerol | 20% Glycerol | 20% Glycerol | 20% Glycerol | 20% Glycerol | |
| | | 0.1% Triton-X-100 | 1mM EDTA | 0.1% Triton-X-100 | |
| | | | | 1mM EDTA | |

### 3.2.10 Protein purification

The purification was performed at Synthetic Biomaterials Lab, Assistant Prof. Kelly Velonia. For this procedure three KPi buffers were prepared: Buffer A, containing 50 mM KPi and 500 mM NaCl at pH 7.5, Buffer B with 50 mM KPi, 500 mM NaCl, and 300 mM imidazole at pH 7.5 and Buffer C containing 50 mM KPi at pH 7.5. Buffers B and C were used to prepare an additional buffer of 15 mM imidazole, to be used as a lysis buffer and for the first washing

steps. The Ni-NTA column (HisTrapTM FF, GE Healthcare) had a volume of 5 mL, while for the desalination was used three columns of 5 mL in line (HisTrapTM desalting columns GE Healthcare). The column equilibration was performed with 2 column volumes, the wash was performed in 5 column volumes (10% buffer B), and the elution with 100% buffer B in 4 column volumes. The flow of the column was determined at 1.5 mL/min. After the protein purification, desalting was performed with 100% of C at a flow rate of 3 mL/min for 3 column volumes. The columns were cleaned with deionized water and then a solution of 20% ethanol and stored in the refrigerator (+4°C). Samples containing the desalted protein were collected and concentrated in Amicon® Ultra Centrifugal Filters MWCO 3 or 10 kDa at a final volume, 2mL, using sodium phosphate (5 mM, pH 8). Finally, the samples were stored at 20°C.

### 3.2.11 SDS-Polyacrylamide gel electrophoresis

Sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) is a method that is used for the separation of proteins by their molecular weight. It is based on the quality of charged molecules to move in response to the electric field applied. The high expression of proteins is detected by the presence of bands at the expected height of the gel, according to known molecular weight marker standards. The marker used was the Serva Unstained Protein Marker 6- 200 kDa. The gels of the electrophoresis consist of 2 parts, the resolving gel, and the stacking gel. The gels were prepared according to Table 7.

Table 7. Components of one 15% gel for SDS-PAGE.

| Components | Resolving gel (15 %) | Stacking gel (4 %) |
|---|---|---|
| Tris/HCl buffer, pH=7.4 | 2 mL | - |
| Tris/HCl buffer, pH=6.8 | - | 1 mL |
| AB mix 30% | 4 mL | 0.53 mL |
| dH2O | 2 mL | 2.67 mL |
| APS (10%) | 40 µL | 40 µL |
| TEMED | 4µL | 4µL |

All samples and the protein marker were prepared using 10 µL of proteins, mixed with 20 µL loading buffer, incubated at 95°C for 5-10 min (Eppendorf thermomixer comfort). The gels were loaded with 10-15 µL for each sample. The intensity in the electrophoresis device was set at 25 mA per gel and allowed to run for about 2 h. Afterward, the gels were stained with SDS-PAGE staining (methanol (45% v/v), acetic acid (7% v/v) και Coomassie Brilliant Blue R-250 (0.25-1 % w/v) for at least 2 h, and then were left in SDS-PAGE de-staining solution (methanol 40% v/v and acetic acid 7% v/v), until the gel was discolored.

### 3.2.12 Protein concentration determination

Bovine serum albumin (BSA) solutions with concentrations 0.1 mg/mL to 1mg/mL were used for the standard curve. The Bradford staining solution was prepared using 0.1 g Brilliant Blue G250 dissolved in 50 mL ethanol, 100 mL 85 % phosphoric acid, and 100 mL $dH_2O$. Subsequently, the staining was dissolved with $dH_2O$ in a volume ratio of 4 mL Bradford solution: 12 mL $H_2O$ and filtered. The samples of BSA, and the unknown concentration sample, were prepared by mixing 200 µL of Bradford staining and 15 µL sample in triplicates. The unknown samples were diluted based on the standard curve, and the absorption was measured at 595 nm using the photometer (Thermo Scientific, Multiskan Sky).

## 3.3 Sweet taste assay with *Drosophila melanogaster*

In the framework of this study, *Drosophila melanogaster* is used as a model organism for the sweet taste assay. Accurately, two different taste activity assays take place, using the adult fly (Paragraph 3.3.1) and the third instar larva (Paragraph 3.3.2).

### 3.3.1 Taste preference assay for adult *Drosophila*

This protocol was divided into two days. The first day was starvation, and the adult flies (~25) were added to the starvation vials. The vials had a cotton ball with nanopure water at the bottom of the standard fly vial. Best results are obtained with animals that are less than 5 days old. In this experiment, the animals are 1-2 days old. The vials were closed with a cotton ball and placed on an environmentally controlled incubator (12 h : 12 h L : D photoperiod at 21 ± 2°C temperature and 65 ± 5% relative humidity. The next day, the taste preference assay was performed. All tastants should be prepared on the same day. Subsequently, the assay chambers were a standard 100 mm x 15 mm plastic petri dish (or 55mm x 15mm), were prepared in the following method: Place three 10 µL dashes of tastants on three sides of the plates, according to (Bantel & Tessier, 2016) (Figure 9).

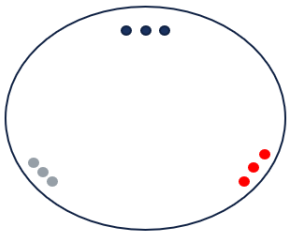


Figure 9: An illustration of the assay chambers. With red and dark blue are depicted the tastant (thaumatin and sucrose) and with gray the water.

Ensure that the spacing between drops is similar. Repeat this procedure for each tastant and plate. The observation of the assay performed at the Laboratory of Prof. Michael Pavlidis of the Department of Biology, at the DanioVision observation chamber (Figure 10). To anesthetize adult flies, their vial was left on ice for less than 5 min, avoiding behavioral effects from even extended exposure (Bantel & Tessier, 2016). To start the observation, the assay chamber with the adult flies was placed in the DanioVision, and the recording last 8 min.

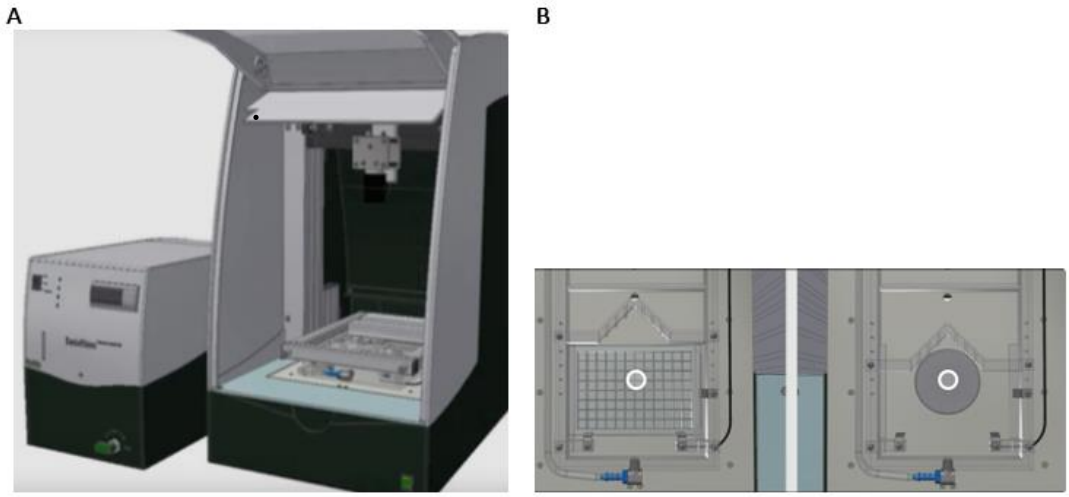



Figure 10 : Representation of the observation chamber DanioVision. A) DanioVision offer a controlled environment for high-throughput testing. B) The interior of the chamber. Data sources: https://www.noldus.com/daniovision/observation-chamber

### 3.3.2 Taste preference assay for larva *Drosophila*

The first step of this protocol was the preparation of the population. The adult flies laid the eggs for 24 h and then they were transferred into a new vial. The vial incubated in the controlled incubator (12 h : 12 h L : D photoperiod at 21 ± 2 °C temperature and 65 ± 5% relative humidity. After 5-6 days, the third instar larvae were visible. Larvae were almost ready for the experiments. In order to take the larva from the vial, 50 mL of 10-20% sucrose were added into it, and was let to sit for 20 min (Nichols et al., 2012). Larvae floated to the top. Next, it was necessary to wash the population three times with deionized $H_2O$ before the trials. Finally, the animals were collected in a clean and empty vial. Subsequently, the assay chamber, a standard 35 mm x 15 mm plastic petri dish, was prepared in the following manner: 20 mL of a 1 % agarose solution was poured into the petri dish (Nichols, et al., 2012) Once the agarose was solid, the assay chambers were ready. All tastants were prepared on the same day. The assay was performed at the Laboratory of Prof. Michael Pavlidis of the Department of Biology, at the DanioVision observation chamber. In order to start the observation, 100 µL of each tastant were added to the assay chamber with the testing larvae. Then, the petri dish was placed in the DanioVision. The recording lasted 8 min.

# 4. Results and Discussion

The results of this thesis are be divided into four major parts: Firstly, are presented the outcomes from the meta-analysis, identifying amino acids and motifs critical for the elucidation of the sweet taste (Paragraph 4.1). Afterward, the results of the *in-silico* analysis, from the construction of the homology models to the docking simulations, are be described (Paragraph 4.4). Following, the heterologous expression, purification, and characterization of those proteins with the SDS-PAGE (Paragraph 4.5). In the last part of this chapter, we discuss the implementation of a sweet taste assay using *Drosophila melanogaster* (Paragraph 4.6).

## 4.1 Meta-analysis

The meta-analysis started searching for amino acids critical for the sweet sensation of thaumatin. Thaumatin is the most analyzed sweet protein, with around 6.400 references recorded in google scholar. In the early studies, chemical modifications of lysine residues revealed that the positive charges on the surface of thaumatin play an essential role in the elicitation of its sweet taste and suggest the importance of lysine residues on the cleft-containing region. Acetylation of lysine residues declared a significant reduction in sweetness which depends on the number of lysine residues (Kaneko, 2001). The same results are exported by the modification of lysines, using a succinic acid anhydride and pyridoxal 5′-phosphate (PLP). Notably, the effect of phosphopyridoxalation of K106 suggested that the structure of this residue is crucial as well since PLP-K106 and dephosphorylated-PLP (PLP) receive the same threshold value (260 nm) (Kaneko, 2001). Site-directed mutagenesis studies focused on the two basic residues lysine, and arginine, in the cleft-containing region. Substitution of four lysines (Lys19, Lys49, Lys67, Lys106, Lys163) and three arginines (Arg76, Arg79, Arg82) were evaluated for their sweetness threshold values by sensory analysis in humans, and the results showed a substantial reduction in sweetness, particularly Lys67 and Arg82 (Masuda, et al., 2018,) (Figure 11). Exploring the reason for this remarkable reduction, further mutations in these two residues were performed in cell-based assays using HEK293 cells expressing human sweet receptors, and the results confirmed the significance of these two residues. Primarily, the EC50 (half maximal effective concentration) value of the mutant R82E could not be calculated in the cell-based assay, highlighting the critical role of the positive charge of Arg82 (Ohta, et al.,2011).

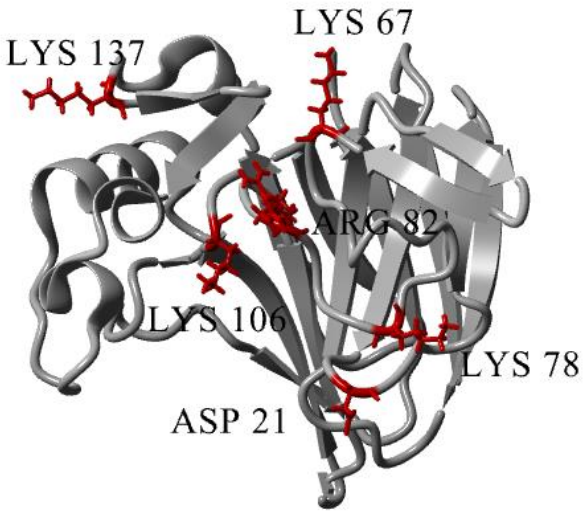



Figure 11: Overall structure of Thaumatin (PDB ID: 3AL7). The structure are colored gray and the critical amino acids red.

Structural analysis by X-ray crystallography at atomic resolution in the three mutants K78A, K106A, and K137A disclosed the significance of a large positively charged surface on the interaction. Also, the calculation of the threshold values of the mutants K78A and K137A showed an acquit increase from 45 nM to 220 nM and 230 nM, respectively (Masuda, et al., 2014). These results implied that the flexible conformations at the side-chains of K78 and K137 might enhance the interaction with the receptor (Figure 11). Next, mutagenesis studies were accomplished in six acidic amino acids (Asp21, Glu42, Asp55, Asp59, Asp60, and Glu89) all around the two critical basic residues, Lys67 and Arg82 (Masuda, et al., 2016). However, only the mutant D21N showed an extreme reduction in the threshold value at 31 nM lower than wild-type thaumatin, gaining the title of the sweetest thaumatin mutant (Figure 11) (Masuda, et al., 2019, Masuda, et al., 2016).

Furthermore, X-ray crystallography of mabinlin-II revealed unique structural elements for the interaction with the sweet-taste receptor. The results demonstrated that the ‘‘amphipathic design” of the protein could play a crucial role in the initial interaction with the receptor. Also, 3D-docking analysis between sweet receptor hT1R2/T1R3 and native mabinlin II, individual B-chain, and A-chain declared that only the native mabinlin II and individual B-chain respond to the receptor (Li, et al., 2008) .The analysis advocates that the uncommon [NL/I] tetralet motif in the B54-B64 possesses an essential structural element for the interaction with the sweet receptor. Moreover, the sweetness profile between mabinlin II and B-chain differs (Figure 12). The sweetness of the B-chain vanished quickly instead of the sweetness of native protein that lasts for half an hour, suggesting that the A-chain is critical for the long after-taste. Wintjens and coworkers performed structural analysis for mabinlins and exposed promising candidate residues for sweetness elicitation, two in A-chain, R7, and R24, and five in the B-chain, namely R8, P20, P42, R46, and N57 (Wintjens, et al., 2011) (Figure 12).

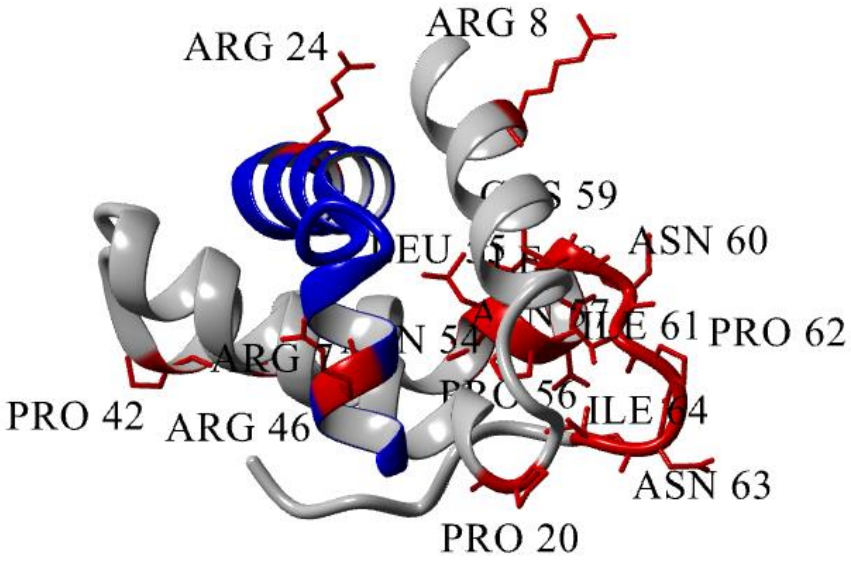



Figure 12: Overall structure of Mabinlin (PDB ID: 2DS2). The a-chain is showed by blue color and the b-chain by gray. The critical amino acids are represented by red.

The meta-analysis was continuous with the Brazzein. Extensive site-directed mutagenesis studies, combined with sensory analysis, have been conducted to elucidate the structure-function relationships of brazzein. It has been classified, based on mutagenesis data and solution structure, that brazzein has three essential sites responsible for the sweetness: site 1 (Loop III), site 2 (N-,C-terminal regions, residue Glu36, and Loop33), and site 3 (Loop residues 9-19) (Figure 13) (Assadi-Porter, et al.,2010, Jin, et al., 2003). Site 1 consists of five amino acids Tyr39, Asp40, Glu41, Lys42, and Arg34 located between the β2 and β3 strands (Walters, et al., 2009). Also, the same analysis suggested that Glu41 and Arg43 have a specific charge and size, essential for the interaction with the sweet receptor. As mentioned above, site 2 comprises the N- and C-terminal, the central part of the molecule (pGlu1, Asp2, Lys5, Lys6, Tyr8, His31, Lys42, Asp50, and Tyr54) and the three critical residues Lys30, Arg33, and Glu36 (Assadi-Porter, et al.,2005). The side-chains of Lys6 and Glu36 are exposed to the solvent, suggesting that these could directly interact with the receptor. Additionally, substitutions of Glu36 to Ala, Lys, and Gln and Lys6 to Asp concluded to abolish sweetness, highlighting the significance of these amino acids. The mutants K30A and K30R had undoubtedly reduced sweetness, and the mutation to Lys30 to the negatively charged residue Asp provides a tasteless mutant. Comparable results indicated mutations at Arg33. A good explanation for the sweetness behavior of these mutants lies in the analysis of the crystal structure. So, according to the 3D structure, the side-chain of Lys30 formed a cation-π interaction with the side-chain of Typ8 and the side-chain of Arg33 with the Tyr54 and a hydrogen bond with the backbone carbonyl group of the Cys52. Consequently, the charge and the orientation of the side chain were essential for sweet-taste activity. Another dominant residue of Loop33 is His31 (Lee, et al., 2013, Yoon, et al., 2011). Mutation of His31 to a more positive Lys exhibited a substantial rise in sweetness, while changing His into a neutral Ala decreased the threshold value, suggesting the necessity of the positive charge at this position. Side-directed mutations of the negatively charged Asp29 to either a neutral or positively charged residue (D29A, D29N, and D29K) notably enhanced sweetness (Jin, et al., 2003). Also, at the N-terminal, the side-chain of Lys5 had an electrostatic interaction with the side-chains of Glu36 and Asp50, and in comparison, with mutations studies, it was clear that the charge of this residue, rather than its size, was important (Assadi-Porter, et al.,2003). The significance of the C-terminal was revealed by the

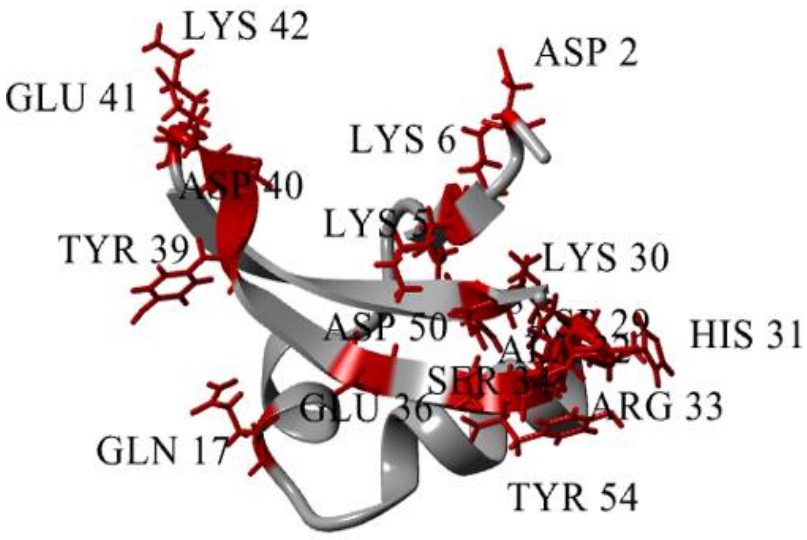



Figure 13: Overall structure of Brazzein (PDB ID: 4HE7). The structure are colored gray and the critical amino acids red.

deletion of Tyr54, which abolished the sweetness of brazzein. Further site-directed mutations of Tyr54 to Trp or His had a slight increase of sweetness, revealing that a bulky residue at this position could be important for interaction with the receptor. However, the mutation of Asp50 to Ans improved the sweet taste and the crucial role of the positive charge was repeatedly highlighted. At site 3, Lys 15 formed electrostatic interaction with the side chains of Gln21 and Asp24, while the side-chain of Gln17 was fully exposed to the solvent and located nearby the disulfide bond C16-C37. Mutation of Q17N had a meaningful reduction in sweetness, and its side-chain could be involved in re receptor identification. Additionally, mutations that determine the size and the position of the disulfide bond provide non-sweet molecules. Furthermore, the tasteless mutants, C16A/C37A, C16A/Q17C, and C16A/C37A/L18_A19insRI, illustrated the value of the Cys16–Cys37 disulfide bond in the correct folding of brazzein (Yoon, et al., 2011). Lee and collaborators described those three double mutations (H31R/E36D, H31R/E41A, E36D/E41A) and one triple mutation (H31R/E36D/E41A) produced molecules more delicious than brazzein and principally the triple one (Lee, et al., 2013, Walters, et al., 2009). Also, the results strongly confirmed the hypothesis that brazzein binds to the receptor at various sites. Ultimately, they recommended that mutations, reducing the overall negative charge and increasing the positive favor sweet-tasting protein potency.

The interaction mechanism between this sweet protein and its analogous with the sweet-taste receptor was still unknown, although numerous mutagenesis studies have been performed on monellin, SCM, and MNEI (Somoza, et al., 1995, Templeton, et al., 2011,Yang, et al., 2019, Zhao, et al., 2018). Earlier structure-based site-directed mutagenesis in native monellin and SCM revealed the importance of three regions essential for the sweet taste. The first one contained four residues (6-9), at the N-terminal, the C-terminal, and the region formed by residues in the fourth and fifth β-strands (Arg86, Arg70, and Arg72). Especially the β-carboxyl group of the AspB7 cooperated with the sweet receptor through electrostatic interaction, while IleB6 should be involved in a hydrophobic interplay (Mizukoshi,etal., 1997).Further mutagenesis studies focused to report the sweet-activity of MNEI. The solution structure of MNEI revealed three distinctive regions: the Loop34 (residues 66-69), the N-terminal (residues 6-9), and C-terminal (94-96) (Figure 14) (Somoza, et al., 1995 ,Templeton, et al., 2001). These regions were in agreement with the results from the docking analysis of MNEI with the sweet receptor, although the models of MNEI docked on the external cavity of the receptor

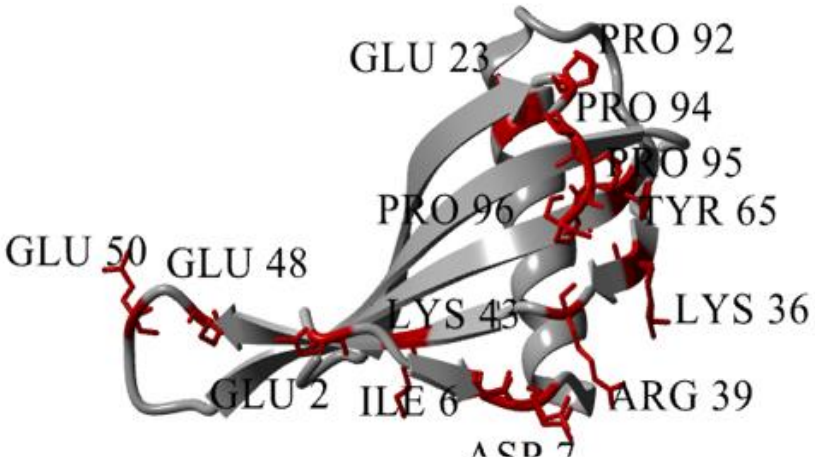



Figure 14: Overall structure of MNEI (PDB ID: 2O9U). The structure are presented with gray and the critical amino acids red.

indicated that there were several more residues responsible for the interaction. Furthermore, various mutants led to an extreme decrease in sweetness, for example, I6, D7, G9, R39, and 43. In contrast, two mutants, Y65R and E2N, seemed to enhance the sweetness by 1.7-fold improvement and 3-fold, respectively (Cai, et al., 2016, Rega, et al., 2017). Recently, loop23 residue 47-56) tasted as novel sweetness determinate site of the sweet-tasting protein. Interestingly, in both 50 and 54 positions, mutation of the negative charge into positive charge exhibited a significant increase in sweetness. E50N variant gained a 2.8-fold increment while E50K and E50M have a 3-fold (Yang, et al., 2019). The sweetness potency of E48N was partially similar to the wild-type MNEI (1.5-fold), confirming that the flexibility of the C-terminal poly-(L-proline) II was essential for the sweet-taste activity (Templeton, et al., 2011. Moreover, a positive impact was found for the sweetness of 2 double-site mutants (E50N/ Y65R and E2N/E50N) that indicated enriched sweetness than their single-site mutants.(Yang, et al., 2019). Recently, was determined a double-site mutant (E2N/E23A) of the MNEI with improved sweetness (around 3-fold) and thermostability (Zhao, et al., 2018).

To elucidate the sweet-taste properties of lysozyme, Masuda and coworkers accomplished extensive studies about the role of lysine and arginine. Alanine substitution showed that Lys13, Lys96, Arg14, Arg21, and Arg73 had a key role in the sweetness of the molecule (Figure 15) (Maehashi & Udaka, 1998, Masuda, et al.,2005 A, Masuda, et al., 2005B). Besides, the results from structure-activityand mutagenesis researches reported the importance of a positively broad surface region for the sweetness but did not the size of the particular positions. Manato and coworkers obtained a quadruple mutant of bovine stomach lysozyme (bs-LZ)(bs-E50R/D75N/E83A/E125Q) elicit 75% of the sweetness human lysozyme, suggesting that the reduction of the negative charge was decisive for the sweetness (Matano, et al., 2015). Furthermore, this study illustrated that the sweet-taste activity was independent of the enzymatic activity (Masuda, et al., 2001). The results from the meta-analysis for the sweet-taste proteins are represented in the Table 8 below.

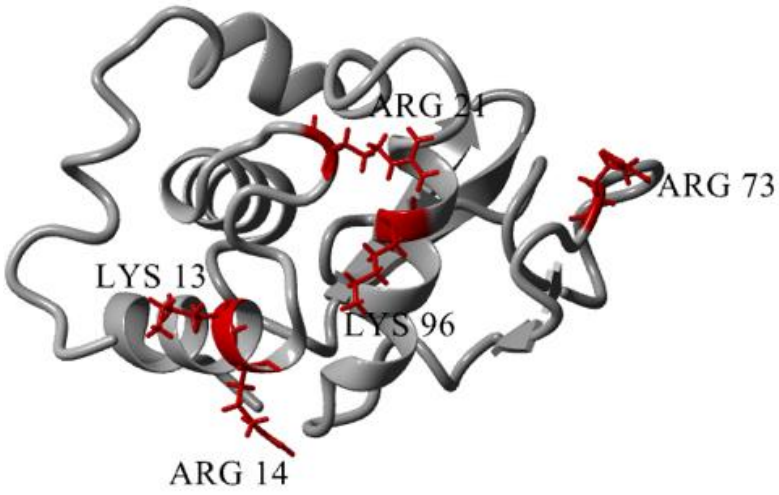



Figure 15: Overall structure of Lysozyme (PDB ID: 19EL). The structure are presented with gray and the critical amino acids red.

Table 8: Critical amino acids for Brazzein, Thaumatin, Monellin-MNEI, and Lysozyme (Hen-egg white).

| Brazzein | Thaumatin | MNEI | Mab-II | Lysozyme |
|---|---|---|---|---|
| Asp2, Lys5, Lys6, Tyr8, Gln17, Aso29, Lys30, His31,Arg33, Arg34, Glu36,Tyr39, Asp40, Glu41, Lys42, Asp50, Asp54 | Asp21,Lys67, Lys78, Arg82, Lys106, Lys137 | Glu2, Ile6, Asp7, Glu23, Lys36,Arg39 Lys 43, Glu48 Glu50,Tys65, | A-chain : Arg 7, Arg 24 B-chain :Arg8, Pro20, Pro42, Arg 46, Asn54,Leu55, Pro56, Asn57, Ile58, Cys59, Asn60, Ile61, Pro62, Asn63, Ile64 | Lys13, Arg14, Arg21, Arg73, Lys96 |

Meanwhile, the meta-analysis continued with the identification of critical amino acids for the receptor. According to the literature, the N-terminal VFD of T1R2 was required for response to monellin, the VFD of T1R2 and the T1R3-CRD for brazzein, and thaumatin the T1R3-CRD. However, molecular modeling combined with docking studies had proposed several amino acids that could interact with sweet-taste proteins. Temmusi's topological approach highlighted thirteen residues that probably interact with the MNEI and Brazzein. Eight were located in T1R2 (Asp173, Lys174, Asp188, Gln221, Gln441, Asp456, Lys457, and Ser458) and five in the T1R3 (Glu47. Glu48, Ser59, Asp 215, and Arg247) (Temussi, P. A.,2011). (Table 9) Another investigation indicated six more residues close enough to allow interactions namely Asp169, Glu170, Arg172, Arg176, Arg 217, and Asp218, all from T1R2 (Esposito et al., 2006). (Table 9). Later, Zhao et al. confirmed most of these interactions between the mutant E2N/E23A- MNEI and the receptor. In parallel, Mesuda and co-workers found some critical residue that could contact. In their complexes, thaumatin and the three mutants were located adjacent to three residues of the T1R3 subunit of sweet receptor, Glu45_T1R3, Glu47_T1R3, Asp215_T1R3, and to two of the T1R2 subunit Asp173_T1R2, Asp433_T1R2 (Mesuda T et al., 2018). (Table 9). Site-directed mutation and detailed molecular docking in the CRD of T1R3 were showed five amino acid residues vital for the response to thaumatin (Masuda T et al., 2013). Recently, another molecular docking underlined several critical binding sites of the VFD domains. Individually, brazzein could interact with Gln109, Glu145, Asn152, Ser155 of T1R2 and Arg177, Glu178, Glu428, Ser446 of T1R3, whereas thaumatin with Tyr131, Asn152, Ser155, Arg176, Trp418, Glu422 of T1R2 and Ser158, Arg177, Asp419, Val421, Lys422 of T1R3, and MNEI with As152, Asp173, Leu156, Gln419, Ile450 pf T1R2, and Arg177 of T1R3 (Acevedo, et al., 2018). (Table 9).

Table 9: Critical amino acids for the sweet taste receptor.

| T1R2 | T1R3 |
|---|---|
| Gln109, Tyr131, Glu145, Asn152, Ser155,Leu156 ,Asp169, Glu170, Arg172, Asn173, Lys174, Arg176, Asn213, Arg217, Asn218, Gln221, Trp418, Gln419, Glu422, Asp433, Gln441, Ile450, Asn456, Arg457, Ser458 | Glu45, Glu47, Glu48, Ser59, Ser158, Glu178, Arg177, Asp215, Asp419, Val421, Lys422 , Glu428, Arg247, Ser 446 |

Subsequently, our research advanced in the investigation of wedge motifs in protein surfaces. The results showed that Receptor-like Protein-Tyrosine Phospatases (RPTPs) preserves a conserved wedge motif in the D1 catalytic domain at the N-terminal. This wedge motif consists of helix-loop-helix (HLH) with some variations in the sequence. From these results, using the software Geneious (Version R9), we identified a wedge-shaped sequence (Figure 16) that was further examined by bioinformatic analysis for the sweet proteins. Unfortunately, we did not correlate this motif with the sweet proteins.

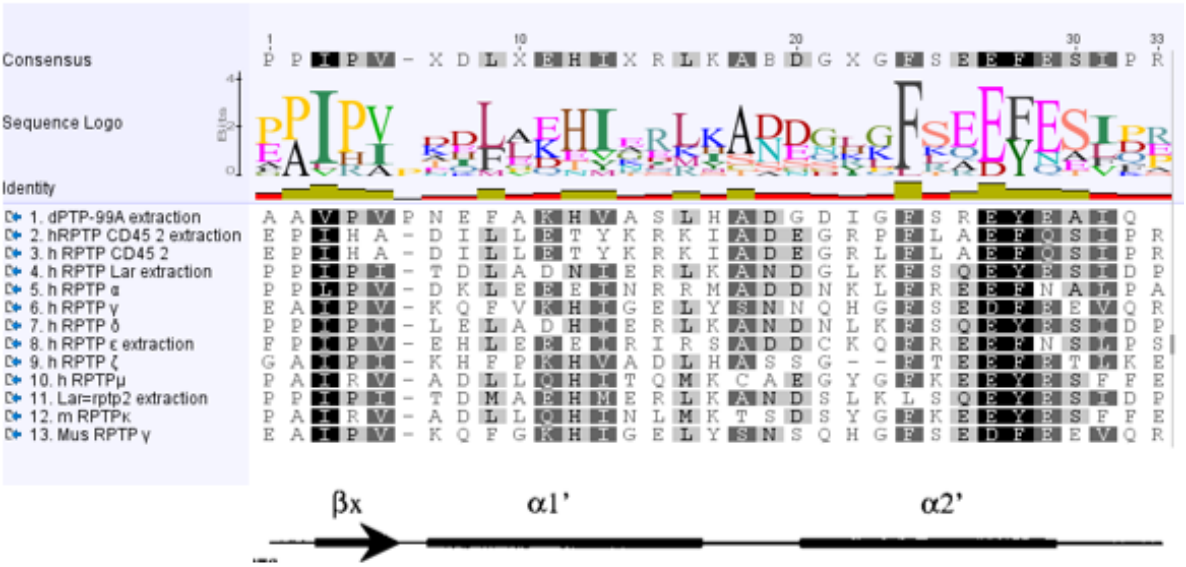



Figure 16: Identification of wedge-shaped sequence from RPTPs. The table was received for the Geneious Software.

The Basic Local Alignment Search Tool (BLSTP) was used to recognize sequences with high identity matches with the wedge-shaped motif from RPTPs. The algorithm yielded 57 alignments with E-value from 6 to 344, suggesting alignments between the wedge-shaped motif from RPTPs and the sweet proteins, the results indicated no correlation and did not further continue this analysis minimal homology. The final results were presented in Figure 17 Although we performed multiple alignments with the wedge-shaped motif from RPTPs and the sweet proteins, the results showed no correlation, so we did not continue this analysis further.

| | Name | Description | Sequence Length |
|---|---|---|---|
| | PNR41267.1:328-337 | hypothetical protein PHYPA_018670 [Physcomitrium patens] | 10 |
| | XP_024395113.1:389-398 | uncharacterized protein LOC112291641 isoform X1 [Physcomitrium patens] | 10 |
| | XP_024395114.1:389-398 | uncharacterized protein LOC112291641 isoform X2 [Physcomitrium patens] | 10 |
| | XP_024395115.1:389-398 | uncharacterized protein LOC112291641 isoform X3 [Physcomitrium patens] | 10 |
| | XP_024395116.1:376-385 | uncharacterized protein LOC112291641 isoform X4 [Physcomitrium patens] | 10 |
| | XP_024395117.1:328-337 | uncharacterized protein LOC112291641 isoform X5 [Physcomitrium patens] | 10 |
| | XP_031261336.1:39-52 | uncharacterized protein At1g76070-like [Pistacia vera] | 14 |
| | XP_031265525.1:39-52 | uncharacterized protein At1g76070-like [Pistacia vera] | 14 |

Figure 17 Summary of sequence alignments for wedge-motif. The table was received for the Geneious Software

## 4.2 Selection of genes of known sweetener proteins.

Afterword in the current thesis, the research was focused on two known sweetener proteins, MNEI and Mabinlin as they were the proteins that could be expressed in *Escherichia coli* at high yields, according to literature.

## 4.3 Selection of genes of putative sweetener proteins.

The results from the meta-analysis for the known sweetener proteins attain all the information on the significant structural elements and positions on the sweetener proteins. This knowledge enables us to proceed to further steps, of identifying putative sweetener proteins from the sequence space. The recent years, through the advances in next-generation sequencing technologies, metagenomic and genomic projects produced a vast amount of sequencing data. However, the bottleneck lies now on the characterization of the products (i.e., proteins) of the sequenced genes. In contemplation of choosing the genes of putative sweetener proteins, we performed BLASTP. This algorithm compares target sequences to sequences from databases, calculates the statistical significance of matches, and supports the recognition of the gene families. As already mentioned in the present master thesis, MNEI and Mab-II had been selected for further investigation. The results for MNEI yielded matches with percentage identity from 48,8% to 27,2% for MNEI (23 sequences). Subsequently, sequences with identity lower than 30%, and theoretical pI lower than 8,0 were rejected, and from the 6 remaining sequences was selected the gene with the accession number: KAF0910838.1 (Cryza.M). (Table 10)

Table 10: Results from BLASTP for MNEI.

| ACCESSION | Organism | Full-Length | Per. Identical | pI |
|---|---|---|---|---|
| 1. XP_023731197.1<br>Custeine proteinase inhibitor 1- like | Lactuca sativa | 115 | 27.96% | 9.31 |
| 2. XP_024983929.1<br>Custeine proteinase inhibitor 5- like | Cynara cardunculus<br>Var. scolymus | 119 | 30.00% | 10.02 |
| 3. KAF0910838.1<br>Hypothetical protein E2562_004804 | Cryza meyeriana<br>var.granulata | 118 | 35.42% | 9.16 |
| 4. XP_027125045.1<br>Potative custeine proteinase inhibitor 7 | Coffea arabica | 145 | 34.00% | 9.69 |
| 5. ABK22339.1<br>Unknown | Pichea stchensis | 231 | 33.82% | 9.61 |
| 6. CAB4082797.1<br>Unnamed protein product | Lactuca Saligna | 104 | 27.96% | 8.28 |

For Mab-II has found 87 sequences initially, and after the pI filter, only 20 were left. Then the sequences were of a similar size (≥105 amino acids), and we had 14 lefts. The remaining genes

were examined, performing multiple alignments, to clarify which sequences have the appropriate [NL/I] motif (Figure 18), as this seemed to be responsible for the sweetness.

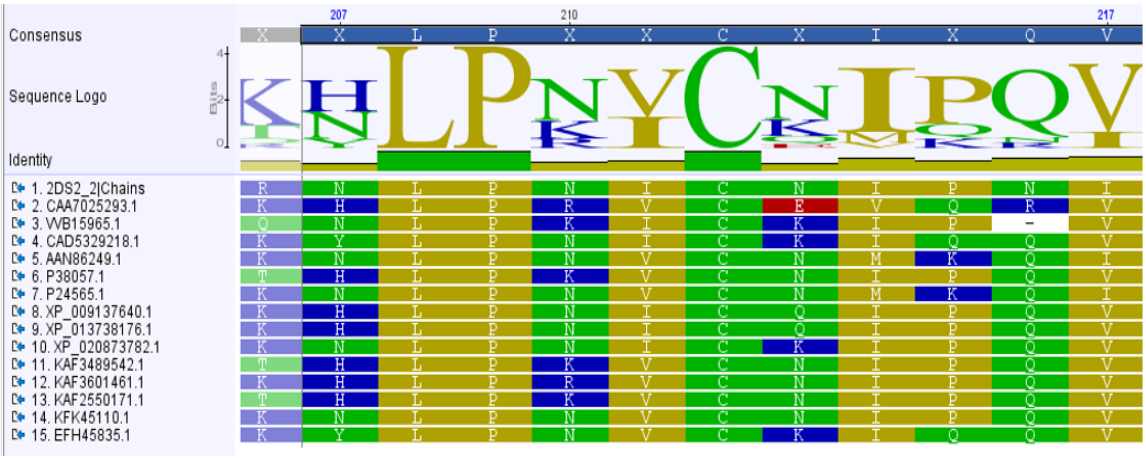



Figure 18: Identification of [NL/I] motif. The analysis performed in Geneious software.

In compliance with the results, only one sequence had this motif: KFK45110.1 (ArabisAlp) (Table 11). However, among the results of Mabinlin-II, were also identified three interesting sequences from the Cretan plant, *Brassica cretica* (KAF3489542, KAF3601461, KAF2550171, BrasCret1-3, respectively). Finally, it was decided to study these genes together with the other one.

Table 11. Results from BLASTP for Mab-II

| ACCESSION | Organism | Full-Length | Per. Identical | pI |
|---|---|---|---|---|
| 1.KFK45110.1 Hypothetical protein AALP_AA1G345700 | Arabis alpina | 183 | 48.61% | 10.01 |
| 2. KAF3489542.1 Hypothetical protein F2Q69_00052222 | Brassica cretica | 166 | 40.26% | 8.21 |
| 3. KAF3601461.1 Hypothetical protein F2Q69_00034154 | Brassica cretica | 180 | 42.86% | 8.21 |
| 4. KAF2550171.1 Hypothetical protein F2Q68_00033573 | Brassica cretica | 101 | 40.00% | 9.18 |

## 4.4 Bioinformatic analysis

In this chapter, docking experiments were conducted for all target proteins (ligands) and the T1R2-T1R3 receptor (Paragraph 4.4.3). To did that, we had downloaded and refined the 3D structures of the MNEI and Mab-II (PDB ID: 2O9U and 2DS2, respectively) and created and refined homology models for ArabisAlp, BrasCret1, BrasCret2, BrasCret3, and Cryza.M proteins (Paragraph 4.4.1). In regards to the receptor, a homology model was selected from the literature. (Paragraph 4.4.2).

### 4.4.1 Ligand preparation

#### 4.4.1.1 Homology modeling

The homology modeling was accomplished by YASARA. The software evaluated all possible templates, crystal structures for PDB for each protein (data not presented) and finally a hybrid model based on the best parts that emerged from the analysis was created. Hybrid models were rated with a quality Z-score, as shown in Figure 18-22. Z-scores was calculated as "the weighted averages of the individual Z-scores using the equation :
Overall = 0.145*Dihedrals + 0.390*Packing1D + 0.465*Packing3D.
Also, these scores represent the overall quality of the hybrid models, involving the good matching of the main chain of the proteins, the dihedral angles from the side chains, and the overall "packaging" of the proteins. According to the YASARA manual, z-scores range from -5 to> 0, considering prohibitive to optimal, respectively, to the target protein.

| Check type | Quality Z-score | Comment |
|---|---|---|
| Dihedrals | 0.958 | Optimal |
| Packing 1D | 0.284 | Optimal |
| Packing 3D | -1.155 | Satisfactory |
| Overall | 0.287 | Good |

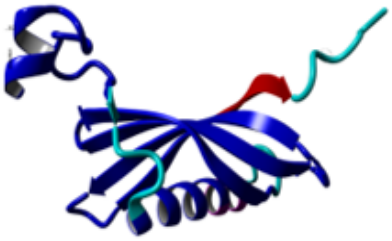


Figure 18 Quality results of Cryza.M hybrid model from YASARA: The table contains three quality factors: Dihedrals, packing 1D, Packing 3D, and the Overall Z-score. Dihedrals define the normality of dihedral angles conforming to the force field. Packing 1D is the normality of 1D distance-dependent packing interactions in the force field, including all molecules. Packing 3D is the normality of 3D direction-dependent packing interactions in the force field, for common atom types found in proteins only. On the right side is depicted as the hybrid model for Cryza.M as prepared by YASARA. Colored regions include the sequences used by the templates to create the final model.

The results for Cryza.M exhibited a good overall Z-score. Likewise, the 3D packing achieved the minimum score among the other parameters, although it was not a significant problem as this value was not prohibitive. According to these outcomes, this hybrid model could be used for further experimentation. Similar to Cryza.M, BrasCret1 (Figure 19) quality factors showing a good overall Z-score, reaching -0,714

| Check type | Quality Z-score | Comment |
|---|---|---|
| Dihedrals | -0,596 | Good |
| Packing 1D | -0,614 | Good |
| Packing 3D | -0,892 | Good |
| Overall | -0,741 | Good |

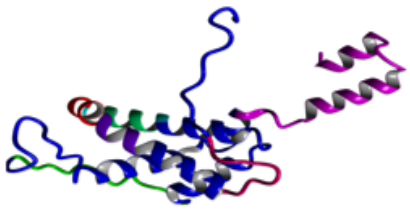


Figure 19 Quality results of BrasCret1 hybrid model from YASARA: The table contains three quality factors: Dihedrals, packing 1D, Packing 3D, and the Overall Z-score. Dihedrals define the normality of dihedral angles conforming to the force field. Packing 1D is the normality of 1D distance-dependent packing interactions in the force field, including all molecules. Packing 3D is the normality of 3D direction-dependent packing interactions in the force field, for common atom types found in proteins only. On the right side is depicted as the hybrid model for BrasCret1 as prepared by YASARA. Colored regions include the sequences used by the templates to create the final model.

| Check type | Quality Z-score | Comment |
|---|---|---|
| Dihedrals | 0.336 | Optimal |
| Packing 1D | 0.026 | Optimal |
| Packing 3D | -0,153 | Good |
| Overall | -0,012 | Good |

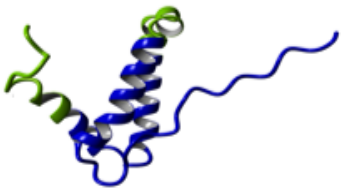


Figure 20 Quality results of BrasCret3 hybrid model from YASARA: The table contains three quality factors: Dihedrals, packing 1D, Packing 3D, and the Overall Z-score. Dihedrals define the normality of dihedral angles conforming to the force field. Packing 1D is the normality of 1D distance-dependent packing interactions in the force field, including all molecules. Packing 3D is the normality of 3D direction-dependent packing interactions in the force field, for common atom types found in proteins only. On the right side is depicted as the hybrid model for BrasCret3 as prepared by YASARA. Colored regions include the sequences used by the templates to create the final model

The results for BrasCret3 displayed a good overall Z-score, closely to the optimal score (0). Positive values or near zero indicated that the homology model looked better than a high-resolution X-ray structure. Nevertheless, the model included an optimal score for dihedrals and 1D packing. Figure 21 illustrated the results for BrasCret2. The model also had a good overall Z-score, near -0,5. The last model was performed for ArabisAlp (Figure 22). The result revealed a satisfactory Z-score, -1,062, classifying as the worst hybrid models among the previous.

| Check type | Quality Z-score | Comment |
|---|---|---|
| Dihedrals | 0,107 | Optimal |
| Packing 1D | -0,362 | Good |
| Packing 3D | -0,865 | Good |
| Overall | -0,528 | Good |

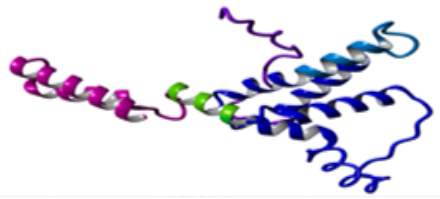


Figure 21.Quality results of BrasCret2 hybrid model from YASARA: The table contains three quality factors: Dihedrals, packing 1D, Packing 3D, and the Overall Z-score. Dihedrals define the normality of dihedral angles conforming to the force field. Packing 1D is the normality of 1D distance-dependent packing interactions in the force field, including all molecules. Packing 3D is the normality of 3D direction-dependent packing interactions in the force field, for common atom types found in proteins only. On the right side is depicted as the hybrid model for BrasCret2 as prepared by YASARA. Colored regions include the sequences used by the templates to create the final model

| Check type | Quality Z-score | Comment |
|---|---|---|
| Dihedrals | -0,232 | Good |
| Packing 1D | -1,117 | Satisfactory |
| Packing 3D | -1,274 | Satisfactory |
| Overall | -1,062 | Satisfactory |

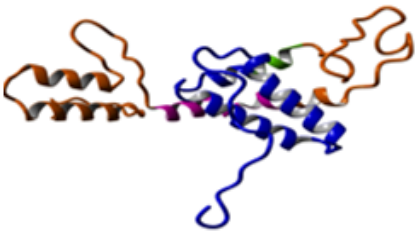


Figure 22. Quality results of ArabisAlp hybrid model from YASARA: The table contains three quality factors: Dihedrals, packing 1D, Packing 3D, and the Overall Z-score. Dihedrals define the normality of dihedral angles conforming to the force field. Packing 1D is the normality of 1D distance-dependent packing interactions in the force field, including all molecules. Packing 3D is the normality of 3D direction-dependent packing interactions in the force field, for common atom types found in proteins only. On the right side is depicted as the hybrid model for ArabisAlp as prepared by YASARA. Colored regions include the sequences used by the templates to create the final model

### 4.4.1.2 Evaluation and optimization of hybrid models and Crystal structures

To verify the stereochemical quality of the hybrid models as well as the two crystal structures for MNEI and Mab-II were performed an evaluation using MolProbity's web service. As described in detail in the methods section, the PDB files were uploaded in MolProbity, and the resulting evaluation data were collected and analyzed.

First was executed the analysis for Cryza.M. As seen in Table 12, there was a flip applied in the homology model (His121). Generally, the hybrid models' clash scores were good (colored green in the summary of the multi-criterion chart). However, protein sterics and geometry displayed significant deviations from the permissible limits. According to the Ramachandran plot, a plot that described the torsional angles, phi (ϕ) and psi (ψ), of the residues, 2.59% of the amino acids were in the allowed region while 0,93% were in the prohibitive region. In addition, the percentage of bad bonds and angles were high and deviated significantly from the desired limit. After the evaluation, the model was optimized at YASARA, the parameters and conditions are reported in method section, and the macro used was the md_refine.mcr file. Snapshot 12 had the lowest force field energy of all snapshots, and all data (Z-score) and force field energy are shown collectively in Table 12.

Table 12: Refinement Z-scores and force field energies for Cryza.M

| Check type | Quality Z-score | Comment |
|---|---|---|
| Dihedrals | 1.44 | Optimal |
| Packing 1D | -0.28 | Good |
| Packing 3D | -0.38 | Good |
| Overall | 0.26 | Good |
| Energy | -62764.72 | |

The overall Z-score remained at the same levels, obtaining a slight reduction from 0,287 to 0,26. Dihedrals and Packing 3D Z-scores improved while Packing 1D score worsen. Afterward, Cryza.M was evaluated through MolProbity. Data of the second evaluation were pictured in Table 13 .No flips were needed after the refinement. Clash scores remain the same, although the Ramachandran outliers decreased to zero.

Table 13 MolProbity structure evaluation data for homology models of Cryza.M before and after refinement.

| Summary Statistics | | Before Refinement | After Refinement |
|---|---|---|---|
| **All-Atom Contacts** | Clashscore, all atoms | **0,59** [100th percentile* (N=677, 2.03Å ± 0.25Å)] | **0,59** [100th percentile* (N=677, 2.03Å ± 0.25Å)] |
| | Poor rotamers | 0, 0,00% Goal: <0.3% | 3,3.49%, Goal: <0.3% |
| | Favored rotamers | 82,95.35% Goal: >98% | 82,95.35% Goal: >98% |
| | Ramachandran outliers | 1, 0.93% Goal: <0.05% | 0, 0.00%, Goal: <0.05% |
| | Ramachandran favored | 100,92.59% Goal: >98% | 104,96.30%,Goal: >98% |
| **Protein Geometry** | Rama distribution Z-score | -1,33 ± 0.68 Goal: abs(Z score) < 2 | ± 0.72 Goal: abs(Z score) < 2 |
| | MolProbity score | 1.16[100th percentile* (N=12152, 2.03Å ± 0.25Å)] | 1.36[99th percentile* (N=12152, 2.03Å ± 0.25Å)] |
| | Cβ deviations >0.25Å | 0, 0.00%, Goal: 0 | 0, 0.00%, Goal: 0 |
| | Bad bonds: | 6 / 877 , 0.68% Goal: 0% | 6 / 877 , 0.68% Goal: 0% |
| | Bad angles: | 8 / 1189,0.67%, Goal: <0.1% | 6 / 1189, 0.50%, Goal: <0.1% |
| **Peptide Omegas** | Cis Prolines: | 0 / 2, 0.00%,Expected: ≤1 per chain, or ≤5% | 0 / 2, 0.00%,Expected: ≤1 per chain, or ≤5% |
| **Additional validations** | Chiral volume outliers | 0/129 | 0/129 |
| | Waters with clashes | 0/0 0.00% | 0/0 0.00% |

The same analysis was performed for the other four hybrid models and the crystal structure of MNEI and Mab-II. The BrasCret3 hybrid model was evaluated using the MolProbity, and the results were illustrated in Table 14 .The model had a good clash score, and also, it was positive that there was no outlier in the Ramachandran plot, while the vast majority were located in favored regions.

Table 14 MolProbity structure evaluation data for homology models of BrasCret3 before and after refinement

| Summary | | Before Refinement | 1st Refinement |
|---|---|---|---|
| **All-Atom Contacts** | Clashscore, all atoms | 0, 100th percentile* (N=819, 1.70Å ± 0.25Å) | 0, 100th percentile* (N=819, 1.70Å ± 0.25Å) |
| | Poor rotamers | 2, 2.33% Goal: <0.3% | 0, 0.00% Goal: <0.3% |
| | Favored rotamers | 84 , 97.67% Goal: >98% | 84, 97.67% Goal: >98% |
| | Ramachandran outliers | 0, 0.00% Goal: <0.05% | 0, 0.00% Goal: <0.05% |
| | Ramachandran favored | 87, 95.60 % Goal: >98% | 90, 98.90% Goal: >98% |
| **Protein Geometry** | Rama distribution Z-score | -0,64 ± 0.81 Goal: abs(Z score) < 2 | 0.50 ± 0.81 Goal: abs(Z score) < 2 |
| | MolProbity score | 1.07 10th [percentile* (N=12152, 2.03Å ± 0.25Å)] | 0.50, 100th percentile* (N=9248, 1.70Å ± 0.25Å) |
| | Cβ deviations >0.25Å | 0, 0.00%, Goal: 0 | 0, 0.00%, Goal: 0 |
| | Bad bonds: | 2 / 745 , 0.13% Goal: 0% | 0/745, 0.45% Goal: 0% |
| | Bad angles: | 0/ 1012, 0.29%, Goal: <0.1% | 1 / 1012, 0.10% Goal: <0.1% |
| **Peptide Omega** | Cis Prolines: | 0 / 11, 0.00%,Expected: ≤1 per chain, or ≤5% | 0 / 11, 0.00% Expected: ≤1 per chain, or ≤5% |
| | Twisted peptides | - | - |
| **Additional validations** | Chiral volume outliers | 0/129 | 0/111 |
| | Waters with clashes | 0/0 0.00% | 0/0 0.00% |

The refinement was performed under the same conditions, and the results were shown in Table 15 .Snapshot 15 had the lowest force field energy and maximum quality. The overall Z-score and all the quality factors were upgraded. After refinement, no flips were needed, and the overall clash score remained the same, while the percentage of Ramachandran favored was improved.

Table 15 Refinement Z-scores and force field energies for BrasCret3

| Check type | Quality Z-score | Comment |
|---|---|---|
| Dihedrals | 1.41 | Optimal |
| Packing 1D | 0.12 | Optimal |
| Packing 3D | 0.54 | Optimal |
| Overall | 0.69 | Optimal |
| Energy | -52818.31 | |

The hybrid model for ArabisAlp, before the refinement, appeared one flip (GLN167), had a good Clashscore, two Ramachandran outliers, and 94,55% of the amino acids are in the favorable region (Figure 16). However, ArabisAlp refined structure was close to the end of the simulation, and thus a second refinement was applied, and the best-refined structure was snapshot 14.

Table 17 MolProbity structure evaluation data for homology models of ArabisAlp before and after 2nd refinement.

| Summary | | Before Refinement | 2nd Refinement |
|---|---|---|---|
| **All-Atom Contacts** | Clashscore, all atoms | 0,67, 99th percentile* (N=819, 1.70Å ± 0.25Å) | 0, 100th percentile* (N=819, 1.70Å ± 0.25Å) |
| | Poor rotamers | 4, 2.42% Goal: <0.3% | 2, 1.21% Goal: <0.3% |
| | Favored rotamers | 154 , 94.55% Goal: >98% | 163, 98.79% Goal: >98% |
| | Ramachandran outliers | 2, 1,10.00% Goal: <0.05% | 0, 0.00%, Goal: <0.05% |
| | Ramachandran favored | 177, 97.79 % Goal: >98% | 178, 97.80% Goal: >98% |
| **Protein Geometry** | Rama distribution Z-score | -0,85 ± 0.60 Goal: abs(Z score) < 2 | 0.03 ± 0.57 Goal: abs(Z score) < 2 |
| | MolProbity score | 1.55 89th [percentile* (N=12152, 2.03Å ± 0.25Å)] | 0.61 100th percentile* (N=9248, 1.70Å ± 0.25Å) |
| | Cβ deviations >0.25Å | 0, 0.00%, Goal: 0 | 0, 0.00%, Goal: 0 |
| | Bad bonds: | 2 / 1544 , 0.13% Goal: 0% | 1 / 1544, 0.06% Goal: 0% |
| | Bad angles: | 6/ 1087, 0.29%, Goal: <0.1% | 1 / 2087, 0.05% Goal: <0.1% |
| **Peptide Omegas** | Cis Prolines: | 0 / 15, 0.00%,Expected: ≤1 per chain, or ≤5% | 0 / 15, 0.00% Expected: ≤1 per chain, or ≤5% |
| | Twisted peptides | - | - |
| **Additional validations** | Chiral volume outliers | 0/214 | 0/214 |
| | Waters with clashes | 0/0 0.00% | 0/0 0.00% |

The data collected from YASARA are shown, in Table 18. The Overall Z- score after the 2nd refinement had a significant increase reaching 0,12, improvement confirming and from the Clashscore of MolProbity.

Table 18 Refinement Z-scores and force field energies for ArabisAlp

| Check type | Quality Z-score | Comment |
|---|---|---|
| Dihedrals | 1.63 | Optimal |
| Packing 1D | -1.04 | Satisfactory |
| Packing 3D | -0.23 | Good |
| Overall | 0.12 | Optimal |
| Energy | -106795.62 | |

The results for BrasCret1 and BrasCret2 were comparable to ArabisAlp. As seen in Table 19 and Table 20, both hybrid models' clash scores before refinement were good. Also, two outliers needed to be both in the two models, while 95.56% for BrasCret2 and 96.84% BrasCret1 of amino acids were located in allowed regions in the Ramachandran plot. Therefore, improvement was expected to be seen in the second evaluation after the refinement.

Table 19: MolProbity structure evaluation data for homology models of BrasCret1 before and after 2nd refinement.

| Summary Statistics | Before Refinement | 2 Refinement |
|---|---|---|
| Clashscore, all atoms | **0.38** 100th percentile*, (N=792, 1.69Å ± 0.25Å) | **0,** 100th percentile* (N=792, 1.69Å ± 0.25Å) |
| Poor rotamers | 5, 3,31%, Goal: <0.3% | 1, 0.63% Goal: <0.3% |
| Favored rotamers | 141, 93.38% Goal: >98% | 155, 98.10%Goal: >98% |
| Ramachandran outliers | 2, 1.20% Goal: <0.05% | 0, 0.00% Goal: <0.05% |
| Ramachandran favored | 153, 92.17% Goal: >98% | 173, 96.11% Goal: >98% |
| Rama distribution Z-score | -1.66 ± 0.62 Goal: abs(Z score) < 2 | -1.15 ± 0.54 Goal: abs(Z score) < 2 |
| MolProbity score | 1.51 90th percentile* (N=8895, 1.69Å ± 0.25Å) | 0.76, 100th percentile* (N=8895, 1.69Å ± 0.25Å) |
| Cβ deviations >0.25Å | 0, 0.00%, Goal: 0 | 0, 0.00%, Goal: 0 |
| Bad bonds: | 1 / 1368 0.07%, Goal: 0% | 4 / 1475, 0.27% Goal: 0% |
| Bad angles: | 2/ 1846,0.11%, Goal: <0.1% | 3 / 1983 0.15% Goal: <0.1% |
| Cis Prolines: | 0 / 12, 0.00%,Expected: ≤1 per chain, or ≤5% | 0 / 12, 0.00%,Expected: ≤1 per chain, or ≤5% |
| Twisted peptides | 1 / 181 0.55% Goal: 0 | 1 / 181 0.55% Goal: 0 |
| Chiral volume outliers | 0/194 | 0/211 |
| Waters with clashes | 0/0 0.00% | 0/0 0.00% |

Table 20:MolProbity structure evaluation data for homology models of BrasCret2 before and after 2nd refinement.

| Summary Statistics | Before Refinement | 2 Refinement |
|---|---|---|
| Clashscore, all atoms | **1,06** ,99th percentile* [(N=792, 1.69Å ± 0.25Å)] | **0,** 100th percentile* (N=792, 1.69Å ± 0.25Å) |
| Poor rotamers | 2, 1.27%, Goal: <0.3% | 1, 0.63% Goal: <0.3% |
| Favored rotamers | 153, 96.84% Goal: >98% | 155, 98.10%Goal: >98% |
| Ramachandran outliers | 2, 1.11% Goal: <0.05% | 0, 0.00% Goal: <0.05% |
| Ramachandran favored | 172, 95.56% Goal: >98% | 173, 96.11% Goal: >98% |
| Rama distribution Z-score | -1.15 ± 0.57 Goal: abs(Z score) < 2 | -1.15 ± 0.54 Goal: abs(Z score) < 2 |
| MolProbity score | 1.19, 99th percentile* (N=8895, 1.69Å ± 0.25Å) | 0.76, 100th percentile* (N=8895, 1.69Å ± 0.25Å) |
| Cβ deviations >0.25Å | 0, 0.00%, Goal: 0 | 0, 0.00%, Goal: 0 |
| Bad bonds: | 2 / 1470 , 0.14% Goal: 0% | 4 / 1475, 0.27% Goal: 0% |
| Bad angles: | 2 / 1992, 0.10% Goal: <0.1% | 3 / 1983 0.15% Goal: <0.1% |
| Cis Prolines: | 0 / 14, 0.00% Expected: ≤1 per chain, or ≤5% | 0 / 14, 0.00%,Expected: ≤1 per chain, or ≤5% |
| Twisted peptides | 1 / 181 0.55% Goal: 0 | 1 / 181 0.55% Goal: 0 |
| Chiral volume outliers | 0/212 | 0/211 |
| Waters with clashes | 0/0 0.00% | 0/0 0.00% |

For the BrasCret1 and BrasCret2, snapshot 1 had the lowest force field energy of all snapshots. All data (Z-scores) and force field energies are depicted collectively in Table 21 The force field energy for the twice refined structures of BrasCret1 and 2 got lower than the initial, suggesting remarkably advanced stability. The Z- scores, Dihedrals, and Packing 3D for both models indicated a slight gain, while the packing 1D score for BrasCret1 got worse, but the change was negligible. As expected, the Overall Z-score for BrasCret1 improved from -0.724 to 0.01, reaching the optimal, and for BrasCret2 from -0.528 to -0,15. After the second evaluation, the MolProbity data collected for the two models showed significantly improved Clash scores. The value for the BrasCret2 model changed from 1,06 to 0 and for BrasCret1 from 0,38 to 0 (green-colored boxes, respectively). Therefore, the outcome was positive, and the double refined structures became more reliable for further applications.

Table 21: Refinement Z-scores and force field energies for A) BrasCret1 and B) BrasCret2

A

| Check type | Quality Z-score | Comment |
|---|---|---|
| Dihedrals | 0.78 | Optimal |
| Packing 1D | -0.65 | Good |
| Packing 3D | -0.10 | Good |
| Overall | 0.01 | Optimal |
| Energy | -95038.90 | |

B

| Check type | Quality Z-score | Comment |
|---|---|---|
| Dihedrals | 0.79 | Optimal |
| Packing 1D | -0.76 | Good |
| Packing 3D | -0.48 | Good |
| Overall | -0.15 | Good |
| Energy | -101476.34 | |

Finally, the results from MolProbity analysis for the crystal structure of Mab-II were illustrated in Table 22. Five flips were applied: GLN8, GLN9, HIS23, GLN31, and ASN57. The class score showed a significantly low quality, attaining 15.05. However, the model had only one Ramachandran outlier and a good percentage of amino acids in the allowed location in the Ramachandran plot.

Table 22: MolProbity structure evaluation data for homology models of Mab-II before and after refinement.

| Summary Statistics | | Before Refinement | After Refinement |
|---|---|---|---|
| **All-Atom Contacts** | Clashscore, all atoms | 15.05, 37th percentile* (N=819, 1.70Å ± 0.25Å) | 0, 100th percentile* (N=819, 1.70Å ± 0.25Å) |
| | Poor rotamers | 3, 1.84% Goal: <0.3% | 0, 0.00% Goal: <0.3% |
| | Favored rotamers | 149 , 91.47% Goal: >98% | 174, 100% Goal: >98% |
| | Ramachandran outliers | 1, 0.55% Goal: <0.05% | 0, 0.00% Goal: <0.05% |
| | Ramachandran favored | 178, 98.34 % Goal: >98% | 179, 100% Goal: >98% |
| **Protein Geometry** | Rama distribution Z-score | -1.355 ± 0.55 Goal: abs(Z score) < 2 | 0.93 ± 0.87 Goal: abs(Z score) < 2 |
| | MolProbity score | 1.88 65th [percentile* (N=12152, 2.03Å ± 0.25Å)] | 0.50 100th percentile* (N=9248, 1.70Å ± 0.25Å) |
| | Cβ deviations >0.25Å | 0, 0.00%, Goal: 0 | 0, 0.00%, Goal: 0 |
| | Bad bonds: | 2 / 1544 , 0.13% Goal: 0% | 6 / 687 0.87% Goal: 0% |
| | Bad angles: | 18/ 1576, 0.29%, Goal: <0.1% | 2 / 2083, 0.10% Goal: <0.1% |
| **Peptide Omegas** | Cis Prolines: | 0 / 13, 0.00%,Expected: ≤1 per chain, or ≤5% | 0 / 9 0.00% Expected: ≤1 per chain, or ≤5% |
| | Twisted peptides | - | 1 / 80 1.25% Goal: 0 |
| **Additional validations** | Chiral volume outliers | 0/222 | 0/98 |
| | Waters with clashes | 0/0 0.00% | 0/0 0.00% |

After refinement snapshot 12 (Table 23) was selected and analyzed with the MolProbity platform, and the results show an improved structure. More accurately, the percentage of amino acids in the favorable region of the Ramachandran plot has increased, and no amino acid is in the prohibitive location. Likewise, the percentage of favored rotamers reached 100%, and no Ramachandran outlier was applied. However, four bad bonds were introduced into the structure of the protein without affecting the quality.

Table 23: Refinement Z-scores and force field energies for Mab-II

| Check type | Quality Z-score | Comment |
|---|---|---|
| Dihedrals | 1.65 | Optimal |
| Packing 1D | 0.17 | Optimal |
| Packing 3D | 0.38 | Optimal |
| Overall | 0.73 | Optimal |
| Energy | -51196.93 | |

For the crystal structure of MNEI, there were no results from the MolProbity as the platform could not ready the file for the crystal structure. Although the structure was refined under the usual conditions, and the results are demonstrated in Table 24.

Table 24: MolProbity structure evaluation data for homology models of MNEI after refinement.

| Summary Statistics | | After Refinement |
|---|---|---|
| **All-Atom Contacts** | Clashscore, all atoms | 0.62, 99th percentile* (N=263, 1.15Å ± 0.25Å) |
| | Poor rotamers | 1, 1.19% Goal: <0.3% |
| | Favored rotamers | 82, 97.62% Goal: >98% |
| | Ramachandran outliers | 0, 0.00% Goal: <0.05% |
| | Ramachandran favored | 93, 98.94% Goal: >98% |
| **Protein Geometry** | Rama distribution Z-score | -0,54 ± 0.72 Goal: abs(Z score) < 2 |
| | MolProbity score | 0.76 100th percentile* (N=9248, 1.70Å ± 0.25Å) |
| | Cβ deviations >0.25Å | 0, 0.00%, Goal: 0 |
| | Bad bonds: | 0 / 834 0.00% Goal: 0% |
| | Bad angles: | 0 / 1125, 0.10% Goal: <0.1% |
| **Peptide Omegas** | Cis Prolines: | 2/ 6 0.00% Expected: ≤1 per chain, or ≤5% |
| | Twisted peptides | 1 / 89 1.12% Goal: 0 |
| **Additional validations** | Chiral volume outliers | 0/109 |
| | Waters with clashes | 0/62 0.00% |

Snapshot 03 was used for further evaluation (Table 25). According to MolProbity, no flips were needed, the structure had a good class score, and 98.94% of residues were found in favored location in the Ramachandran plot.

Table 25: Refinement Z-scores and force field energies for MNEI

| Check type | Quality Z-score | Comment |
|---|---|---|
| Dihedrals | 1.79 | Optimal |
| Packing 1D | -0.04 | Optimal |
| Packing 3D | -0.19 | Optimal |
| Overall | 0.52 | Optimal |
| Energy | -69448.47 | |

### 4.4.2 Receptor preparation

So far, the T1R2-T1R3 sweet taste receptor did not crystalize yet. For this reason, in the current section, a new model of the sweet taste receptor tried to be built. In an effort to find the most recent templates for the whole sequence of the T1R2 and T1R3 monomers, using the BLAST tool, the results were unsuccessful, results that were in accordance with the

literature. Hence the sequences for the hT1R2 and hT1R3 subunits, including both the VFT domain and the CRD, were retrieved from UniProt (Access Code: Q8TE23 and Q7RTX0, respectively). Initially, the research started searching for possible templates through Blastp. Table 26 and Table 27 are presented the five best crystal structures with acceptable identity to T1R2 and T1R3, respectively. Based on the total scores for T1R2, were found 55 sequences with Per. identical from 23.66 % to 37.39%. To conclude, the sequences with the best max score, e-value, cover, and right molecular weight were chosen.

Table 26 :The top five possible templates for T1R3 from Blastp.

| NO | PDB ID | Total score | Query cover | E-value | Per. Ident |
|---|---|---|---|---|---|
| 1 | 5X2M_B | 322 | 79% | 1e-103 | 37.29% |
| 2 | EK5T_A | 274 | 91% | 1e-83 | 31.66% |
| 3 | EKES_A | 274 | 91% | 2e-83 | 31.49% |
| 4 | 5FBH_A | 226 | 80% | 5e-66 | 30.42% |
| 5 | 5X2M_A | 204 | 79% | 5e-59 | 30.17% |

Table 27:The top five possible templates for T1R2 from Blastp.

| NO | PDB ID | Total score | Query cover | E-value | Per. Ident |
|---|---|---|---|---|---|
| 1 | EK5T_A | 265 | 98% | 1e-80 | 30.58% |
| 2 | EKES_A | 263 | 98% | 9e-80 | 30.41% |
| 3 | 5X2M_B | 253 | 85% | 2e-77 | 33.89% |
| 4 | 5X2M_A | 236 | 86% | 2e-71 | 32.57% |
| 5 | 5FBH_A | 219 | 86% | 7e-64 | 29.28% |

In parallel, for T1R3, were seen 37 sequences with Per. identical from 25.05 % to 38.89%, and were selected the five with the best characteristics. However, all the templates showed a satisfactory coverage of their amino acid sequence with the two subunits. Obviously, the percentage identity was significantly low, assuming that the hybrid models will be of low quality. According to Bornscheuer, and Höhne the template identity should be > 30% of the target protein, resulting in a model with high structural quality (Bornscheuer, and Höhne, 2018). Moreover, the BLSTp analysis for both sequences, indicated that the new X-ray structures of medaka fish sweet taste receptor (5X2M_A/B) could be a progressive template, building a different model from the already existing in the literature.

Subsequently, the hybrid model was constructed in YASARA Figure 23 The program combined the best parts of the 25 models to obtain the hybrid model for the T1R3 and T1R2 extracellular domains.

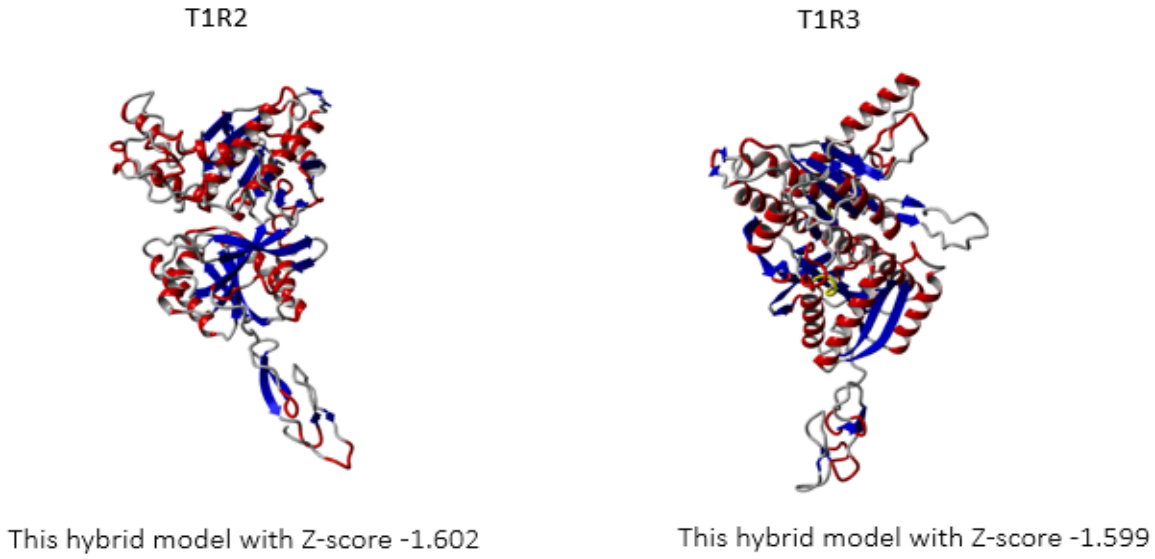



Figure 23 :Hybrid models for T1R2 and T1R3. Structure graphics were illustrated with the YASARA program.

Then, each model was further evaluated by quality Z-score (Table 28). As depicted in the tables below, the two hybrid models have a satisfactory Z-score indicating that the heterodimer will also have satisfactory quality.

Table 28: Results from the hybrid models of A) T1R2 and B)T1R3

**A**

| Check type | Quality Z-score | Comment |
|---|---|---|
| Dihedrals | -0.466 | Good |
| Packing 1D | -1,044 | Satisfactory |
| Packing 3D | -2,424 | Poor |
| Overall | -1,602 | Satisfactory |

**B**

| Check type | Quality Z-score | Comment |
|---|---|---|
| Dihedrals | -2.829 | Poor |
| Packing 1D | -0.938 | Good |
| Packing 3D | -1,770 | Satisfactory |
| Overall | -1,599 | Satisfactory |

The next step was the construction of the hT1R2-hT1R3 heterodimeric complex. For this purpose, the ClusPro 2.0 docking protein-protein webserver was used. ClusPro 2.0 algorithm determines the energy-like scoring function by the sum of inter-and intra- molecular contributions that approximate the thermodynamic stability of the ligand with the receptor according to the formula :

$E=0.40E_{rep}+-0.40E_{att}+600_{Eelec}+1.00E_{DARS}$

whereas the $E_{rep}$ and $E_{att}$ represent the repulsive and attractive contributions to the van der Waals interaction energy, $E_{elec}$ the electrostatic energy, and $E_{DARS}$ denotes the desolvation contribution.

In order to select the appropriate complex for our experiments, we searched in literature for significant interaction between the two subunits. As highlighted by the recent literature, the two subunits formed two significant intermolecular salt bridges between E126-T1R2 and R137-T1R3 and D127-T1R2 and R123 (Perez-Aguilar, et al., 2019). Also, hydrophilic interactions were mainly established by the amide residues N52-T1R2 and Q425-T1R3, although Y128_T1R3 also participates. Residue L51 from T1R2 formed a hydrophobic interaction with the residues L427_T1R3, L161_T1R3, P136_T1R3, and the –CH2– groups in the side chain of E428_T1R3 (Perez-Aguilar, et al., 2019).Furthermore, were identified several intramolecular interactions which are represented in the Supplementary (Appendix II). Between the two subunits, there were four intermolecular disulfide bonds and eight intramolecular (Kashani-Amin, et al., 2019). Two disulfide bonds were built between C233-C513 for T1R2 and C236-C517 for T1R3, two disulfide bonds were also formed between residues C59-C102 and C62-C103 for T1R2 and T1R3, respectively . Four in T1R2:C495-C514, C499-C517, C520-535, C538-C551 and C499-C518, C503-521, C524-C538, and C541-C554 in T1R3. So, to selected the best candidate for the T1R2-T1R3 complex, all the models from ClusPro, were examined to identify which one had the most interactions. The results concluded with the model of Figure 24.

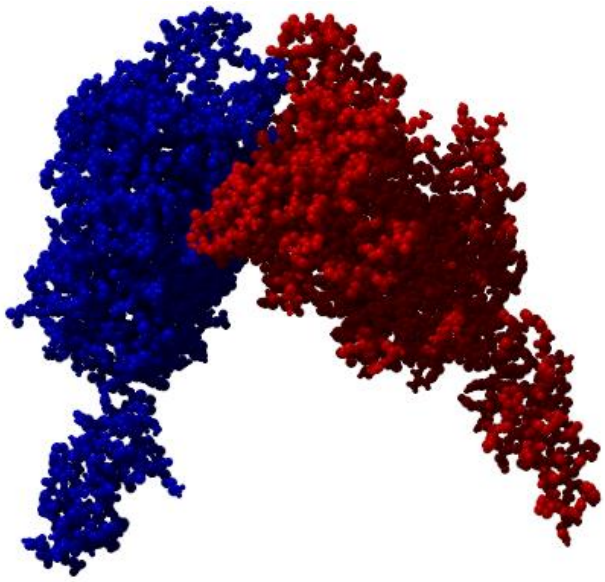


Figure 24 :The best candidate for the T1R2-T1R3 complex according to ClusPro, visualized using YASARA. The T1R2 subunit is presented with red and the T1R3 with blue. Structure graphics were illustrated with the PyMol program.

### 4.4.3 Docking simulations

To interpret the interaction between the sweet taste receptor T1R2/T1R3 and MNEI, Mab-II, and the five putative proteins, docking simulations were performed employing the program ClusPro 2.0. It has been demonstrated that MNEI and their mutant interacts with multiple sites in the surface of sweet taste receptors, and especially with the T1R2 subunit. Therefore, taking advantage of the already existing knowledge about MNEI-T1R2/T1R3 complexes, our docking experiments commenced with MNEI. Specifically, the modeled structure of sweet receptor T1R2/T1R3 was used as the receptor, crystal structure (PDB: 2O9U) after refinement was used as ligand, and the interface residues: N152, D169, E170, R172, D173, K174, R176, D188, D213, R217, D218, D456 and R457 for T1R2, R177, D190, R191 and D216 for T1R3, and D7, K36, R39, K43, R72 and R88 (6 7 9 13 28 36 39 43 65 72 88 92-96) for MNEI. Recent experimental data confirmed that MNEI interacts with the T1R2-T1R3 through the so-called wedge model. So, according to this mechanism, the protein interacts in an external cavity formed by the extracellular domains. The results from our docking experiments are illustrated in Figure 25, indicating that our complexes were apparently inconsistent with the wedge model.

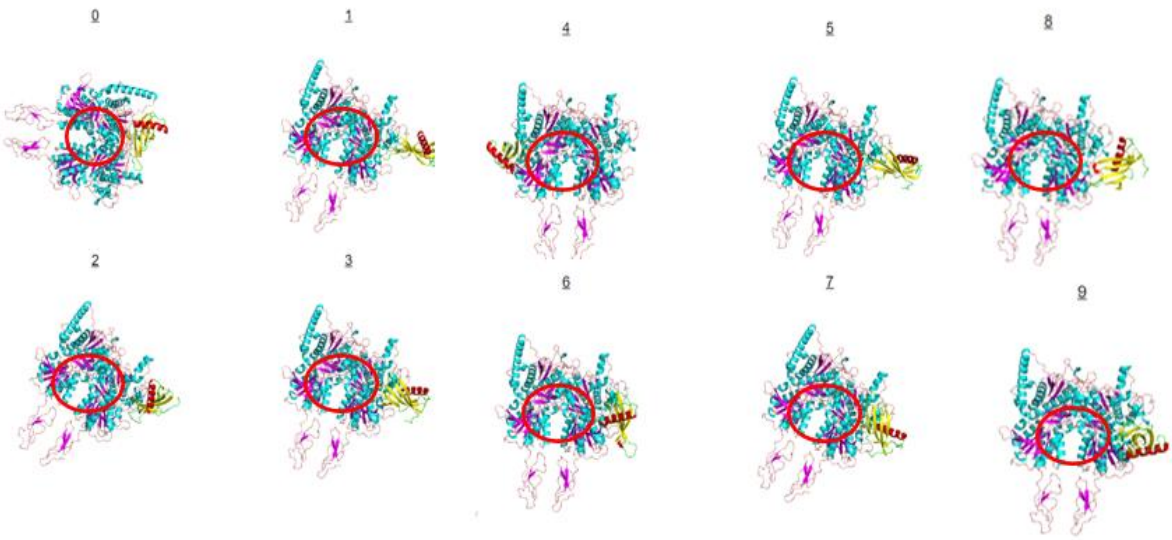



Figure 25: Results from docking simulations between MNEI and our model for T1R2/T1R3. The visualization performed in PyMol. The red cycle represent the interacting area according to wedge model.

It seems that MNEI did not fit in the external cavity and was docked in different locations. To further continue with our experiments, as a homology model for the receptor was used a model from Acevedo et coworkers.

After changing the model for the receptor, our investigation further continued under the same docking parameters. This time the results for MNEI were efficient. Figure 26 showed the interaction residues between the receptor and the sweet protein, which was outlined in Table 29. The choice from the ClusPro 2.0 results was based on the number of interactions. So, the complex with the higher number was selected. In agreement with the literature, our

model presents several interactions between critical amino acids, both from MNEI and the receptor, for sweetness.

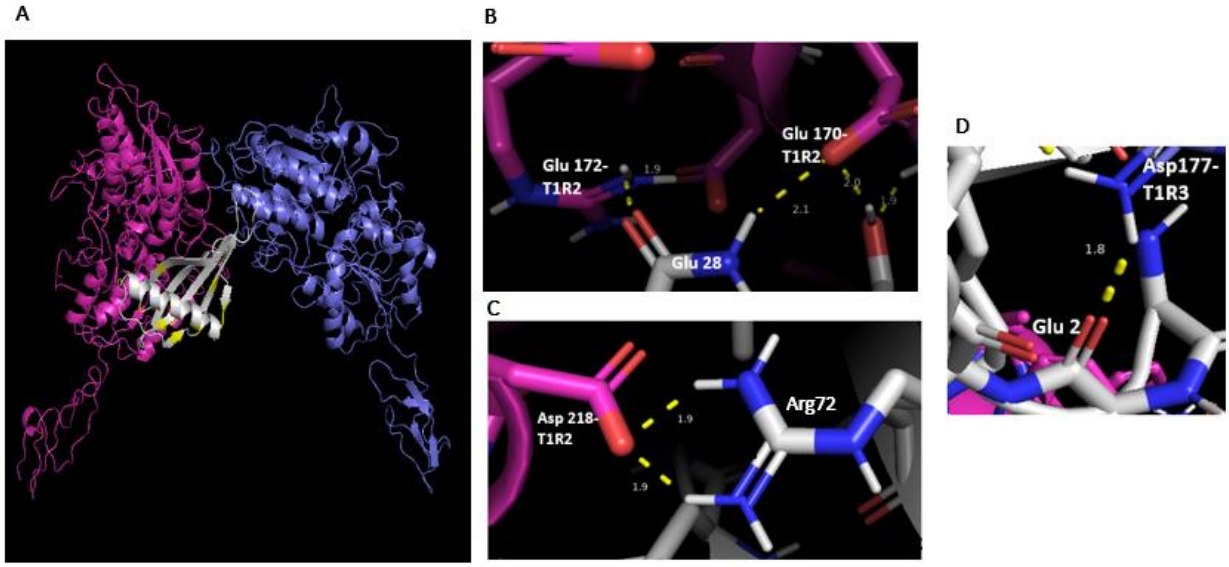



Figure 26: Interaction between MNEI and sweet taste receptor. (A) the docking complex structure of MNEI and receptor. The T1R2, T1R3, and MNEI were shown in cartoon models and colored in magenta, blue, and gray, respectively. With yellow are shown the critical amino acids. (B) interaction of Gln 28 of MNEI with the receptor. The residues were labeled and rendered as a stick model. The important hydrogen bonds were denoted as yellow dashed lines. (C) interaction of Arg72 of MNEI with the receptor; (D) interaction between Glu 2 of MNEI and Arg177 of T1R3. Structure graphics were illustrated with the PyMol program.

Remarkably, the Glu2 has a strong hydrogen bond with Arg177 of T1R3, as shown in Figure 26, two hydrogen bonds are formed between the Gln 28 and the Glu 170 (2.1 Å), and the side chain of Arg 172 (1.9 Å) of T1R2, respectively. Furthermore, Arg72 was directly double hydrogen-bonded with side chain O Asp218. Other hydrogen bonds can be found between Tyr 65, Pro 92, Gln 28, Arg 88, and the T1R2. Also, several newly established interactions between the protein and receptor could account for the interpretation of the interaction.

Table 29:Results from docking simulations between MNEI and the receptor. With red are colored the interacting pairs between critical amino acids, with light blue was colored critical amino acids for the receptor, and with green critical amino acids for the sweet protein.

| Critical Bonds | | Other Bond | Receptor |
|---|---|---|---|
| • Glu 2 O- Arg 177 HH11 | H-Bond 1.8 A | • Thr 45 HG1-Glu 178 OE2 | |
| • Tyr 58 HH- Arg 177 O | H-Bond 2.0 A | • Tyr 47 HH- Glu 178 OE2 | |
| • Glu 4 OE1- Arg 177 HH12 | H-Bond 1.9 A | • Glu 54 OE2- Lys 155 HZ1 | T1R3 |
| • Glu 4 OE2 - Arg 177 HH22 | H-Bond 1.8 A | | |
| • Tyr 65 HH- Gln 237 OE1 | H-Bond 1.9 A | • Lys 69 HZ2- Glu 238 OE2 | |
| • Tyr 65 OH- Gln 237 HE22 | H-Bond 2.1 A | • Lys 69 HZ1- Glu 225 OE1 | |
| • Pro 92 O- Gln 221 HE21 | H-Bond 2.0 A | • Lys 44 HZ3- Pro 110 O | |
| • Gln 28 O- Arg 226 HH22 | H-Bond 1.8 A | • Tyr 29 O- Arg 226 HH11 | |
| • Gln 28 O- Arg 226 HH12 | H-Bond 1.8 A | | |
| • Tyr 29 OH- Glu 170 OE2 | H-Bond 2.0 A | | |
| • Gln 28 HE21-Glu 170 OE2 | H-Bond 2.1 A | | T1R2 |
| • Phe 89 H- Glu 170 OE1 | H-Bond 2.0 A | | |
| • Gln 28 OE1- Arg 172 HH11 | H-Bond 1.9 A | | |
| • Arg 88 HE-Glu 145 OE1 | H-Bond 2.0 A | | |
| • Leu 87 O- Arg 176 HH11 | H-Bond 1.8 A | | |
| • Arg 88 HH22- Val 108 O | H-Bond 2.1 A | | |
| • Arg 88 HH12- Val 108 O | H-Bond 1.8 A | | |
| • Arg 88 HH21- Glu 145 O | H-Bond 2.0 A | | |
| • Arg 72 HH12- Asp 218 OD2 | H-Bond 1.9 A | | |
| • Arg 72 HH22- Asp218 OD2 | H-Bond 1.9 A | | |

Overviewing the literature only one work was found for docking tests for Mab-II with the sweet taste receptor. According to this research, Mab-II seemed to interact with the receptor in line with the ‘‘wedge model”, but in a different location from MNEI. Likewise, in this framework, they claimed that the unique motif (B54–B64) of Mab -II could be one of the critical sites for the interaction. Therefore, the selected complex from the 3D-docking experiment for Mab-II is represented in Figure 27. Notably, the model showed a hydrogen bond between two essential amino acids, the main chain O atom of Asn63 formed a hydrogen bond (1.8 Å) with the said chain H atom of Arg 177 of T1R3. Also, Asn152, one of the substantial amino acids of T1R2, was strongly interacting with the main chain O of Arg15 and Gln16. Another hydrogen bond involved Asn57 and Glu 178 of T1R3. Next, residues 124 and 129-131 seemed to interact with the T1R2, confirming the results from the previous work. In other words, it appeared that these crucial residues were in a pivotal position, supporting the interaction. A summary of these interactions is demonstrated in Table 30.

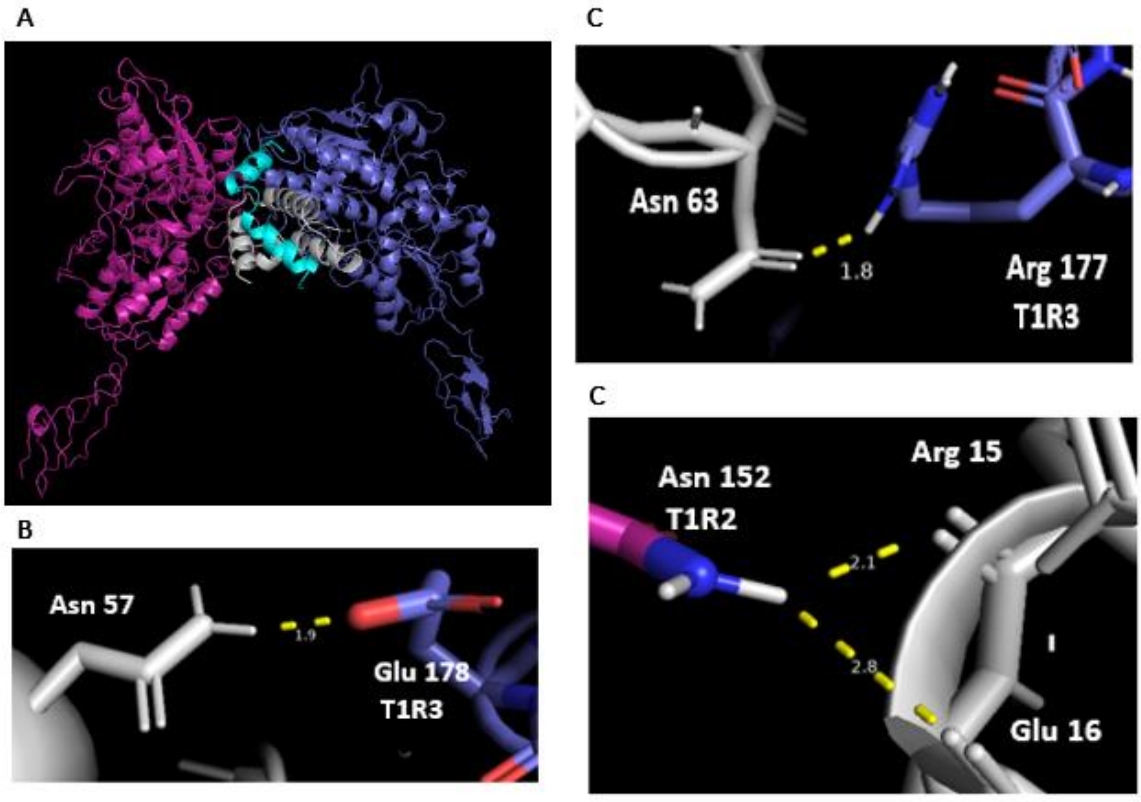



Figure 27: Significant interactions between Mab-II and sweet taste receptor. (A) the docking complex structure of Mab-II and receptor. The T1R2, T1R3, were shown in cartoon models and colored in magenta, blue. Mab-II also presented in cartoon model and A-chain colored light blue and B-chain gray.(B) interaction of Asn 57 of Mab-II with the receptor. The residues were labeled and rendered as a stick model. The critical hydrogen bonds were denoted as yellow dashed lines. (C) interaction of Asn63 of Mab-II with the receptor.(D) interaction between Arg15 and Glu of sweet protein and Asn152 of T1R2. Structure graphics were illustrated with the PyMol program.

Table 30: Results from docking simulations between Mab-II and the receptor. With red are colored the interacting pairs between critical amino acids, with light blue was colored critical amino acids for the receptor, and with green critical amino acids for the sweet protein.

| Critical Bonds | | Other Bond | | Receptor |
|---|---|---|---|---|
| • Asn 63 OD1 - Arg 177 HH11 | H-Bond 1.8 A | • Asn 57 HD21- Glu 178 OE2 | H-Bond 1.9 A | T1R3 |
| • Gln 16 O- Asn 152 HD21 | H-Bond 2.8 A | • Arg 4 O-Tyr 131 HH | H-Bond 2.0 A | |
| • Arg 15 O -Asn 152 HD21 | H-Bond 2.1 A | • Arg 4 O- TYR 131 HH | H-Bond 2.8 A | |
| | | • Arg 4 HH12- Ile 124 O | H-Bond 2.5 A | |
| | | • Arg 7 HH11- Asn 130 OD1 | H-Bond 1.9 A | T1R2 |
| | | • Arg 7 HH12 – Ser 129 O | H-Bond 1.8 A | |
| | | • Arg 7 HH22- Ser 129 O | H-Bond 1.9 A | |
| | | • Arg 7 HE- Tyr 131 OH | H-Bond 2.1 A | |
| | | • Gln11 HE22- Leu 156 O | H-Bond 2.1 A | |

The docking experiments for the five putative proteins were performed under the same conditions as the previous trials, and the selections were based on the number of interaction

pairs between critical residues from the receptor. The results are shown in Figures 28-32.The one-to-one contacts are overviewed in Table 30-35 in the Appendix section. Especially for BrasCret1 (Figure 30) His141, Lys144, Gln150, and Val15 amino acids that correspond to the unique motif (B54–B64) of Mab -II, were hydrogen-bonded with residues from the T1R3 subunit.

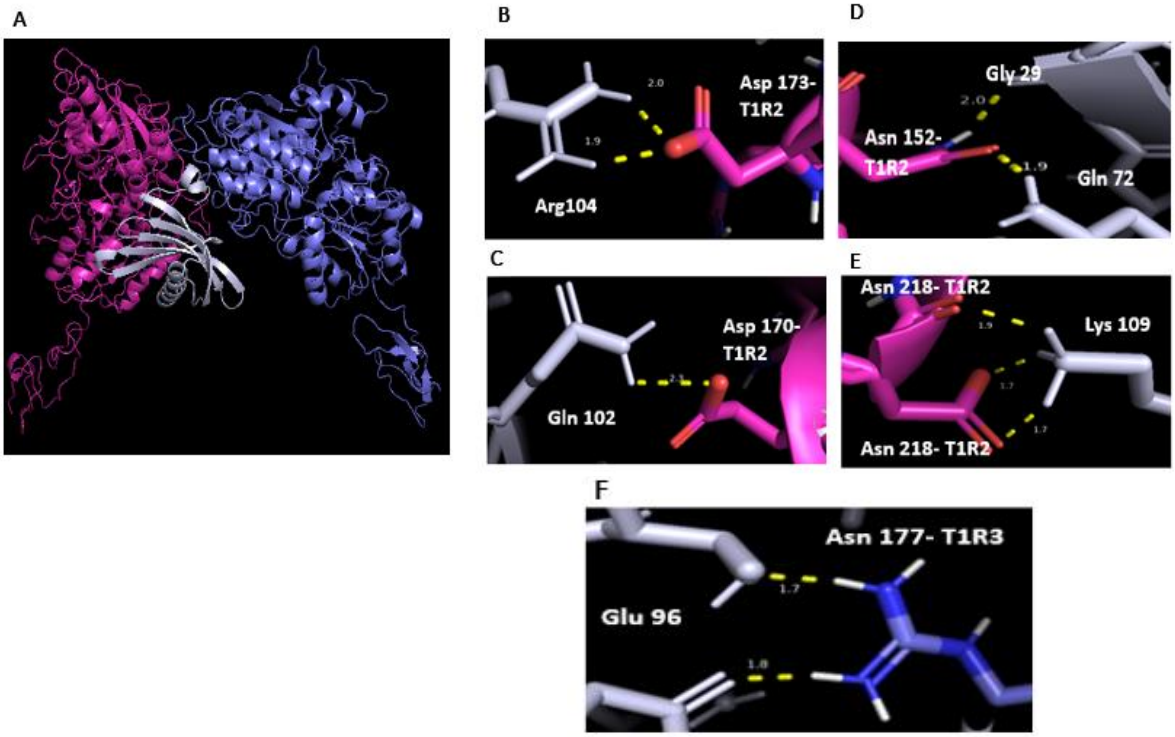



Figure 28: Interaction between Cryza.M and sweet taste receptor. (A) the docking complex structure. The T1R2, T1R3, and Cryza.M were shown in cartoon models and colored in magenta, blue, and gray, respectively. (B) interaction of Arg104 of sweet protein with the Asp172 of T1R2. The residues were labeled and rendered as a stick model. The necessary hydrogen bonds were denoted as yellow dashed lines. (C) interaction of Gln102 with Asp 170 of T1R2 (D) and (E) present interactions between Lys109 and Gly29, Gln 72 with critical amino acids from T1R2. (F) interaction of Glu96 with Arg177 of T1R3. Structure graphics were illustrated with the PyMol program.

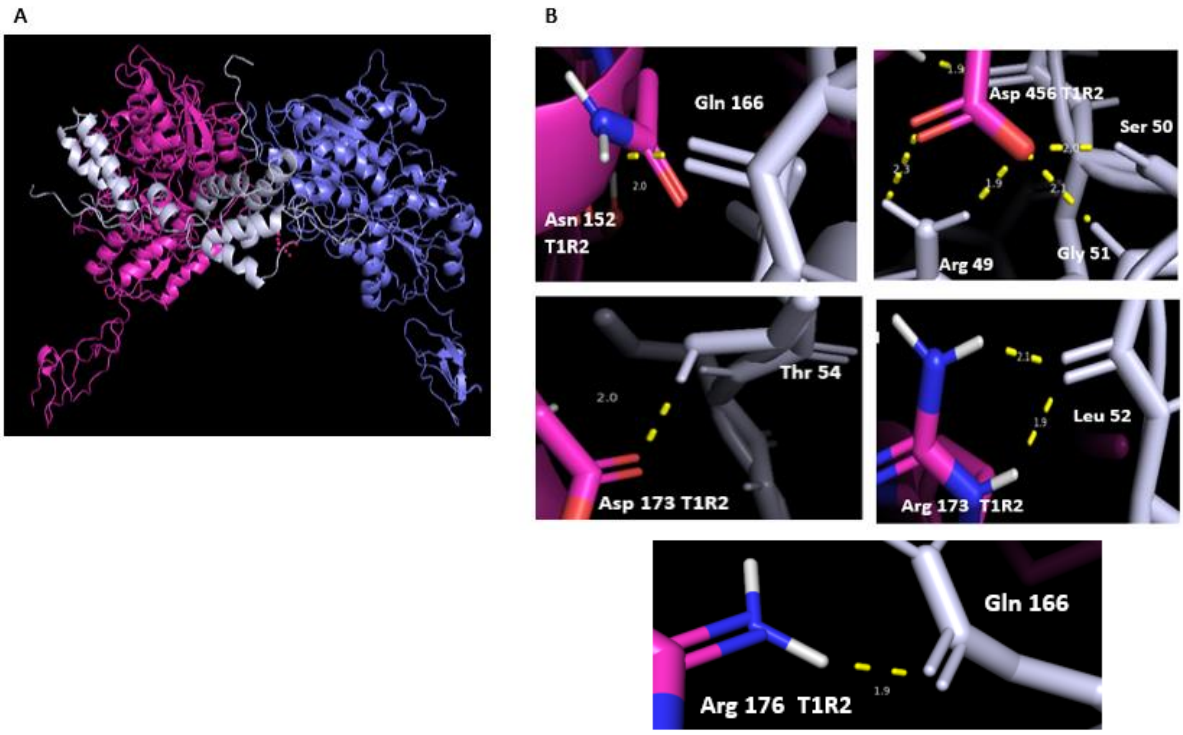



Figure 29: Interaction between ArabisAlp and sweet taste receptor. (A) the docking complex structure. The T1R2, T1R3, and ArabisAlp were shown in cartoon models and colored in magenta, blue, and gray, respectively. (B) interactions between BrasCret3 and critical amino acids of T1R2.The residues were labeled and rendered as a stick model. The necessary hydrogen bonds were denoted as yellow dashed lines.

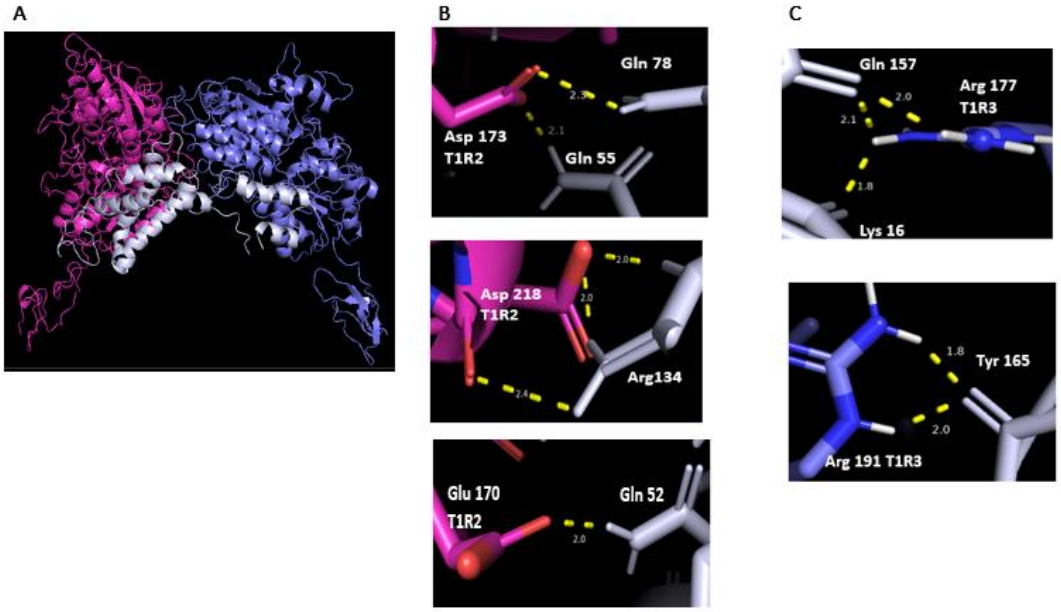



Figure 30 : Interaction between BrasCret1 and sweet taste receptor. (A) the docking complex structure. The T1R2, T1R3, and BrasCret1 were shown in cartoon models and colored in magenta, blue, and gray, respectively. (B) interactions between BrasCret1 and critical amino acids of T1R2.The residues were labeled and rendered as a stick model. The necessary hydrogen bonds were denoted

as yellow dashed lines. (C) interaction between BrasCret1 and significant amino acids of T1R3. Structure graphics were illustrated with the PyMol program.

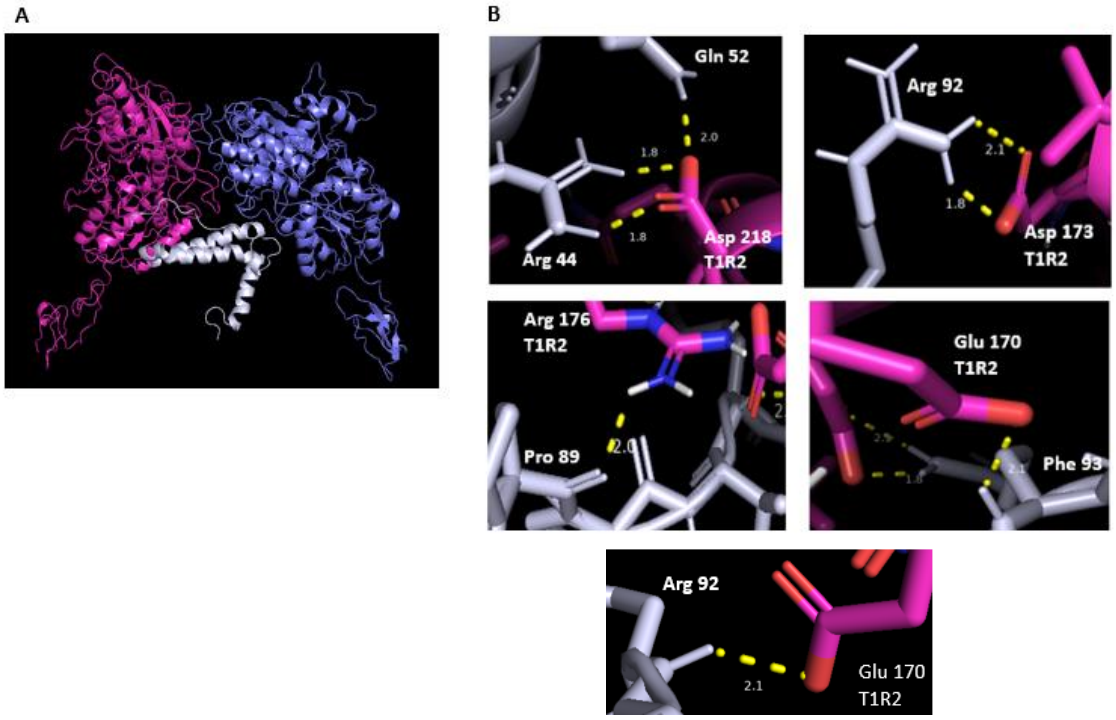



Figure 31: Interaction between BrasCret3 and sweet taste receptor. (A) the docking complex structure. The T1R2, T1R3, and BrasCret3 were shown in cartoon models and colored in magenta, blue, and gray, respectively. (B) interactions between BrasCret3 and critical amino acids of T1R2.The residues were labeled and rendered as a stick model. The necessary hydrogen bonds were denoted as yellow dashed lines. Structure graphics were illustrated with the PyMol program.

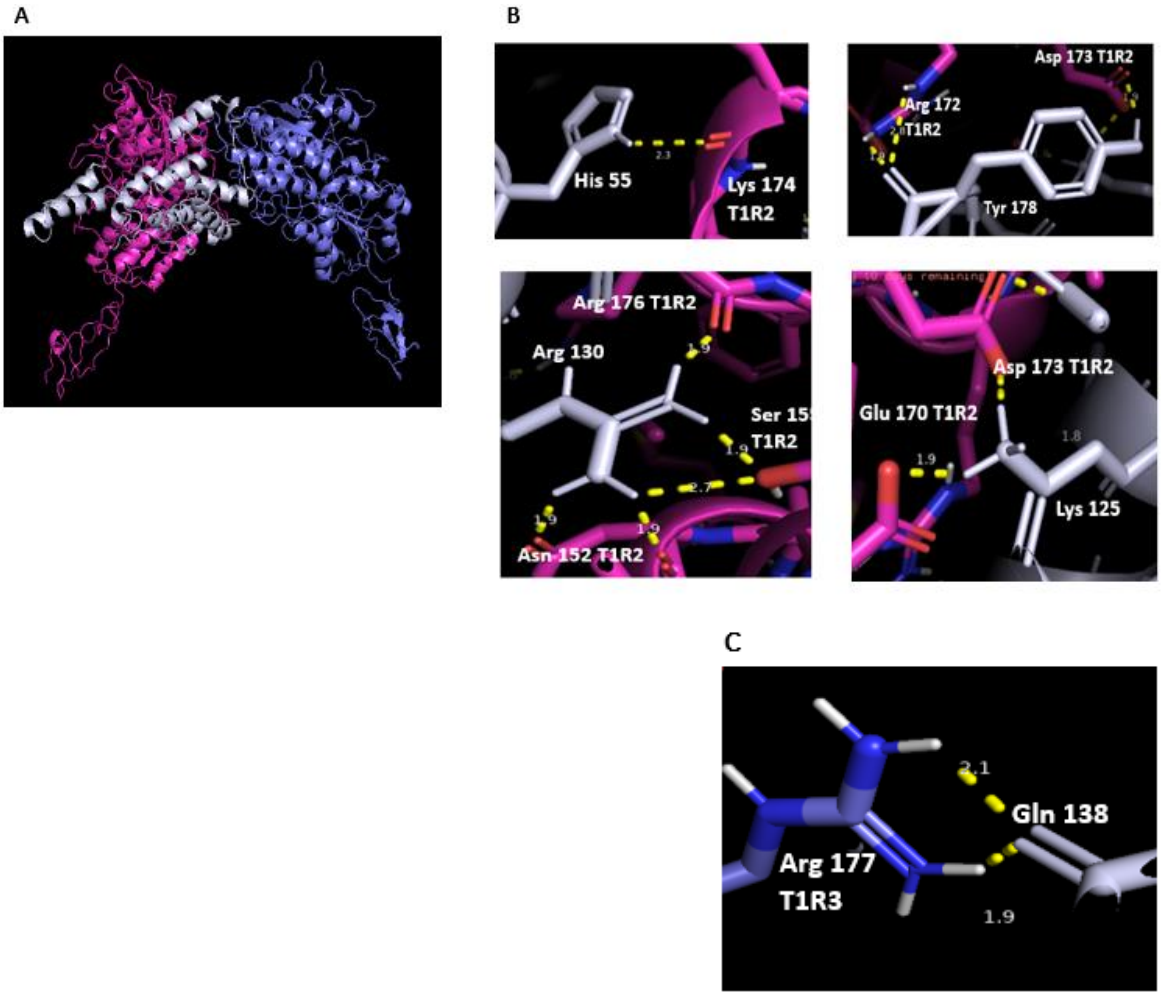



Figure 32: Interaction between BrasCret2 and sweet taste receptor. (A) the docking complex structure. The T1R2, T1R3, and BrasCret2 were shown in cartoon models and colored in magenta, blue, and gray, respectively. (B) interactions between BrasCret1 and critical amino acids of T1R2.The residues were labeled and rendered as a stick model. The required hydrogen bonds were denoted as yellow dashed lines. (C) interaction between BrasCret2 and significant amino acids of T1R3. Structure graphics were illustrated with the PyMol program.

Finally, the results from the docking analysis conclude in five potentially critical amino acids for the interaction. Arg177 of T1R3 and Alu170, Asp173 of T1R2 were found to interact in five of the seven complexes, while Asn152 and Asp 218 of T1R2 were found in four. Particularly, Arg177 was interacting with critical amino acids for MNEI and Mab-II.

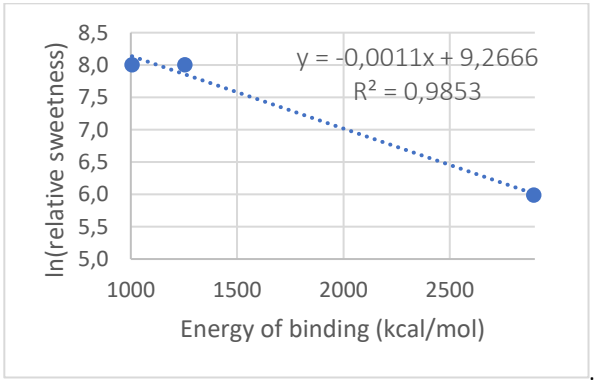
Eventually, in order to encounter more information for the elucidation mechanism of sweet taste, utilizing these complexes, a probable correlation of binding energy and the sweet sensation was performed. The relative sweetness of the evaluated sweet proteins was obtained from the literature (Izawa, et al.,2010) (Table 31). As illustrated in Figure 33A, the calculated binding energy negatively correlated with the sweetness intensities for sweeteners. Evidence from Table 31 showed that Mab-II had the greater binding energy (2895,15 kcal/mol) and the lowest relative sweetness, while MNEI and Thaumatin had similar relative sweetness, higher than Mab-II, appeared significantly lower binding energies.

Table 31 : Calculated Binding energies and the relative sweetness.

| Proteins | Binding Energy (kcal/mol) | ln(relative sweetness)[a] |
|---|---|---|
| Thaumatin | 1254 | 8,0 |
| MNEI | 1005 | 8,0 |
| Mab-II | 2895 | 5,9 |
| ArabisAlp | 2596 | 6,4 |
| BrasCret1 | 1068 | 8,1 |
| BrasCret2 | 1205 | 7,9 |
| BrasCret3 | 1531 | 7,6 |
| Cryza.M | 511 | 8,7 |

a: Izawa, et al.,2010

These results were in accordance with similar analyses performed by Acevedo and co-workers (Acevedo, et al., 2018) (Figure 33B). In their analysis, they used only four points to build their equation, as there are only eight sweet taste and taste-modifying proteins. In our analysis, three points are used, as we research, the plant-based sweet taste proteins and there are only five. Based on the equation (Figure 33A), the relative sweetness for the putative proteins was calculated, giving a quick assessment. Subsequently, the relative sweetness needs to be examined with the sweet taste assay to validate our results



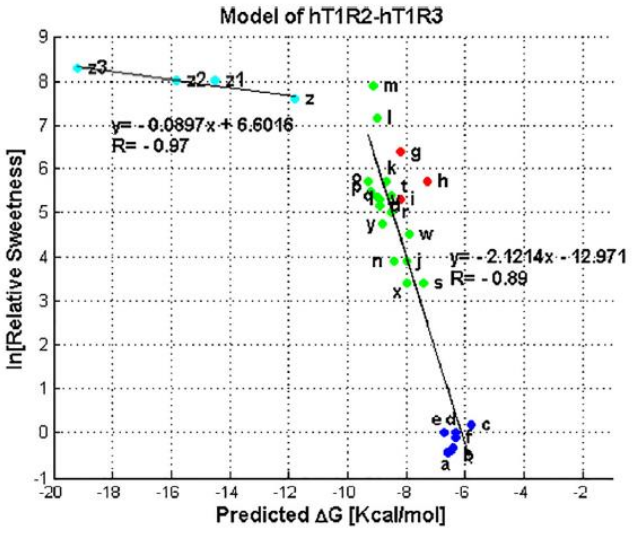




Figure 33: A) Relationship between relative sweetness of sweeteners and the calculated binding energy from molecular docking.B) Relationship between relative sweetness of sweeteners and ΔG binding (binding energy) calculated from molecular docking onto the hT1R2-hT1R3 sweet receptor. The data corresponds to: (a) glucose, (b) galactose, (c) fructose, (d) xylose, (e) sucrose, (f) tagatose, (g) sucralose, (h) saccharin, (i) aspartame, (j) monatin 2S4,S,(k) monatin 2S4,R, (l) monatin 2R4,S, (m) monatin 2R4,R, (n) glycyrrhizic acid, (o) mogroside V, (p) stevioside, (q) reb A, (r) reb B, (s) reb C, (t) reb D, (u) reb E, (v) reb F, (w) steviolbioside, (x) dulcoside A, (y) rubusoside, (z) thaumatin, (z1) brazzein, (z2) monellin and (z3) neoculin.

## 4.5 Heterologous expression in *E.coli*

The first technological objective was the establishment of the protocol for the heterologous expression of the seven target genes. In the present master thesis, specific plasmid vectors, *E. coli* strains, and several chemicals were used to validate the suitable conditions for the expression and the proper folding.

The gene of the known sweetener proteins (MNEI and Mab-II) and putative sweetener proteins (Cryza.M, BrasCret1, BrasCtret2, BrasCret3, and ArabisAlp) were ordered as synthetic genes, codon-optimized for the expression host. The bacterial constructs are situated downstream of the T7 expression and IPTG as inducer. 6xHis-tag had already been added during cloning at the 5 'end of the genes from the cloning vector. The affinity tags, against the Nickel column, were used for easy purification. Then, we selected the *E. coli* cells that are suitable for each protein. For MNEI and Cryza.M, we choose BL21 (DE3), while for the other disulfide-bonded proteins, the SHuffle T7. SHuffle T7, chemically competent cells, support the formation and the correct folding in proteins with multiple disulfide bonds in the cytoplasm.

### 4.5.1 Small scale expression test

After the successful transformation, small-scale expression tests were performed to examine the protein expression. The expressions of the target proteins were observed by SDS-PAGE, under denaturing conditions. Comparing the profile of proteins before and after induction, we saw that the intensities differ, indicating the proper lysate. As illustrated in Figure 34 Mab-II, displayed a strong protein band at 17,2 kDa that matches its molecular weight at sample 3, indicating that the expression performed at the insoluble fraction.

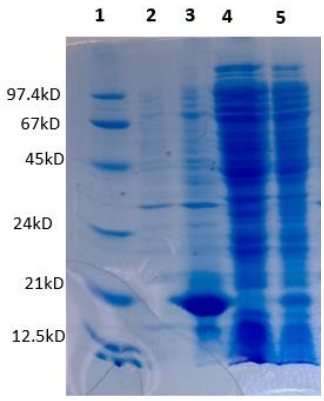



Figure 34: 15 % SDS-PAGE of (A) Mabinlin-II (17,2 kDa), expressed in *E.coli*, Shuffle T7 : 1) marker, 2) pellet before induction 3) pellet after induction, 4) soluble before induction, 5)soluble after induction

ArabisAlp and BrasCret3 presented bands at 22,3 kDa and 12,4 kDa, respectively (Figure 35). So similar to Mab-II, these two proteins were also expressed in the pellet. Figure 36 exhibited the SDS-PAGE for BrasCret1 and 2. According to sample 7, BrasCret1 had a clear band at 19,9 kDa, revealing the aggregation of the protein.

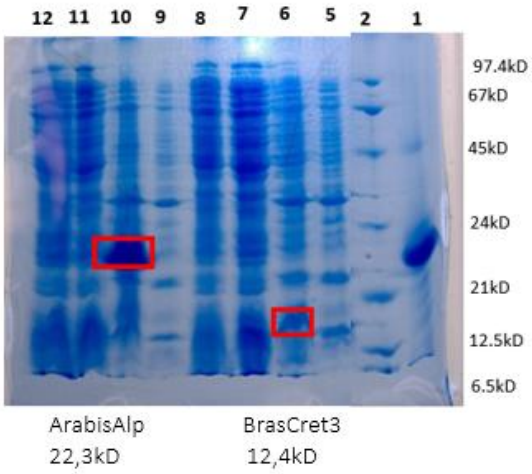



Figure 35: 15 % SDS-PAGE of (E) BrasCret3 (12,387 kDa),and ArabisAlp (22,334 kDa) expressed Shuffle *T7*, 2) marker, 5&9) pellet before induction 6&10) pellet after induction, 7&11) soluble before induction, 8&12)soluble after induction

However, the expression of MNEI was successful. It was clear that the protein is presented in sample 10 at 12,5 kDa (Figure 37), and there was no corresponding strong band for the before induction fraction (sample 9).

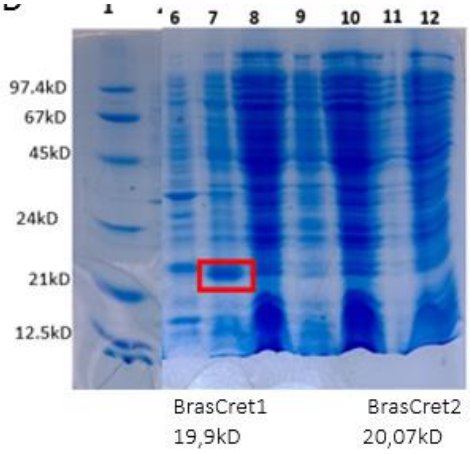



Figure 36 :15 % SDS-PAGE of (C) BrasCret1 (19,972 kDa),and BrasCret2 (20,07 kDa) expressed Shuffle *T7*, 1) marker, 6&9) pellet before induction 7&10) pellet after induction, 8&11) soluble before induction, 9&12)soluble after induction

In Figures 38 (samples 2-5) and 36 (samples 9-12), non Cryza.M nor BrasCret2 were found, concluded that the expressions were unsuccessful.

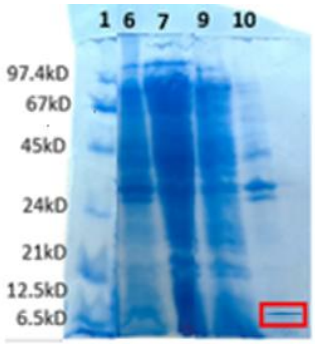



Figure 37: 15 % SDS-PAGE of (B) MNEI (12,5 kDa), expressed in *E.coli*, BLE21 : 1) marker, 6) pellet before induction 7) pellet after induction, 9)soluble before induction, 10)soluble after induction

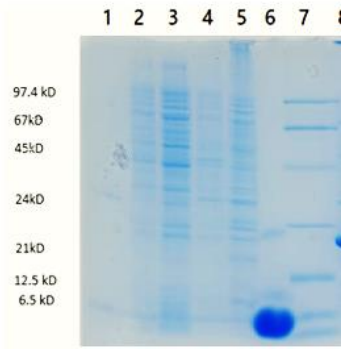



Figure 38: 15 % SDS-PAGE of (C) Cryza.M (13,810 kDa), expressed in *E.coli*, BLE21 : 1)-, 2) pellet before induction, 3) pellet after induction, 4)soluble before induction, 5)soluble after induction, 6) Lysozyme (14,4 kDa), 7) Marker

### 4.5.2 Solubility teste: Co-expression of target genes with chaperones

Insoluble protein expression was one of the various drawbacks for the heterologous gene expression in E. coli. The insoluble protein usually aggregated and formed inclusion bodies (IBs). Inclusion bodies are inactive misfolded protein aggregates that form due to improper intra- or inter-molecular interactions. In our first attempt to increase the solubility, we co-expressed our genes (Mab-II, BrasCret1, BrasCret2, BrasCret3, and ArabisAlp) with chaperons. Molecular chaperones are involved in protein folding, reducing the rate of protein synthesis and the formation of inclusion bodies, and maximizing soluble protein. Plasmids pG-Tf2 and pKJE7 were used in this work. The analysis proved that none of the proteins were expressed in the soluble fraction. We followed a one-step protocol, transforming, simultaneously, a chaperone plasmid and the target protein. These results are displayed in Figure 39 samples 3, 8, 13, 18, and 21. According to the co-expression the chaperones did not prevent the formation of inclusion bodies or insoluble aggregates.

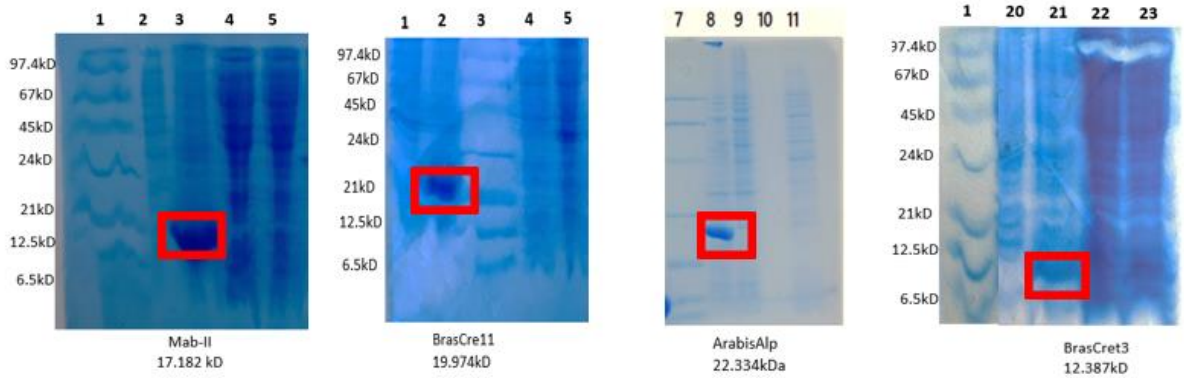



Figure 39: 15 % SDS-PAGE of Mab-II, BrasCret1, BrasCret2, ArabisAlp, and BrasCret3 1, 7, 14) marker, 2,12,17,9, and20) pellet before induction, **3, 8, 13, 18, and 21** pellet after induction, 4,15,19,10,and 22) soluble before induction, 4, 11, 16, 20,and 23 soluble after induction

### 4.5.3 Solubility teste: Cell lysis

Among other approaches to increasing solubility, the selection of the lysis buffer seems to be a dominant factor. In our second solubility test, the optimal lysis buffer for each protein was tested. The harvesting culture were analyzed with five different cell lysis buffers. The previous lysis buffer was a sodium phosphate buffer with pH=8. This pH was indicated as the proper since that differs by one unit from the pI of the proteins. So, four additions (sodium chloride-NaCl, glycerol, Triton-X-100, and Ethylenediaminetetraacetic acid-EDTA) were performed in this buffer attempt to increase the solubility. NaCl controls the pH and osmolarity of the lysate. The ionic strength is correlated with protein solubilization. The concentrations of the salts should be sufficient to disrupt interactions of soluble components from non-soluble cellular. So, examined two concentrations 100 mM and 200 mM. Then, glycerol was added in these lysis buffers to reduce the number of hydrophilic interactions that could interfere with the proper folding of a protein. In lysis buffers three and five, triton-X-100, non-ionic detergents, was used to extract protein and other cellular organelles or to permeabilize the living cell membrane for transfection. Also, in lysis four and five, EDTA was selected to prevent proteolysis. The last substitution was the increment of the amount of the buffer, about 10 mL of lysis buffer per gram of cells. Considering the previous expression's results, was decided to test only the after-induction fractions. As illustrated in Figure 40, Mab-II appears protein bands in the soluble samples (2, 4, 6, 8, and 10) at the corresponding molecular weight.

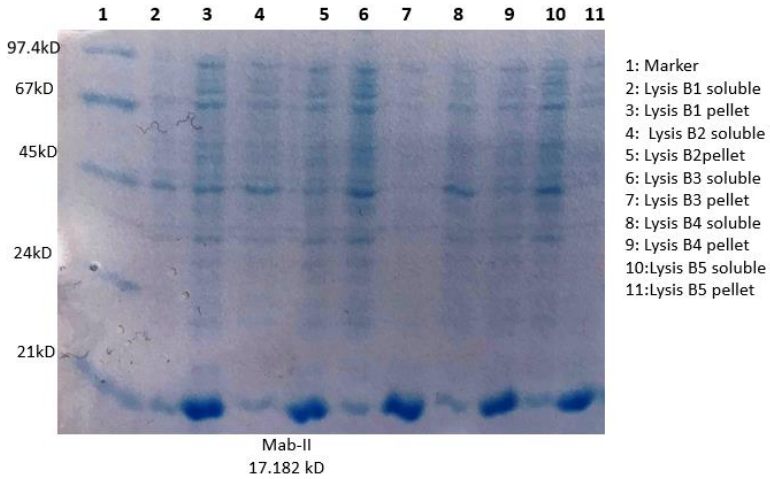



Figure 40 : 15 % SDS-PAGE of Mabinlin-II (17.182 kDa) with different Lysis Buffer

The experiments with Mabinlin-II expression, show that the most crucial parameter was the proper dilution of the proper dilution of the cells, and all 5 lysis buffers yielded similar results. However, again the most protein was in insoluble form. In parallel, lysis buffers 3 and 5 increase the solubility of BrasCret1 and 3, according to samples 6, 10, and 1,5, proportionately (Figure 41).

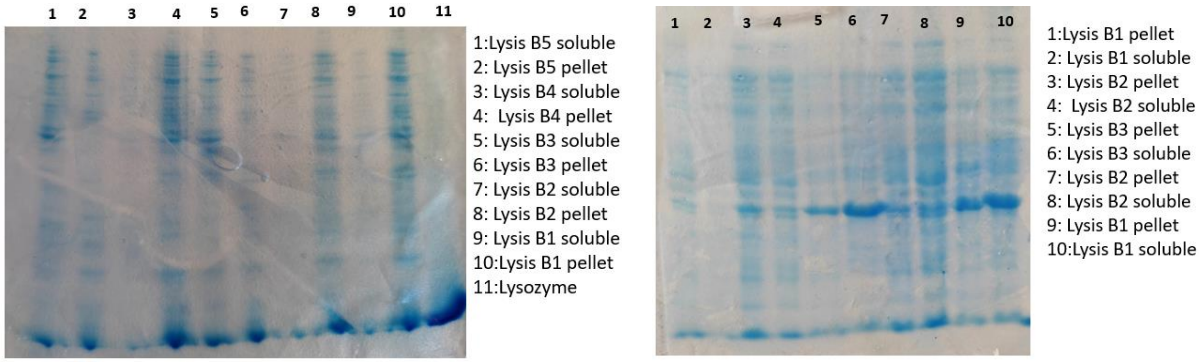



Figure 41: 15 % SDS-PAGE of BrasCret3 (12.387 kDa)on the left and BrasCret1 (9.974 kDa) on the right, with different Lysis Buffer

These shreds of evidence demonstrate the value of Triton-X-100. The last ArabisAlp had the more successful outcomes. As depicted in Figure 42, ArabisAlp displayed bands only in the soluble samples (2, 4, 6, 8, and 10).

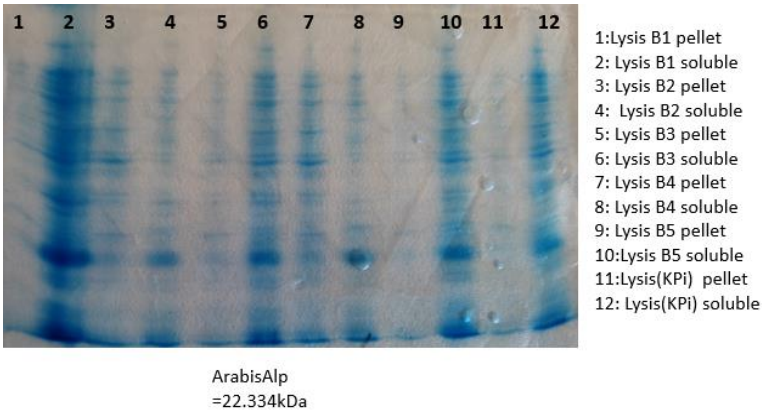



Figure 42 : 15 % SDS-PAGE of ArabisAlp (22.334 kDa) with different Lysis Buffer

### 4.5.4 Large scale production and purification via metal affinity chromatography

Large scale expressions (1 L culture) for MNEI, Mab-II, BrasCret1, BrasCret3, and Arabis were prepared under standard expression conditions. Following the cultivation, cell disruption was performed, using the proper lysis buffer for each protein (Table 6). For the purification, metal affinity chromatography in the FPLC system, was used and the desalting of the pure fraction was performed with 10 kDa concentrators for Mab-ii, BrasCret1, and Arabis and 3 kDa for MNEI and BrasCret3, by washing it three times with KPi buffer (5 mM, pH 8). The analysis of the purification performed by SDS-PAGE. Figure 43 illustrated the results for ArabisAlp (4,5), BrasCret3 (1-3), and Mab-II (sample 9-11). According to sample 11, a strong band at 17 kDa was visible, corresponding to the molecular weight of Mab-II. In the flow-through fraction (sample 10) and the wash-fraction, no Mab-ii was found, indicating that the expression and purification were successful. Similar to Mab-II, MNEI appeared clear bands only in the elution fraction, sample 5 Figure 44. In the case of the putative proteins, unfortunately, the purification was unsuccessful as the elution fractions were not clean enough. Figure 43 present also the results for ArabisAlp.

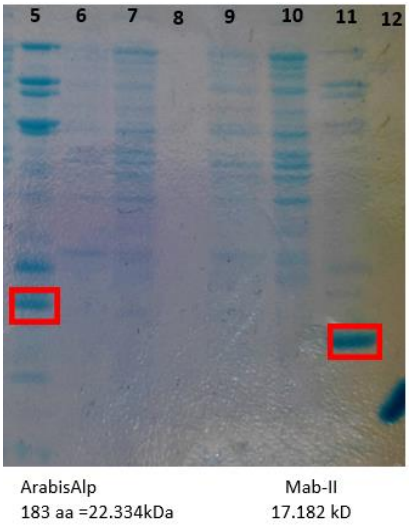



Figure 43: 15 % SDS-PAGE of Mabinlin-II (17,182 kDa) after purification. Samples: 5) ArabisAlp pure, 6) ArabisAlp washing, 7) ArabisAlp flow through, 9) Mabinlin-II washing, 10) Mabinlin-II Flow through, 11) Mabinlin-II pure, 12) marker (Lysozyme 14,4 kDa)

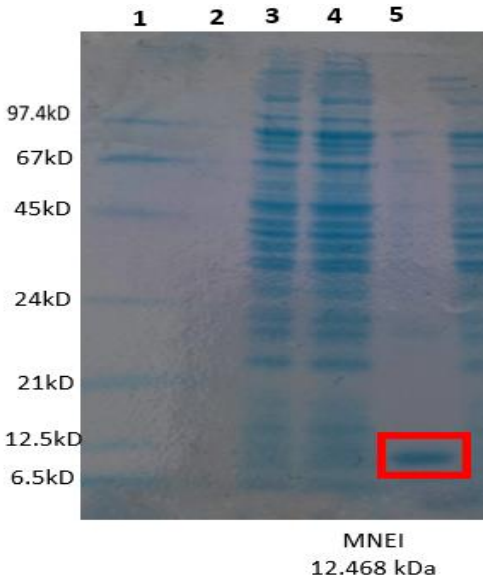



Figure 44: 15 % SDS-PAGE of MNEI (12,486 kDa) after purification. Samples: 1) Marker 3) MNEI washing, 4) MNEI flow through, 5) MNEI pure

ArabisAlp had a good affinity to the Nickel resin, although the pure fraction, sample 5, showed additional bands. These bands could be proteins with some affinity to the Nickel or chaperones that copurified. During heterologous proteins' expression, the chaperons that already exist in the E.coli[9], usually bind to the target protein with very high affinity. Therefore, when purifying the desired protein, the chaperone remains attached to and copurified. As evident from Figure 45, BrasCret1 and BrasCtre3, facing the same problem with ArabisAlp.

[9] These include, but may not be limited to DnaK (Mr ~ 70 000), DnaJ (Mr ~ 37 000), GrpE (Mr ~ 40 000), GroEL (Mr ~ 57 000), and GroES (Mr ~10 000).

Attempting to overtake this problem, a separate step was added after the desalting, using Amicon® Ultra Centrifugal Filters MWCO 30 kDa. However, the results from the separation were not satisfactory. Therefore, this extra cleaning step was discarded. Finally, the concentrations of purified Mab-II and MNEI were determined with the Bradford method. This method was more reliable as the absorption at 280 nm is caused only by aromatic residues of the protein. The yield of MNEI is 0,72 mg/L of cultivation and for Mab-II 2,1 mg/L.

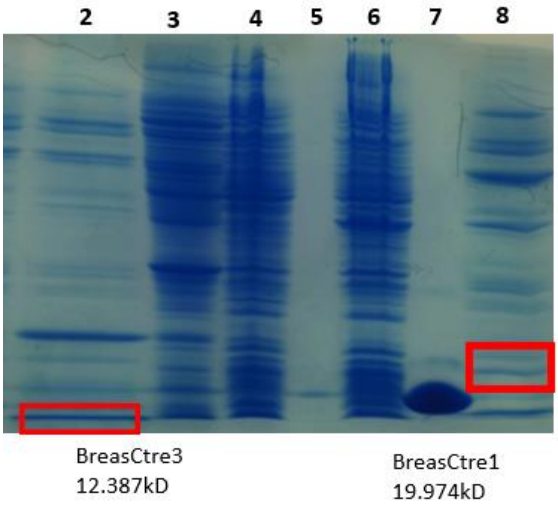



Figure 45: 15 % SDS-PAGE of BrasCret1 (19,974 kDa) and BrasCret3 (12,387 kDa) after purification. Samples: 2) BrasCret3 pure 3) BrasCret3 washing, 4) BrasCret3 flow through 6) BrasCret1 flow through 7) Marker (Lysozyme 14,4 kDa), 8) BrasCret1 pure

## 4.6 Sweet Taste Assay

The end goal of this project was to verify the sweetness of MNEI and Mab-II as well as to identify if our putative proteins were sweet, performing some sensory tests. Consequently, it was unable with this expression system and infrastructure to prepare enough amount of our proteins for taste analysis in a sensory panel, so a preliminary sweetness test with *D. melanogaster* was established . Being small, producing many progenies, growing rapidly, and it perceives the taste very similar to humans are crucial advantages of housing *Drosophila*. First, the adult fly is used for the detection of the sweet taste activity. (Paragraph 4.6.1). As this assay appeared many difficulties, the protocol was changed using the third instar larvae. (Paragraph 4.6.2). This taste preference assay was good at determining how a population of flies responds to different sweet tastants.

### 4.6.1 First Behavioral assay

In our first attempt, adult flies (1-2 days aged) were used in combination with the petri dish 90 mmx 15 mm as an assay chamber. The petri dish was a common product found in laboratories and, therefore, helps facilitate the testing without time or money investment

in new equipment. Thaumatin and sucrose, at 1 mM concentration respectively, applied as tastants and water as the control. After 24 h starvation (Figure 46), 3 adults flies were chosen and simultaneously tracked, using the DanioVision™ Observation Chamber. DanioVision is a video tracking-based system connected to the EthoVision® XT software to digitize the analog signal. Also, this observation chamber preserves the internal environment stable.

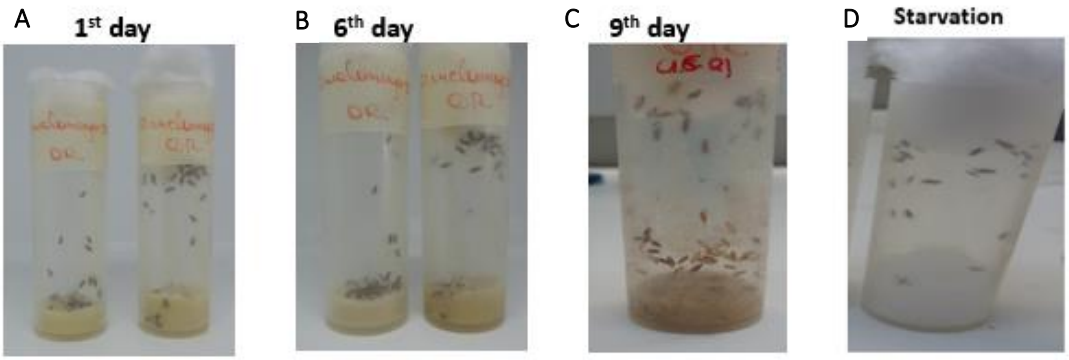



Figure 46: Fly population in different stages. (A) represents the embryonic stage. (B) are shown third inter larva. (C) Pupas start the metamorphosis stage. (D) is illustrated the starvation vials.

The animals were tracked in the closed observation chamber for 8 min, at ambient temperature, and the lighting was set at 20% illumination. However, the results from this assay were abortive. Anesthesia seems to be a critical step for the failure of this assay. The animals showed very stressful behavior, gathering on the lid and stopping flying. One hypothesis was that exposure to ice disrupts physiology and behavior, and thus ample recovery time is required before tracking. The only positive feedback for these results was that the flies sound to recognize Thaumatin as the sweetest, as was the only tastant that flies taste.

### 4.6.2 Second Behavioral assay

Studies were conducted to enhance the sweet taste assay. Third-instar *D. melanogaster* larvae were used throughout the study, and the Petri dishes of 90 mm inner diameter as the assay chamber. The easy handling of larvae allowed to taste 10-15 animals simultaneously, maintaining the rest of the experimental conditions identical (Figure 47).

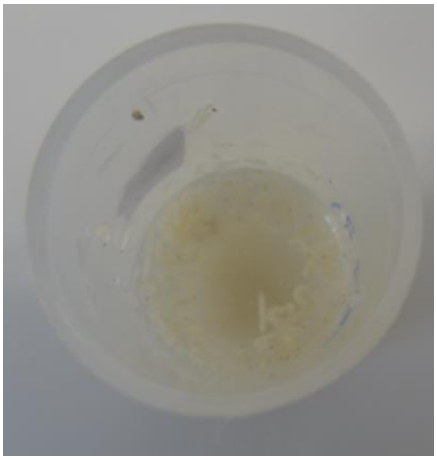


Figure 47: An illustration of third instar larvae.

The first tracking was performed between Thaumatin at 0.23 mM against water. One-half of the Petri dishes containing 100 µL water, whereas the other half contained 100 µL of Thaumatin as illustrated in Figure 48.

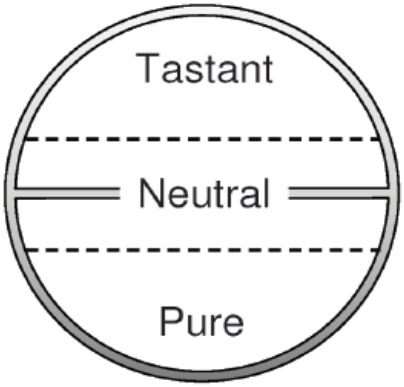



Figure 48: Petri dish in assay. The neutral zone (about 1 cm between both sides).

In order to quantify the results, the gustatory preference index (PREF) was calculated, according to the equation:

$$PREF= ( \#_{tastant} - \#_{pure} )/ \#_{total}$$

#Pure is the number of larvae on the waterside #tastant is the number of larvae on the tastant side and divided this difference by the total number of larvae on the dish (thus excluding the larvae on the lid). The PREF values were restrained between -1, indicating aversion of the tastant, and 1 indicating a preference for the tastant. The bar chart (Figure 49) showed the results from larvae for Thaumatin. According to the grey bars, the animals exhibited an aversion to the sweet protein every minute. These results were in contrast with the results from the literature. Similarly, MNEI and Mab-II at 0.04 mM and 0,03 mM, respectively, were tested. The preference scores for MNEI were negative for the first 4 min and then started increasing until the end. The outcome for Mab-II followed the result

for Thaumatin. Only in one minute, the larvae seem to prefer the protein against the water.

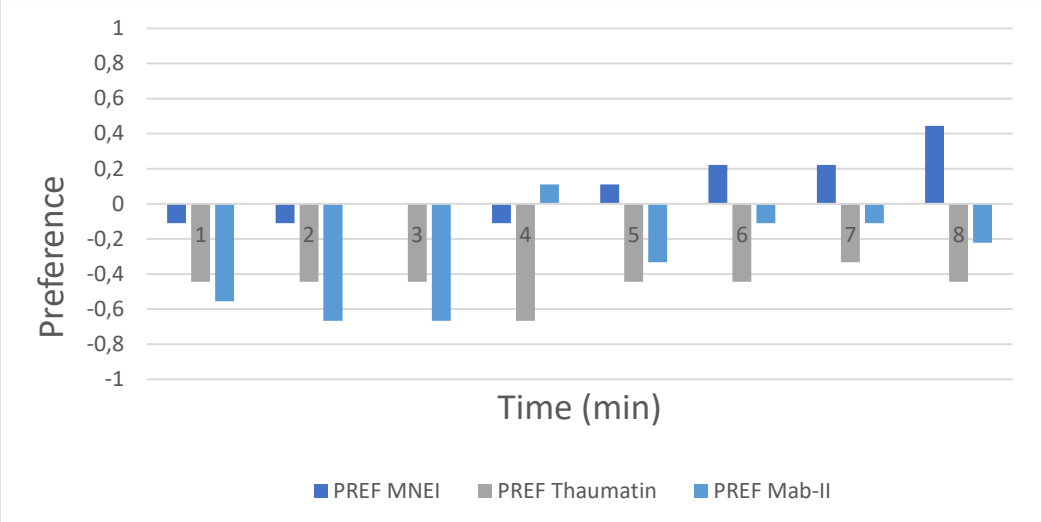



Figure 49: Gustatory choice of third instar larvae for MNEI, Mab-II, and Thaumatin. Gustatory preference indexes were calculated based on the distribution of larvae 1- 8 min after assay onset. Thaumatin is represented with gray, MNEI with dark blue, and Mab-II with light blue.

To that extent, the following experiments was continued analogous to those of König, et al., 2014 with modest variations in tastants and concentrations. The main difference between the new assay and our previous one was the usage of a smaller petri dish, providing the animals less space to react. Therefore, to establish a working concentration for sweet taste assays using DanioVision, larvae were tasted with 1 mM Thaumatin against water (Figure 50) and 1 mM Thaumatin against 1 mM sucrose (Figure 51).

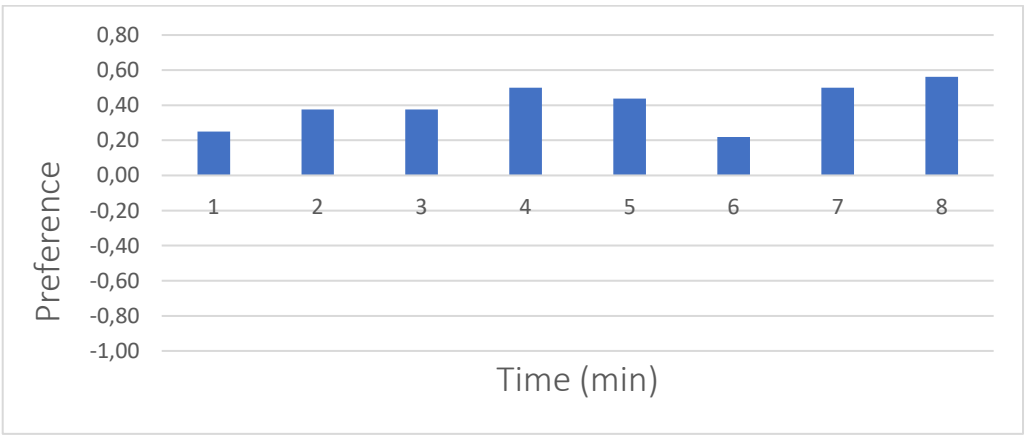



Figure 50 : Gustatory choice of 33 third instar larvae for Thaumatin (1 mM) against water. Gustatory preference indexes were calculated based on the distribution of larvae 1- 8 min after assay onset.

This concentration from the sweet protein caused a strong attraction towards larvae, and the gustatory preference scores were positive for all time points. More accurately, the results from the Thaumatin/sucrose trial highlighted the significant response to sweet protein against the more usual sweetener.

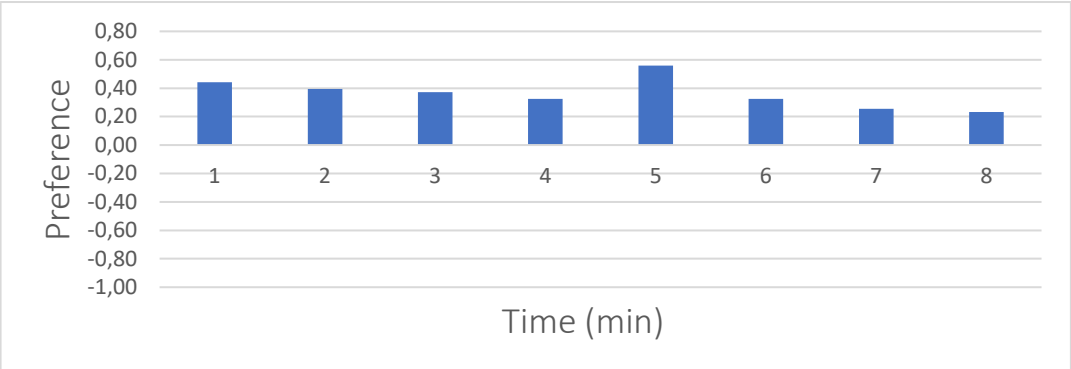



Figure 51 : Gustatory choice of 43 third instar larvae for 1 mM Thaumatin against 1 mM of Sucrose. Gustatory preference indexes were calculated based on the distribution of larvae 1- 8 min after assay onset.

Additionally, MNEI (Figure 52) and Mab-II (Figure 53) tested opposite water. Due to the lack of high concentration for the proteins, 0.11 mM was selected to use in further studies. Consistent with our earlier results from Thaumatin, the animals showed a definite trend in favor of sweet proteins.

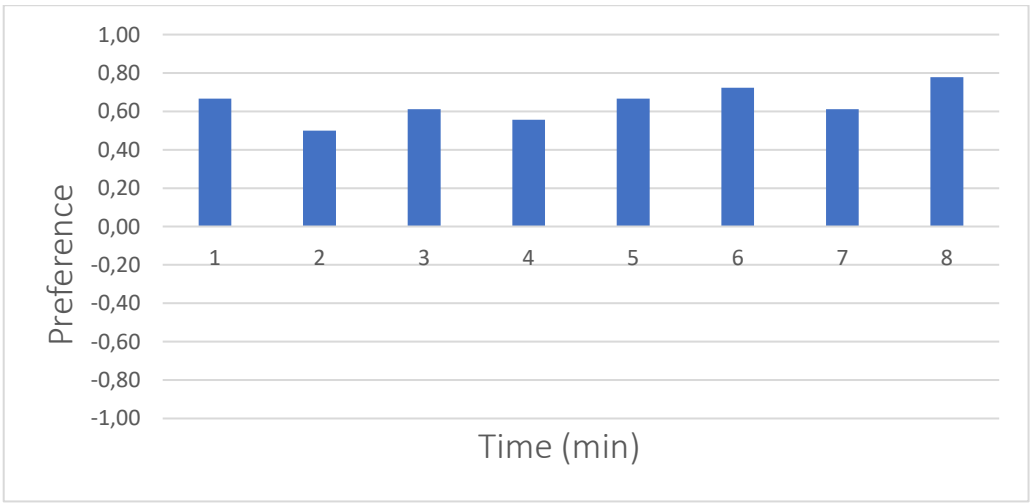



Figure 52: Gustatory choice of 18 third instar larvae for 0.11 mM MNEI against water. Gustatory preference indexes were calculated based on the distribution of larvae 1- 8 min after assay onset

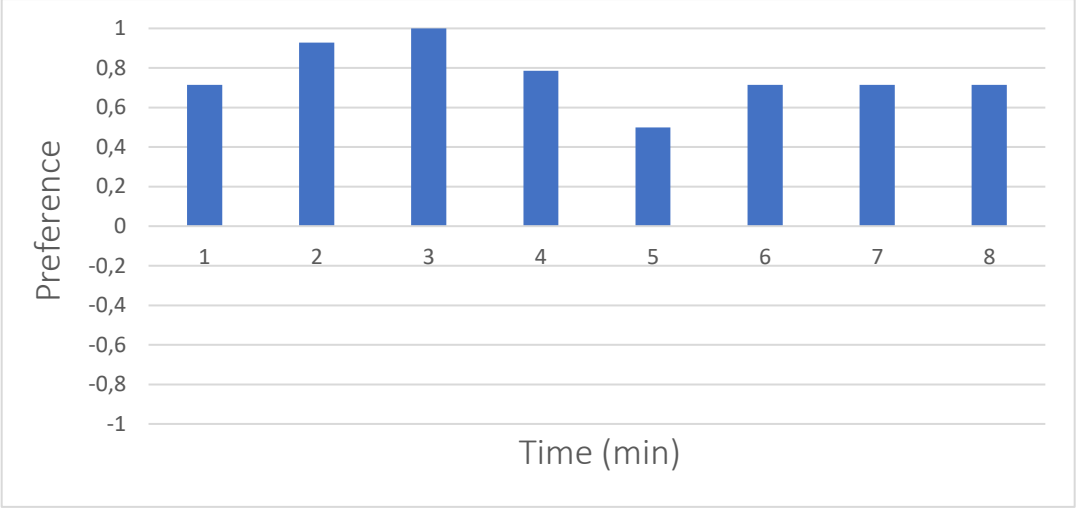



Figure 53: Gustatory choice of 18 third instar larvae for 0.11 mM Mab-II against water. Gustatory preference indexes were calculated based on the distribution of larvae 1- 8 min after assay onset.

However, MNEI and Mab-II were dissolved in sodium phosphate buffer, although, to clarify if the buffer contributed to the larvae's preference, the same experiments was performed between water and sodium phosphate. To directly compared the effects of Kpi, two concentrations were tested, 50 mM and 5 mM. Two different concentrations were tested because the exact concentration of the buffer after the desalting step was unknown. Thus, in the experiment presented in Figure 55, the animals did not respond to Kpi, demonstrating an aversion. Inconsistent with the previous results, larvae, in Figure 54, showed a preference that gradually decreases over time. Together, these data indicated that larvae negatively recognize the sodium phosphate, revealing that the outcomes from MNEI and Mab-II were based on sweet taste.

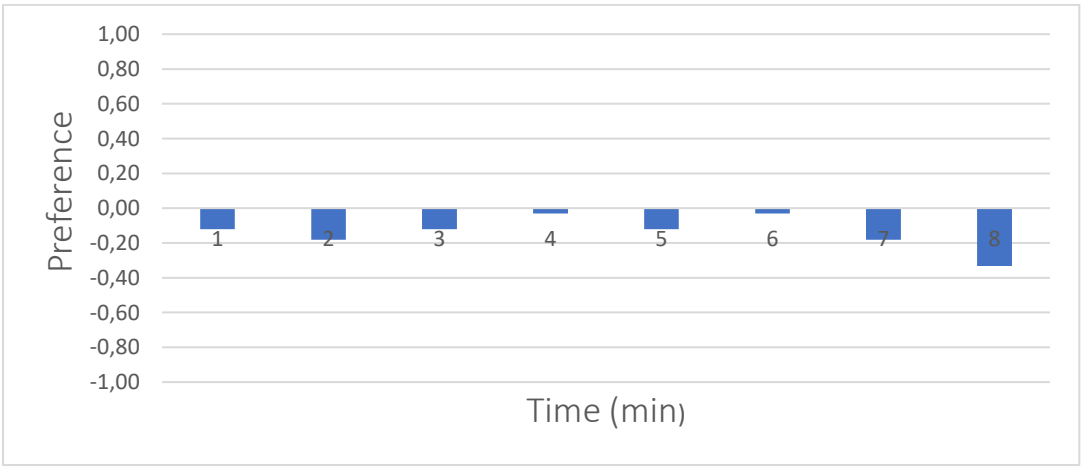



Figure 54: Gustatory choice of 33 third instar larvae for 5 mM sodium phosphate against water. Gustatory preference indexes were calculated based on the distribution of larvae 1-8 min after assay onset.

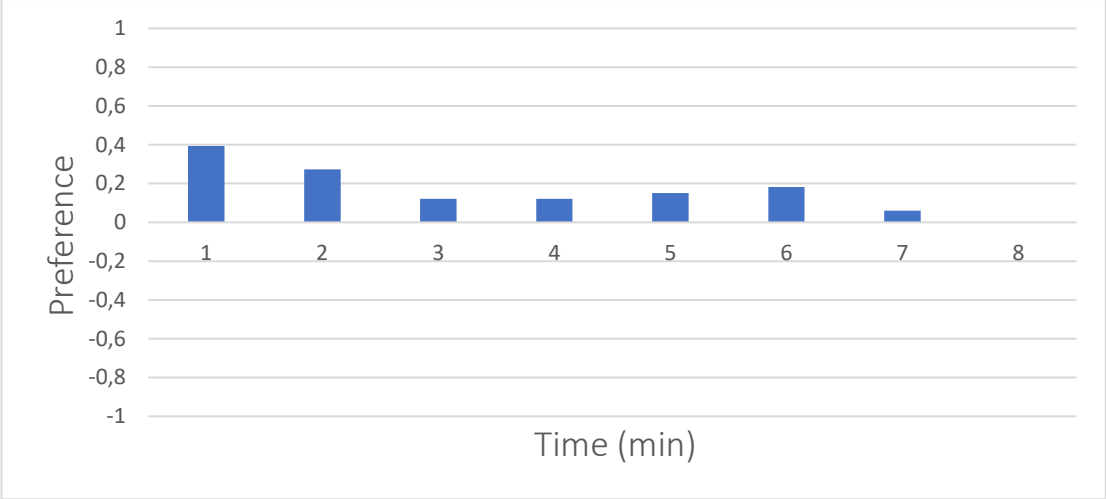



Figure 55: Gustatory choice of 33 third instar larvae for 50 mM sodium phosphate against water. Gustatory preference indexes were calculated based on the distribution of larvae 1- 8 min after assay onset.

# 5. Conclusions and future perspectives

This Master thesis focused on the understanding of the sequence-function relationship of sweet proteins and the identification of novel sweet proteins from the sequence space. The project is divided into three sections: the *in-silico* analysis of sweet proteins and the interaction with the sweet taste receptor, the heterologous expression in *E. coli* of selected putative sweet proteins , and the establishment of a sweet taste assay, using *D. melanogaster* as model organism.

The first scientific objective of this project was the interpretation of the chemical and structural determinants that provide the sweetening effect on the protein. Via the meta-analysis we identified a unique motif of residues with pivotal positions, providing critical interactions with the sweet taste receptor. The interest was focused on the sweet-tasting proteins derived from plants (thaumatin, brazzein, monellin, mabinlin) and on lysozyme.

In summary, according to our results, there are a plethora of critical amino acids vital for the elucidation of sweet sensation. The only common feature among these residues seems to be their charge, confirming the existing literature that the sweet taste relies on electrostatic interactions. However, the sweet proteins deviate in the molecular weight (from 6.4 kDa for brazzein to 24 kDa for neoculin), while the amino acid sequence similarity is low, Indicating a different evolutionary origin. Sweet proteins adopt different three-dimensional structures, and their secondary structure is also entirely different, from only alpha-helical mabinlin to only beta sheets neoculin. Furthermore, the quaternary structure is also quite diverse, observing monomeric, homodimeric, and also heterodimeric states. These structural characteristics could explain the different sweetness profiles, the after taste, as well as the thermal stability between these proteins. One of the common features of sweet-tasting proteins is their alkaline isoelectric point (over 8), except for brazzein, which is acidic.

Taking advantage of the acquired knowledge from the meta-analysis, the research further concentrated on MNEI and Mab-II, as these proteins were reported to be able to be expressed in *E. coli*. Based on their sequences and the meta-analysis, five genes encoding putative sweet proteins were selected from the sequence space. These sequences were ordered as codon-optimized synthetic genes for the expression in *E. coli*.

The interactions between the targeted proteins and the T1R2-T1R3 receptor were studied by docking analysis, using a model from the literature for the receptor, as the protein was not crystallized. The results from docking analysis showed that the proteins interact with the receptor adopting a wedge motif in their structure. Our results establish that Arg177 of T1R3 and Asn152, Glu170, Asp173, and Asp 218 of T1R2 could have a critical role in the interaction, both with the characterized sweet proteins, and also with the putative ones.

Despite trying different strains, medium for expression, chaperons and lysis buffers, the recombinant expression of the selected genes in *E. coli* was proven challenging. Only MNEI was expressed in soluble form, but in low yields. On the other side, Mab-II, ArabisAlp, BrasCret1, and BrasCret3, were expressed mostly in pellet. Cryza.M and BrasCret2 did not

express at all in this expression system. The experiments with Mab-II expression showed that the most crucial parameter is the proper dilution of the cells in the lysis buffer and not the lysis buffer itself. Finally, only two proteins were purified via metal affinity chromatography, but the final yields were very low (2.1 mg/L cultivation for Mab-II and 0.72mg/L for MNEI).

The final goal of this project was to test the sweet taste of the novel proteins. Due to the minimum amount isolated, human tasting panels could not be used. *Drosophila melanogaster*, despite having significant differences in the sweet taste receptor, perceives the taste very similar to humans. In this work, a preliminary sweetness test with *D. melanogaster*'s larvae was established, that allows the calculation of the gustatory preference index. In the established protocol, third instar larvae exhibited a clear preference in 1 mM of thaumatin against sucrose while they detected 0,11 mM of MNEI and Mab-II against water. Despite the fact that these experiments are time-consuming and laborious, as they require growth of the larvae and maintenance of the test animals, and the behavioral tests require a lot of time on the analysis of the videos, we can screen proteins in low amounts, while the assay can also be used to screen multiple samples in parallel, with a medium throughput.

As the results are not conclusive so far, there are several experiments that could be performed in the future. First of all, to test our hypothesis, synthetic genes were ordered for the recombinant expression in *Pichia pastoris*, where the known sweet proteins are expressed in high yields. Genome integration and the expression experiments are already planned in the group. Moreover, the tasting assay could be optimized, with more larvae, better controls and adding in one plate more compounds. This sensory test needs to be compared with tasting panels to prove the correlation and potentially, in a second screening round, trust the *Drosophila* larvae assay and only transfer to the human tasting panels the best candidates. Finally, the structure elucidation of the sweet taste receptors would push forward the whole field, however, this experiment is far from our expertise.

# 6.Appendix

## I Amino acid sequences of the seven tested proteins

**ArabisAlp Wild type** 573 bp DNA with the C-terminal 6xHis-tag

```
5′-ATGACCCAGA  AGCTGTTCCG  TTTTGGTCGT  CGTAGCGACC  AACGTCAGTG  GTGGTACGGC
ACCCGTCGTA  GCAGCAGCAT  CAGCAGCACC  ATTAAGCTGA  AGCCGACCGC  GCTGACCTGC
AGCAGCAGCC  GTACCCGTTT  CCCGTATCGT  AGCGGTCTGG  CGACCCTGCG  TCTGGCGCGT
AAGGAGTTTC  AAGAAAGCCA  GCACCTGCGT  GCGTGCCAGC  AATGGATCCG  TATGCGTGCG
ATGAAACCGC  GTTACGGTCT  GGGCGTGGAC  GATGAGCTGG  ACTTCGAAGA  CGATTTTGAG
AACCCGCAAG  GTCCGCAGCA  ACAGCACCCG  CTGCTGCAAA  AGTGCTGCAG  CGAACTGCAC
CAGGAAGAGC  CGGTGTGCGT  TTGCCCGACC  CTGAAGCAAG  CGGCGAAAGC  GGTGCGTCTG
CAGGGTCAGC  ATGGTCCGTT  CCAGGCGAGC  AAGATTTATC  GTACCGCGAA  AAACCTGCCG
AACGTTTGCA  ACATCCCGCA  GGTGGATGTT  TGCCCGTTTA  AAGCGATTCC  GTTCTTTCCG
CCGTACTATC  TCGAGCACCA  CCACCACCAC  CAC-3′
```

**ArabisAlp Wild type** translation 191 aa with the C-terminal 6xHis-tag

```
MTQKLFRFGR  RSDQRQWWYG  TRRSSSISST  IKLKPTALTC  SSSRTRFPYR  SGLATLRLAR
KEFQESQHLR  ACQQWIRMRA  MKPRYGLGVD  DELDFEDDFE  NPQGPQQQHP  LLQKCCSELH
QEEPVCVCPT  LKQAAKAVRL  QGQHGPFQAS  KIYRTAKNLP  NVCNIPQVDV  CPFKAIPFFP
PYYLEHHHHH  H
```

**BrasCret1 Wild type** 522 bp DNA with the C-terminal 6xHis-tag

```
5′- ATGAAGCGTA  GCGCGAGCGT  GAGCCTGGAC  ACCTGCCGTC  ACGAAATCAA  ACGTCTGATT
TACGTGGCGA  GCTATGTTGG  CGAGGGCGAA  GTTCGTAAGC  AAAAAGAGTT  CCAGCAAGCG
CAGCACCTGC  GTGCGTGCCA  GCAATGGCTG  CACAAGCAAG  CGATGCAGAG  CGGTAGCGGT
CCGAGCTGGA  CCCTGGATGG  CGAGTTCGAT  TTTGAAGACG  ATATGGAGAA  CCCGCAAAGC
CCGCAGCAAC  GTCCGCCGCT  GCTGCAGCAA  TGCTGCAACG  AACTGCACCA  GGAAGAGCCG
CTGTGCGTGT  GCCCGACCCT  GAAAGGTGCG  AGCAAGGCGG  TGAAACAGCA  AGTTCGTCAG
CAACAGGGTC  AACAGGGCCA  ACAGCTGCAA  CAGGTGATCA  GCCGTATTTA  CCAGACCGCG
ACCCACCTGC  CGAAGGTTTG  CAACATCCCG  CAAGTGAGCG  TTTGCCCGTT  TCAGAAAACC
ATGCCGGGTC  CGAGCTATCT  CGAGCACCAC  CACCACCACC  AC-3′
```

**BrasCret1 Wild type** translation 174 aa with the C-terminal 6xHis-tag

```
MKRSASVSLD  TCRHEIKRLI  YVASYVGEGE  VRKQKEFQQA  QHLRACQQWL  HKQAMQSGSG
PSWTLDGEFD  FEDDMENPQS  PQQRPPLLQQ  CCNELHQEEP  LCVCPTLKGA  SKAVKQQVRQ
QQGQQGQQLQ  QVISRIYQTA  THLPKVCNIP  QVSVCPFQKT  MPGPSYLEHH  HHHH
```

**BrasCret2 Wild type** 564 bp DNA with the C-terminal 6xHis-tag

```
5′- ATGGCGAACA  AGCTGTTTCT  GGTGAGCGCG  ACCCTGGCGT  TCTTTTTCCT  GCTGACCAAC
GCGAGCGTGT  ACCGTACCGT  GGTTGAGGTT  GACGAAGACG  ATGCGACCAA  CCCGGCGGGT
CCGTTTCGTA  TCCCGAAGTG  CCGTAAAGAA  TTCCAGCAAG  CGCAGCACCT  GCGTGCGTGC
CAGCAATGGC  TGCACAAGCA  GGCGATGCAA  AGCGGTAGCG  GTCCGAGCCT  GGCGCTGGCG
GGCGAGTTTG  ATTTCGAAGA  CGATATGGAG  AACCCGCAGG  GCAGCCAGCA  ACGTCCGCCG
CTGCTGCAGC  AATGCTGCAA  CGAGCTGCAC  CAGGAAGAGC  CGCTGTGCGT  TTGCCCGACC
CTGAAAGGTG  CGAGCAAGGC  GGTGAAAAGC  CGTGTTCAGC  AACAGGGTCA  GATGCAAGGC
```

CAACAGCAAC AGCAAATCGT GGGCCGTATT TATCAGACCG CGAAGCACCT GCCGCGTGTT
TGCAACATTC CGCAAGTGAG CGTTTGCCCG TTCCGTAAAA CCACCCCGGG CCCGTACTAT
CTCGAGCACC ACCACCACCA CCAC-3′

BrasCret2 Wild type translation 188 aa with the C-terminal 6xHis-tag

MANKLFLVSA TLAFFFLLTN ASVYRTVVEV DEDDATNPAG PFRIPKCRKE FQQAQHLRAC
QQWLHKQAMQ SGSGPSLALA GEFDFEDDME NPQGSQQRPP LLQQCCNELH QEEPLCVCPT
LKGASKAVKS RVQQQGQMQG QQQQQIVGRI YQTAKHLPRV CNIPQVSVCP FRKTTPGPYY
LEHHHHHH

BrasCret3 Wild type 327 bp DNA with the C-terminal 6xHis-tag

5′- ATGGAGAACC CGCAAAGCCC GCAGCAACGT AGCCCGCTGC TGCAGCAATG CTGCAACGAA
CTGCACCAGG AAGAGCCGCT GTGCGTGTGC CCGACCCTGA AGGGTGCGAG CAAGGCGGTG
AAACAGCAAG TTCGTCAGCA ACAGGGTCAA CAGGGCCAAC AGCTGCAACA GGTGATCAGC
CGTATTTACC AAACCGCGAC CCACCTGCCG AAAGTTTGCA ACATCCCGCA AGTGAGCGTT
TGCCCGTTTC AAAAGACCAT GCCGGGTCCG CCGACCCGTT TTCAGACCAA ACCGAGCAGC
GTTCTCGAGC ACCACCACCA CCACCAC-3′

BrasCret3 Wild type translation 109 aa with the C-terminal 6xHis-tag

MENPQSPQQR SPLLQQCCNE LHQEEPLCVC PTLKGASKAV KQQVRQQQGQ QGQQLQQVIS
RIYQTATHLP KVCNIPQVSV CPFQKTMPGP PTRFQTKPSS VLEHHHHHH

Cryza.M Wild type 376 bp DNA with the C-terminal 6xHis-tag

5′- ATGGTGGTGGT GGTGGTGCTCG AGGTGAATCGG GTCAAAGCTGA TCAGTTTACGC GCGTTGCTACG
ATCCTGTTCAT AAACCACCGCC TCGTAACGGCC ATCCTTCGCAC CGCTGCTGCTC GCCTCCAGGGT
CAGACGATAGT TAACACCCGCA ACCACTTGCTT CTCCGCAATGG TCACACGGTTG AAGGTCAGGCC
ACCCAGGCCGG TACGACGGTTC ATTTCCGCAAC CGCCCAACGAC CCAGCTCCTGC ACGTGCGGATC
GTTAATGTCGA TTTGCTGCCAG CTGCCAACGGT CGCCGCGGTCG CACCACGCGCC GCCGCCGCCAG
AACCGCCACGA TGGTCACCGCC GCGAACAGCAG GCTGCTGGTAC GC-3′

Cryza.M Wild type translation 126 aa with the C-terminal 6xHis-tag

MRTSSLLFAA VTIVAVLAAA ARGATAATVG SWQQIDINDP HVQELGRWAV AEMNRRTGLG
GLTFNRVTIA EKQVVAGVNY RLTLEASSSG AKDGRYEAVV YEQDRSNARK LISFDPIHLE
HHHHHH

Mabinlin Wild type 425 bp DNA with the C-terminal 6xHis-tag

5′-ATGAGCATTC AGACCACCGT GATCGAAGTT GACGAGGAAG AGGATAACCA ACTGTGGCGT
TGCCAGCGTC AATTCCTGCA GCACCAACGT CTGCGTGCGT GCCAGCGTTT CATTCACCGT
CGTGCGCAGT TTGGTGGCCA ACCGGACGAA CTGGAGGATG AAGTGGAGGA CGATAACGAC
GATGAGAACC AACCGCGTCG TCCGGCGCTG CGTCAATGCT GCAACCAGCT GCGTCAAGTT
GACCGTCCGT GCGTGTGCCC GGTTCTGCGT CAGGCGGCGC AGCAAGTTCT GCAGCGTCAA
ATCATTCAAG GTCCGCAGCA ACTGCGTCGT CTGTTCGATG CGGCGCGTAA CCTGCCGAAC
ATCTGCAACA TTCCGAACAT CGGCACCTGC CCGTTTCGTA CCTGGCCGCT CGAGCACCAC
CACCACCACC AC-3′

Mabinlin Wild type translation 143 aa with the C-terminal 6xHis-tag

MSIQTTVIEV DEEEDNQLWR CQRQFLQHQR LRACQRFIHR RAQFGGQPDE LEDEVEDDND
DENQPRRPAL RQCCNQLRQV DRPCVCPVLR QAAQQVLQRQ IIQGPQQLRR LFDAARNLPN
ICNIPNIGTC PFRTWPLEHH HHHH

**MNEI Wild type** 315 bp DNA with the C-terminal 6xHis-tag

5′- ATGGGTGAGT GGGAAATCAT TGACATCGGT CCGTTCACCC AGAACCTGGG CAAGTTTGCG
GTGGATGAGG AAAACAAAAT TGGTCAATAC GGCCGTCTGA CCTTCAACAA GGTTATCCGT
CCGTGCATGA AGAAAACCAT TTATGAGAAC GAAGGTTTTC GTGAGATCAA GGGCTACGAA
TATCAGCTGT ACGTGTATGC GAGCGACAAA CTGTTCCGTG CGGACATTAG CGAGGATTAC
AAGACCCGTG GTCGTAAACT GCTGCGTTTT AACGGCCCGG TGCCGCCGCC GCTCGAGCAC
CACCACCACC ACCAC-3′

**MNEI Wild type** translation 105 aa with the C-terminal 6xHis-tag

MGEWEIIDIG PFTQNLGKFA VDEENKIGQY GRLTFNKVIR PCMKKTIYEN EGFREIKGYE
YQLYVYASDK LFRADISEDY KTRGRKLLRF NGPVPPPLEHH HHHH

## II Tables of T1R2-T1R3 interaction

| T1R2 | T1R3 | Type of interaction |
|---|---|---|
| E126 | R137 | intermolecular salt bridges |
| D127 | R123 | intermolecular salt bridges |
| N52 | Q425 and Y128_T1R3 also participates | hydrophilic interactions |
| L51 | L427, L161, P136, and –CH2–of E428 | hydrophobic interaction |
| W543, Y545, and F552 | - | aromatic cluster: |
| | | π–π stacking interactions |
| E542 and either K553 or R554 | - | salt-bridge |
| - | R557 and either o D544 or E545 | polar interactions |
| | W711 interaction with R552 | π-cation interactions |
| C233-C513 | C236-C517 | disulfide bonds |
| C59-C102 | C62-C103 | |
| C495-C514 | C499-C518, | |
| C499-C517 | C503-521, | |
| C520-535, | C524-C538, | |
| C538-C551 | C541-C554 | |

## III Tables of results from docking simulations

Table 31: Results from docking simulations between Cryza.M and the receptor. With light blue are colored critical amino acids for the receptor.

| Critical Bonds | | Other Bond | | Receptor |
|---|---|---|---|---|
| • Glu 96 OE1- Arg 177 HH12 | H-Bond 1.8 A | • Thr 67 HG1-Glu 178 OE3 | H-Bond 2.0 A | |
| • Thr 82 OG1- Arg 177 HH11 | H-Bond 1.7 A | • Arg 94 HH22-Glu 178 OE1 | H-Bond 1.9 A | T1R3 |
| | | • His 117 HD1- Ser 446 OH | H-Bond 1.9 A | |
| | | • Ile 116 O- Ser 446 HG | H-Bond 1.9 A | |
| • Arg 104 HH12- Asp 173 OD2 | H-Bond 1.9 A | • Ser 112 HG -Gln 109 HE1 | H-Bond 1.8 A | |
| • Arg 104 HH22- Asp 173 OD2 | H-Bond 2.0 A | • Asp 114 OD2- Gln 244 HEZ1 | H-Bond 2.1 A | |
| • Gln 102 HE22- Gln170 OD1 | H-Bond 2.3 A | • Asp 114 OD2- Gln 109 HEZ2 | H-Bond 2.3 A | T1R2 |
| • Gln 72 HE22- Asn 152 OD1 | H-Bond 1.9 A | | | |
| • Gly 29 O- Asn 152 HD21 | H-Bond 2.0 A | | | |
| • Lys 109 HZ1 –Asp 218 OD1 | H-Bond 1.7 A | | | |
| • Lys 109 HZ2- Asp 218 O | H-Bond 1.9 A | | | |
| • Lys 109 HZ3- Asp 218 OD2 | H-Bond 1.7 A | | | |

Table 32: Results from docking simulations between ArabisAlp and the receptor. With light blue are colored critical amino acids for the receptor.

| Critical Bonds | | Other Bond | | Receptor |
|---|---|---|---|---|
| | | • Asp 96 OD2 –Lys 155 HZ1 | H-Bond 1.9 A | T1R3 |
| • Gln 166 HE22- Asn 152 OD1 | H-Bond 1.9 A | • Lys 173 HZ1 - Tyr 131 OH | H-Bond 1.7 A | |
| • Arg 49 HH21 - Asp 456 OD2 | H-Bond 1.9 A | • Lys 131 HZ1- Leu156 O | H-Bond 1.6 A | |
| • Arg 49 HH21 - Asp 456 OD1 | H-Bond 2.1 A | • TYR 181/O-Arg 413 NH2 | H-Bond 1.7 A | |
| • Gly 51 H- Asp 456 OD1 | H-Bond 2.1 A | • Pro 180 O- Tyr 128 HH | H-Bond 1.8 A | |
| • Ser 50 H- Asp 456 OD1 | H-Bond 2.0 A | • Arg 59 HH12- Glu 422 OE2 | H-Bond 1.8 A | |
| • Gln 166 HE22-Arg 176 HH11 | H-Bond 1.9 A | • Arg 59 HH22 - Glu 422 OE1 | H-Bond 1.8 A | T1R2 |
| • Thr 54 HG1 -Asp173 OD1 | H-Bond 2.0 A | • Arg 22 HH22- Glu 422 O | H-Bond 1.8 A | |
| • Leu 52 O-Arg 172 HH21 | H-Bond 2.1 A | • Arg 22 HH21- Glu 422 O | H-Bond 1.8 A | |
| • Leu 52 O-Arg 172 HE | H-Bond 1.9 A | • Gln 166 HE21-Glu 145 OE1 | H-Bond 2.1 A | |
| | | • Phe 46H- Glu 449 OE2 | H-Bond 2.3 A | |
| | | • Phe 46H- Glu 449 OE1 | H-Bond 2.1 A | |
| | | • Arg 49 HE-Glu 449 OE1 | H-Bond 2.0 A | |
| | | • Arg 49 HH11-Glu 449 OE1 | H-Bond 1.9 A | |
| | | • Arg 49 HH12- Trp 453 O | H-Bond 2.0 A | |
| | | • Arg 49 HH12- Gln 454 O | H-Bond 2.0 A | |
| | | • Ser 50 H – Asp 456 OD2 | H-Bond 2.0 A | |
| | | • Gly 51 O- Asp 456 OD2 | H-Bond 2.1 A | |

Table 33: Results from docking simulations between BrasCret3 and the receptor. With light blue are colored critical amino acids for the receptor.

| Critical Bonds | | Other Bond | | Receptor |
|---|---|---|---|---|
| | | • Leu 26 O – Arg 443 HH21<br>• Ser 78 HG- Glu 172 OE2 | H-Bond 1.8 A<br>H-Bond 1.9 A | T1R3 |
| • Arg 44 HH22- Asp 218 OD1<br>• Arg 44 HH12 – Asp 218 OD2<br>• Gln 52 HE22- Asp 218 OD2<br>• Arg 92 HHZ1 – Asp 173 OD2<br>• Arg 92 HHZ2 – Asp 173 OD1<br>• Pro 89 O- Arg 176 HH11<br>• Phe 93 H- Glu 170 OE2<br>• Arg 92 H- Glu 170 OE2 | H-Bond 1.8 A<br>H-Bond 1.8 A<br>H-Bond 2.0 A<br>H-Bond 1.8 A<br>H-Bond 2.1 A<br>H-Bond 2.0 A<br>H-Bond 2.1 A<br>H-Bond 2.1 A | • Thr 85 HG1- Gln 109 OE1<br>• Gln 50 HE21- Gln 237 OE1<br>• Gln 50 OE1- Gln 237 HEZ2<br>• Leu 26 O – Arg 443 HH21<br>• Gln 83 HEZ1- Glu 178 OE2<br>• Gln 83 HEZ2- Glu 178 OE1<br>• Lys 37 HZ1- Asn 52 OD1<br>• Lys 37 HZ2- Asn 52 O<br>• Lys 37 HZ30 Gln 241 OE1 | H-Bond 1.8 A<br>H-Bond 1.9 A<br>H-Bond 2.5 A<br>H-Bond 1.8 A<br>H-Bond 2.2 A<br>H-Bond 2.5 A<br>H-Bond 1.7 A<br>H-Bond 1.8 A<br>H-Bond 1.7 A | T1R2 |

Table 34: Results from docking simulations between BrasCret1 and the receptor. With light blue are colored critical amino acids for the receptor, and with green critical amino acids for the protein.

| Critical Bonds | | Other Bond | | Receptor |
|---|---|---|---|---|
| • Gln 157 O - Arg177 HH12<br>• Gln 157 O - Arg177 HH22<br>• Lys 158 O- Arg 177 HH22<br>• Tyr 165 O- Arg 191 HH21<br>• Tyr 165 O- Arg 191 HE | H-Bond 2.1 A<br>H-Bond 2.0 A<br>H-Bond 1.8 A<br>H-Bond 1.8 A<br>H-Bond 2.0 A | • Lys 144 HZ2- Asp 419 OD2<br>• Val 151 O- Lys 422 HZ1<br>• Gln 150 O- Lys 422 HZ2<br>• Tyr 20 OH- Glu 148 OE2<br>• Lys 16 HZ1- Asp 25 OD2<br>• His 141 NE1- Thr 149 HG1 | H-Bond 1.9 A<br>H-Bond 1.7 A<br>H-Bond 1.9 A<br>H-Bond 1.9 A<br>H-Bond 1.8 A<br>H-Bond 2.9 A | T1R3 |
| • Gln 78 HEZ2- Asp 173 OD1<br>• Gln 55 HEZ1 Asp 173 OD2<br>• Arg 134 HE- Asp 218 PD2<br>• Gln 52 HEZ1- Glu 170 OE1<br>• Arg134 HH12- Asp 218 O<br>• Arg 134 HH11- Asp 218 OD2 | H-Bond 2.3 A<br>H-Bond 2.1 A<br>H-Bond 2.0 A<br>H-Bond 2.0 A<br>H-Bond 2.4 A<br>H-Bond 2.0 A | • Glu 67 OE2 – Arg 229 HH11<br>• Glu 67 OE1- Arg 229 HH12<br>• Ser 61 O- Arg 226 HHZ1<br>• Ser 58 OG- Arg 226 HHZ<br>• Gly 14 OE2- Arg 247 HHZ<br>• Thr 63 OG1- Arg 229 HHZ2<br>• Thr 63 HG1- Glu 225 OE2<br>• Gly 57 O- Arg 226 HH12<br>• Gly 57 O- Arg 226 HH22 | H-Bond 1.8 A<br>H-Bond 2.2 A<br>H-Bond 2.0 A<br>H-Bond 2.0 A<br>H-Bond 1.9 A<br>H-Bond 1.7 A<br>H-Bond 1.9 A<br>H-Bond 1.8 A<br>H-Bond 2.0 A | T1R2 |

Table 35: Results from docking simulations between BrasCret2 and the receptor. With light blue are colored critical amino acids for the receptor.

| Critical Bonds | | Other Bond | | Receptor |
|---|---|---|---|---|
| • Gln 138 O - Arg177 HH12 | H-Bond 1.9 A | | | |
| • Gln 138 O - Arg177 HH22 | H-Bond 2.1 A | | | T1R3 |
| • Arg 130 HH2- Asn 152 OD1 | H-Bond 1.9 A | • Arg 130 HH12- Ser 155 HG | H-Bond 1.9 A | |
| • Arg 130 HH2- Asn 152 O | H-Bond 1.9 A | • Arg 130 HH22- Ser 155 HG | H-Bond 2.7 A | |
| • Arg 130 HH11- Arg 176 O | H-Bond 1.9 A | • Gln 134 HE22- Pro 110 O | | |
| • His 55 HD1- Lys 174 O | H-Bond 2.3 A | • Gln 61 HE22- Leu 156 O | | |
| • Tyr 178 O- Arg 172 HHZ1 | H-Bond 1.9 A | • Gln 99 OE1- Trp 418 H | | |
| • Tyr 178 O- Arg 172 HE | H-Bond 2.8 A | • Gln 99 OE1- Gln 419 H | | T1R2 |
| • Tyr 178 O- Asp 173 OD1 | H-Bond 1.9 A | • Gln 61 HE22- LEU 156 O | | |
| • Lys 125 HZ3- Asp 173 OD2 | H-Bond 1.8 A | • Glu 89 OE1- Arg 413 HE | | |
| • Lys 125 HZ1- Glu 170 OE2 | H-Bond 1.9 | • Glu 89 OE2 – Arg 413 HHZ1 | | |
| | | • Glu 89 OE2 Arg 413 HH12 | | |
| | | • Glu 85 OE1- Tyr 128 H | | |

# 7. REFERENCES


Acevedo, W., Ramírez-Sarmiento, C. A., & Agosin, E. (2018). Identifying the interactions between natural, non-caloric sweeteners and the human sweet receptor by molecular docking. *Food Chemistry*, *264*(December 2017),164–171. https://doi.org/10.1016/j.foodchem.2018.04.113

Anja Heilmann, Carolina Machuca Vargas, and Richard G. Watt (2013). *Sugar Consumption and Oral Health, Textbooks in Contemporary Dentistry: Oral Epidemiology*. Retrieved from http://www.https//link.springer.com/chapter/10.1007/978-3-030-50123-5_12

Apostolopoulou, A. A., Rist, A., & Thum, A. S. (2015). Taste processing in Drosophila larvae. *Frontiers in Integrative Neuroscience*, *9*(OCT), 1–9. https://doi.org/10.3389/fnint.2015.00050

Assadi-Porter, F. M., Abildgaard, F., Blad, H., & Markley, J. L. (2003). Correlation of the sweetness of variants of the protein brazzein with patterns of hydrogen bonds detected by NMR spectroscopy. *Journal of Biological Chemistry*, *278*(33), 31331–31339. https://doi.org/10.1074/jbc.M302663200

Assadi-Porter, F. M., Abildgaard, F., Blad, H., Cornilescu, C. C., & Markley, J. L. (2005). Brazzein, a small, sweet protein: Effects of mutations on its structure, dynamics and functional properties. *Chemical Senses*, *30 SUPPL.*(suppl 1), 90–91. https://doi.org/10.1093/chemse/bjh128

Assadi-Porter, F. M., Maillet, E. L., Radek, J. T., Quijada, J., Markley, J. L., & Max, M. (2010). Key Amino Acid Residues Involved in Multi-Point Binding Interactions between Brazzein, a Sweet Protein, and the T1R2-T1R3 Human Sweet Receptor. *Journal of Molecular Biology*, *398*(4), 584–599. https://doi.org/10.1016/j.jmb.2010.03.017

Bantel, A. P., & Tessier, C. R. (2016). Taste preference assay for adult Drosophila. *Journal of Visualized Experiments*, *2016*(115), 1–6. https://doi.org/10.3791/54403

Belenioti, M., & Chaniotakis, N. (2020). Aggressive Behaviour of Drosophila suzukii in Relation to Environmental and Social Factors. *Scientific Reports*, *10*(1), 1–10. https://doi.org/10.1038/s41598-020-64941-1

Belloir, C., Neiers, F., & Briand, L. (2017). Sweeteners and sweetness enhancers. *Current Opinion in Clinical Nutrition and Metabolic Care*, *20*(4), 279–285. https://doi.org/10.1097/MCO.0000000000000377

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–242. https://doi.org/10.1093/nar/28.1.235

Bornscheuer, U. T., & Höhne, M. (2018). Protein engineering - Methods and protocols. *Methods in Molecular Biology*, *1685*, 1–347. https://doi.org/10.1007/978-1-4939-7366-8

Brouwer, J. N., Van Der Wel, H., Francke, A., & Henning, G. J. (1968). Protein from Miracle Fruit. *Nature*, *220*, 373–374.

Cai, C., Li, L., Lu, N., Zheng, W., Yang, L., & Liu, B. (2016). Expression of a high sweetness and heat-resistant mutant of sweet-tasting protein, monellin, in Pichia pastoris with a constitutive GAPDH promoter and modified N-terminus. *Biotechnology Letters*, *38*(11), 1941–1946. https://doi.org/10.1007/s10529-016-2182-4

Chen, V. B., Arendall 3rd, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., & Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular

crystallography. *Acta Crystallographica. Section D, Biological Crystallography*, *66*(Pt 1), 12–21. https://doi.org/10.1107/S0907444909042073

Cheron JB, Golebiowski J, Antonczak S & Fiorucci S (2017) The anatomy of mammalian sweet taste receptors. Proteins 85, 332–341. https://doi.org/10.1002/prot.25228

Dahanukar A, Lei YT, Kwon JY, Carlson JR. Two gr genes underlie sugar reception in *Drosophila*. Neuron. 2007;56:503-516 DOI: 10.1016/j.neuron.2007.10.024

Das, A., & Chakraborty, R. (2016). *An Introduction to Sweeteners*. 1–13. https://doi.org/10.1007/978-3-319-26478-3_1-1

Depoortere, I. Taste receptors of the gut: Emerging roles in health and disease. Gut 2014, 63, 179–190https://doi.org/10.1136/gutjnl-2013-305112

DuBois, G. E. (2008). Sweeteners and sweetness modulators: Requirements for commercial viability. *ACS Symposium Series*, *979*, 444–462. https://doi.org/10.1021/bk-2008-0979.ch029

Esposito, V., Gallucci, R., Picone, D., Saviano, G., Tancredi, T., & Temussi, P. A. (2006). The Importance of Electrostatic Potential in The Interaction of Sweet Proteins with the Sweet Taste Receptor. *Journal of Molecular Biology*, *360*(2), 448–456.

Ferland A, Brassard P, Poirier P. Is aspartame really safer in reducing the risk of hypoglycemia during exercise in patients with type 2 diabetes? *Diabetes Care. 2007*;30:e59. https://doi.org/10.2337/dc06-1888

Fry, J. C. (2012). Natural low-calorie sweeteners. In *Natural Food Additives, Ingredients and Flavourings*. https://doi.org/10.1533/9780857095725.1.41

G. Hellekant, D. Glaser, J.N. Brouwer, H. van der Wel, Gustatory effects of miraculin, monellin and thaumatin in the saguinus midas tamarind monkey studied with electrophysiological and behavioural techniques. *Acta Physiol. Scand. 97 (1976*) 97–105. https://doi.org/10.1111/j.1748-1716.1976.tb10257.x

G. Morini, A. Bassoli, P.A. Temussi, From small sweeteners to sweet proteins: anatomy of the binding sites of the human T1R2–T1R3 receptor. *Journal of Medicinal Chemistry 48* (2005) 5520–5529. https://doi.org/10.1021/jm0503345

Gao, G. H., Dai, J. X., Ding, M., Hellekant, G., Wang, J. F., & Wang, D. C. (1999). Solution conformation of brazzein by 1H nuclear magnetic resonance: Resonance assignment and secondary structure. *International Journal of Biological Macromolecules*, *24*(4), 351–359. https://doi.org/10.1016/S0141-8130(99)00055-0

H. van der Wel, K. Arvidson, Qualitative psychophysical studies on the gustatory effects of the sweet tasting proteins thaumatin and monellin. *Chem. Senses 3 (1978)* 291–299. https://doi.org/10.1038/nbt0592-561

Haque, M., McKimm, J., Sartelli, M., Samad, N., Haque, S. Z., & Bakar, M. A. (2020). A narrative review of the effects of sugar-sweetened beverages on human health: A key global health issue. *Journal of Population Therapeutics and Clinical Pharmacology*, *27*(1), e76–e103. https://doi.org/10.15586/jptcp.v27i1.666

Izawa, K., Amino, Y., Kohmura, M., Ueda, Y., & Kuroda, M. (2010). *4.16 Human–Environment Interactions – Taste*.

J.N. Brouwer, G. Hellekant, Y. Kasahara, H. van der Wel, Y. Zotterman, Electrophysiological study of the gustatory effects of the sweet proteins monellin and thaumatin in monkey, guinea pig and rat. *Acta Physiol. Scand. 89 (1973)* 550–557. https://doi.org/10.1111/j.1748-1716.1973.tb05549.x

Jiang, P., Cui, M., Zhao, B., Snyder, L. A., Benard, L. M. J., Osman, R., … Margolskee, R. F. (2005). Identification of the cyclamate interaction site within the transmembrane domain of the human sweet taste receptor subunit T1R3. *Journal of Biological Chemistry*, *280*(40), 34296–34305. https://doi.org/10.1074/jbc.M505255200

Jiao Y, Moon SJ, Montell C. A *Drosophila* gustatory receptor required for the responses to sucrose, glucose, and maltose identified by mRNA tagging. *Proceedings of the National Academy of Sciences of the United States of America. 2007*;104:14110-14115 https://doi.org/10.1073/pnas.0702421104

Jin, Z., Danilova, V., Assadi-Porter, F. M., Aceti, D. J., Markley, J. L., & Hellekant, G. öran. (2003). Critical regions for the sweetness of brazzein. *FEBS Letters*, *544*(1–3), 33–37. https://doi.org/10.1016/S0014-5793(03)00383-1

Joseph, J. A., Akkermans, S., Nimmegeers, P., & Van Impe, J. F. M. (2019). Bioproduction of the recombinant SWEET protein thaumatin: Current state of the art and perspectives. *Frontiers in Microbiology*, *10*(APR), 1–19. https://doi.org/10.3389/fmicb.2019.00695

Kaneko, R. (2001). Structure-Sweetness Relationship in Thaumatin: Importance of Lysine Residues. *Chemical Senses*, *26*(2), 167–177. https://doi.org/10.1093/chemse/26.2.167

Kant, R. (2005). Sweet proteins - Potential replacement for artificial low calorie sweeteners. *Nutrition Journal*, *4*(February 2005). https://doi.org/10.1186/1475-2891-4-5

Kashani-Amin, E., Sakhteman, A., Larijani, B., & Ebrahim-Habibi, A. (2019). Introducing a New Model of Sweet Taste Receptor, a Class C G-protein Coupled Receptor (C GPCR). *Cell Biochemistry and Biophysics*, *77*(3), 227–243. https://doi.org/10.1007/s12013-019-00872-7

Kaushik, S., & Kain, P. (2020). Understanding Taste Using Drosophila melanogaster . *Animal Models in Medicine and Biology*. https://www.intechopen.com/chapters/69573

Kinghorn, A. D., Chin, Y. W., Pan, L., & Jia, Z. (2010). Natural products as sweeteners and sweetness modifiers. *Comprehensive Natural Products II: Chemistry and Biology*, *3*, 269–315. https://doi.org/10.1016/b978-008045382-8.00077-0

Koizumi A, Nakajima K-I, Asakura T, Morita Y, Ito K, Shmizu-Ibuka A, Misaka T & Abe K (2007) Taste modifying sweet protein, neoculin, is received at human T1R3 amino terminal domain. *Biochem Biophys Res Commun* 358, 585–589. https://doi.org/10.1371/journal.pone.0019448

König, C., Schleyer, M., Leibiger, J., El-Keredy, A., & Gerber, B. (2014). Bitter-sweet processing in larval Drosophila. *Chemical Senses*, *39*(6), 489–505. https://doi.org/10.1093/chemse/bju016

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., … Vajda, S. (2017). The ClusPro web server for protein-protein docking. *Nature Protocols*, *12*(2), 255–278. https://doi.org/10.1038/nprot.2016.169

Kurimoto, E., Suzuki, M., Amemiya, E., Yamaguchi, Y., Nirasawa, S., Shimba, N., … Kato, K. (2007). Curculin exhibits sweet-tasting and taste-modifying activities through its distinct molecular surfaces. *Journal of Biological Chemistry*, *282*(46), 33252–33256. https://doi.org/10.1074/jbc.C700174200

Lee, A. A., & Owyang, C. (2017). Sugars, sweet taste receptors, and brain responses. *Nutrients*, *9*(7), 1–13. https://doi.org/10.3390/nu9070653

Lee, J. W., Cha, J. E., Jo, H. J., & Kong, K. H. (2013). Multiple mutations of the critical amino acid residues for the sweetness of the sweet-tasting protein, brazzein. *Food Chemistry*, *138*(2–3), 1370–1373. https://doi.org/10.1016/j.foodchem.2012.10.140

Lee, S. Y., Lee, J. H., Chang, H. J., Cho, J. M., Jung, J. W., & Lee, W. (1999). Solution structure of a sweet protein single-chain monellin determined by nuclear magnetic resonance and dynamical simulated annealing calculations. *Biochemistry*, *38*(8), 2340–2346. https://doi.org/10.1021/bi9822731

Leone, S., Pica, A., Merlino, A., Sannino, F., Temussi, P. A., & Picone, D. (2016). Sweeter and stronger: Enhancing sweetness and stability of the single chain monellin MNEI through molecular design. *Scientific Reports*, *6*(July), 1–10. https://doi.org/10.1038/srep34045

Li, D. F., Jiang, P., Zhu, D. Y., Hu, Y., Max, M., & Wang, D. C. (2008). Crystal structure of Mabinlin II: A novel structural type of sweet proteins and the main structural basis for its sweetness. *Journal of Structural Biology*, *162*(1), 50–62. https://doi.org/10.1016/j.jsb.2007.12.007

Liu, B., Ha, M., Meng, X.-Y., Kaur, T., Khaleduzzaman, M., Zhang, Z., et al. (2011). Molecular mechanism of species-dependent sweet taste toward artificial sweeteners. *Journal of Neuroscience, 31(30*), 11070–11076. https://doi.org/10.1523/jneurosci.0791-11.2011.

Maehashi, K., & Udaka, S. (1998). Sweetness of lysozymes. *Bioscience, Biotechnology and Biochemistry*, Vol. 62, pp. 605–606. https://doi.org/10.1271/bbb.62.605

Malik, V. S., Popkin, B. M., Bray, G. A., Després, J. P., Willett, W. C., & Hu, F. B. (2010). Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes: A meta-analysis. *Diabetes Care*, *33*(11), 2477–2483. https://doi.org/10.2337/dc10-1079

Masuda K, Koizumi A, Nakajima K, Tanaka T, Abe K, Misaka T & Ishiguro M (2012) Characterization of the modes of binding between human sweet taste receptor and low-molecular-weight sweet compounds https://doi.org/10.1371/journal.pone.0035380

Masuda T et al (2013) Five amino acid residues in cysteine-rich domain of human T1R3 were involved in the response for sweet-tasting protein, thaumatin. *Biochimie* 95:1502–1505 https://doi.org/10.1016/j.biochi.2013.01.010

Masuda, T., & Kitabatake, N. (2006). Developments in biotechnological production of sweet proteins. *Journal of Bioscience and Bioengineering*, *102*(5), 375–389. https://doi.org/10.1263/jbb.102.375

Masuda, T., Ide, N., & Kitabatake, N. (2005A). Structure-sweetness relationship in egg white lysozyme: Role of lysine and arginine residues on the elicitation of lysozyme sweetness. *Chemical Senses*, *30*(8), 667–681. https://doi.org/10.1093/chemse/bji060

Masuda, T., Ide, N., & Kitabatake, N. (2005B). Effects of chemical modification of lysine residues on the sweetness of lysozyme. *Chemical Senses*, *30*(3), 253–264. https://doi.org/10.1093/chemse/bji021

Masuda, T., Kigo, S., Mitsumoto, M., Ohta, K., Suzuki, M., Mikami, B., … Tani, F. (2018). Positive charges on the surface of thaumatin are crucial for the multi-point interaction with the sweet receptor. *Frontiers in Molecular Biosciences*, *5*(FEB), 1–11. https://doi.org/10.3389/fmolb.2018.00010

Masuda, T., Mikami, B., & Tani, F. (2014). Atomic structure of recombinant thaumatin II reveals flexible conformations in two residues critical for sweetness and three consecutive glycine residues. *Biochimie*, *106*, 33–38. https://doi.org/10.1016/j.biochi.2014.07.016

Masuda, T., Ohta, K., Mikami, B., & Kitabatake, N. (2011). A,High-resolution structure of the recombinant sweet-tasting protein thaumatin i. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, *67*(6), 652–658. https://doi.org/10.1107/S174430911101373X

Masuda, T., Ohta, K., Ojiro, N., Murata, K., Mikami, B., Tani, F., … Kitabatake, N. (2016). A Hypersweet Protein: Removal of the Specific Negative Charge at Asp21 Enhances Thaumatin Sweetness. *Scientific Reports*, *6*(October 2015), 1–9. https://doi.org/10.1038/srep20255

Masuda, T., Ohta, K., Tani, F., Mikami, B., & Kitabatake, N. (2011). B.Crystal structure of the sweet-tasting protein thaumatin II at 1.27Å. *Biochemical and Biophysical Research Communications*, *410*(3), 457–460. https://doi.org/10.1016/j.bbrc.2011.05.158

Masuda, T., Okubo, K., Murata, K., Mikami, B., Sugahara, M., Suzuki, M., … Tani, F. (2019). Subatomic structure of hyper-sweet thaumatin D21N mutant reveals the importance of flexible conformations for enhanced sweetness. *Biochimie*, *157*, 57–63. https://doi.org/10.1016/j.biochi.2018.10.020

Masuda, T., Ueno, Y., & Kitabatake, N. (2001). Sweetness and enzymatic activity of lysozyme. *Journal of Agricultural and Food Chemistry*, *49*(10), 4937–4941. https://doi.org/10.1021/jf010404q

Matano, M., Nakajima, K., Kashiwagi, Y., Udaka, S., & Maehashi, K. (2015). Sweetness characterization of recombinant human lysozyme. *Comparative Biochemistry and Physiology Part - B: Biochemistry and Molecular Biology*, *188*, 8–14. https://doi.org/10.1016/j.cbpb.2015.05.009

Mortensen, A. (2006). Sweeteners permitted in the European Union: Safety aspects. *Scandinavian Journal of Food and Nutrition*, *50*(3), 104–116. https://doi.org/10.1080/17482970600982719

Nichols, C. D., Becnel, J., & Pandey, U. B. (2012). Methods to Assay Drosophila Behavior. *Journal of Visualized Experiments*, (61), 3–7. https://doi.org/10.3791/3795

Nookaraju, A., Upadhyaya, C. P., Pandey, S. K., Young, K. E., Hong, S. J., Park, S. K., & Park, S. W. (2010). Molecular approaches for enhancing sweetness in fruits and vegetables. *Scientia Horticulturae*, *127*(1), 1–15. https://doi.org/10.1016/j.scienta.2010.09.014

Ogata, C. M., Gordon, P. F., de Vos, A. M., & Kim, S. H. (1992). Crystal structure of a sweet tasting protein thaumatin I, at 1·65 Å resolution. *Journal of Molecular Biology, 228(3),* 893–908. https://doi.org/10.1016/0022-2836(92)90873-I

Ohta, K., Masuda, T., Ide, N., & Kitabatake, N. (2008). Critical molecular regions for elicitation of the sweetness of the sweet-tasting protein, thaumatin I. *FEBS Journal*, *275*(14), 3644–3652. https://doi.org/10.1111/j.1742-4658.2008.06509.x

Ohta, K., Masuda, T., Tani, F., & Kitabatake, N. (2011). Introduction of a negative charge at Arg82 in thaumatin abolished responses to human T1R2-T1R3 sweet receptors. *Biochemical and Biophysical Research Communications*, *413*(1), 41–45. https://doi.org/10.1016/j.bbrc.2011.08.033

Pearlman, M., Obert, J., and Casey, L. (2017). The association between artificial sweeteners and obesity. *Curr. Gastroenterol. Rep. 19:6*4. http://doi.org/10.1007/s11894-017-0602-9

Perez-Aguilar, J. M., Kang, S. G., Zhang, L., & Zhou, R. (2019). Modeling and Structural Characterization of the Sweet Taste Receptor Heterodimer. *ACS Chemical Neuroscience*, *10*(11), 4579–4592. https://doi.org/10.1021/acschemneuro.9b00438

Picone, D., & Temussi, P. A. (2012). Dissimilar sweet proteins from plants: Oddities or normal components? *Plant Science*, *195*, 135–142. https://doi.org/10.1016/j.plantsci.2012.07.001

Reed DR, Li S, Li X, Huang L, Tordoff MG, Starling-Rony R, Taniguchi K, West DB, Beauchamp GK, Bachmanov AA. 2004. Polymorphisms in the taste receptor gene (Tas1r3) region are associated with saccharin preference in 30 mouse strains. *J Neurosci*. 24:938–946. https://doi.org/10.1093/chemse/bjm088

Rega, M. F., Siciliano, A., Gesuele, R., Lofrano, G., Carpentieri, A., Picone, D., & Guida, M. (2017). Ecotoxicological survey of MNEI and Y65R-MNEI proteins as new potential high-intensity sweeteners. *Environmental Science and Pollution Research*, *24*(10), 9734–9740. https://doi.org/10.1007/s11356-017-8626-0

Sclafani A, Clare RA. 2004. Female rats show a bimodal preference response to the artificial sweetener sucralose. *Chem Senses*. 29:523–528. https://doi.org/10.1093/chemse/bjh055

Shimizu-Ibuka, A., Morita, Y., Terada, T., Asakura, T., Nakajima, K. ichiro, Iwata, S., … Abe, K. (2006). Crystal Structure of Neoculin: Insights into its Sweetness and Taste-modifying Activity.

Somoza, J. R., Cho, J. M., & Kim, S. H. (1995). The taste-active regions of monellin, a potently sweet protein. *Chemical Senses*, *20*(1), 61–68. https://doi.org/10.1093/chemse/20.1.61

Spremulli, L. L. (2014). *From Mammalian Tissues*. (May), 33–49. https://doi.org/10.1007/978-1-59745-365-3

T. Tancredi, A. Pastore, S. Salvadori, V. Esposito, P.A. Temussi, Interaction of sweet proteins with their receptor A conformational study of peptides corresponding to loops of brazzein, monellin and thaumatin. *European Journal of Biochemistry 271* (2004) 2231–2240. https://doi.org/10.1111/j.1432-1033.2004.04154.x

Tancredi, T., Iijima, H., Saviano, G., Amodeo, P., & Temussi, P. A. (1992). Structural determination of the active site of a sweet protein A 1H NMR investigation of pMNEI. *FEBS Letters*, *310*(1), 27–30. https://doi.org/10.1016/0014-5793(92)81138-C

Templeton, C. M., Pour, S. O., Hobbs, J. R., Blanch, E. W., Munger, S. D., & Conn, G. L. (2011). Reduced sweetness of a monellin (MNEI) mutant results from increased protein flexibility and disruption of a distant poly-(L-proline) II helix. *Chemical Senses*, *36*(5), 425–434. https://doi.org/10.1093/chemse/bjr007

Temussi, P. A. (2006). Natural sweet macromolecules: How sweet proteins work. *Cellular and Molecular Life Sciences*, *63*(16), 1876–1888 https://doi.org/10.1007/s00018-006-6077-8

Temussi, P. A. (2011). Determinants of sweetness in proteins: A topological approach. *Journal of Molecular Recognition*, *24*(6), 1033–1042. https://doi.org/10.1002/jmr.1152

V. Danilova, T. Roberts, G. Hellekant, Responses of single taste fibers and whole chorda tympani and glossopharyngeal nerve in the domestic pig, Sus scrofa. *Chem. Senses 24* (1999) 301–316. https://doi.org/10.1093/chemse/243.301

V. Danilova, Y. Danilov, T. Roberts, J.M. Tinti, C. Nofre, G. Hellekant, Sense of taste in a new world monkey, the common marmoset: recordings from the chorda tympani and glossopharyngeal nerves, *J. Neurophysiol. 8*8 (2002) 579–594. https://doi.org/10.1152/jn.00938.2001

Va, D., Sa, A., & Paul, S. (2009). Wonder animal model for genetic for genetic studies -Drosophila Melanogaster– its life cycle and breeding methods–A Review. *Sri Ramachandra Journal of Medicine*, *II*(2), 33.

Walters, D. E., & Hellekant, G. (2006). Interactions of the sweet protein brazzein with the sweet taste receptor. *Journal of Agricultural and Food Chemistry*, *54*(26), 10129–10133. https://doi.org/10.1021/jf062359y

Walters, D. E., Cragin, T., Jin, Z., Rumbley, J. N., & Hellekant, G. (2009). Design and evaluation of new analogs of the sweet protein brazzein. *Chemical Senses*, *34*(8), 679–683. https://doi.org/10.1093/chemse/bjp048

Winnig M, Bufe B, Kratochwil NA, Slack JP & Meyerhof W (2007) The binding site for neohesperidin dihydrochalcone at the human sweet taste receptor. *BMC Struct Biol 7*, 66. https://doi.org/10.1186/1472-6807-7-66

Wintjens, R., Viet, T. M. V. N., Mbosso, E., & Huet, J. (2011). Hypothesis/review: The structural basis of sweetness perception of sweet-tasting plant proteins can be deduced from sequence analysis. *Plant Science*, *181*(4), 347–354. https://doi.org/10.1016/j.plantsci.2011.06.009

World Health Organization (WHO). Diabetes , 2021 [Internet]. Available at: https://www.who.int/news-room/fact-sheets/detail/diabetes

World Health Organization (WHO). Obesity and overweight, 2021 [Internet]. Available at: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight

Yach, D., Stuckler, D., & Brownell, K. D. (2006). Epidemiologic and economic consequences of the global epidemics of obesity and diabetes. *Nature Medicine*, *12*(1), 62–66. https://doi.org/10.1038/nm0106-62

Yang, L., Zhu, K., Yu, H., Zhang, X., & Liu, B. (2019). The flexible loop is a new sweetness determinant site of the sweet-tasting protein: Characterization of novel sweeter mutants of the single-chain Monellin (MNEI). *Chemical Senses*, *44*(8), 607–614. https://doi.org/10.1093/chemse/bjz057

Yoon, S. Y., Kong, J. N., Jo, D. H., & Kong, K. H. (2011). Residue mutations in the sweetness loops for the sweet-tasting protein brazzein. *Food Chemistry*, *129*(4), 1327–1330. https://doi.org/10.1016/j.foodchem.2011.06.054

Zhang, F., Klebansky, B., Fine, R. M., Liu, H., Xu, H., Servant, G., et al. (2010). Molecular mechanism of the sweet taste enhancers. *Proceedings of the National Academy of Sciences of the United States of America, 107(10)*, 4752–4757. https://doi.org/10.1073/pnas.0911660107.

Zhao, M., Xu, X., & Liu, B. (2018). Structure basis of the improved sweetness and thermostability of a unique double-sites single-chain sweet-tasting protein monellin (MNEI) mutant. *Biochimie*, *154*, 156–163. https://doi.org/10.1016/j.biochi.2018.08.010

Tappy, L. (2018). Health Implications of Fructose Consumption in Humans. *References Series in Phytochemistry*, 285–309. https://doi.org/10.1007/978-3-319-27027-2_29

Grembecka, M. (2018). Sugar Alcohols as Sugar Substitutes in Food Industry. *Reference Series in Phytochemistry*, 547–573. https://doi.org/10.1007/978-3-319-27027-2_23

Arnold, D. L. (1983). Two-generation saccharin bioassays. *Environmental Health Perspectives*, *Vol. 50*(19), 27–36. https://doi.org/10.1289/ehp.835027

Morando Soffritti, Angela Guaragna, and Marco Manservigi. Potential Carcinogenic Risks of Aspartame In: Merillon J-M, Ramawat KG, editors. Sweeteners: pharmacology, biotechnology, and applications. *Cham: Springer International Publishing;* 2018. pp. 311–347.

Mazur, R. H., Goldkamp, A. H., James, P. A., & Schlatter, J. M. (1970). Structure-taste relationships of aspartic acid amides. *Journal of Medicinal Chemistry*, *13*(6), 1217–1221. https://doi.org/10.1021/jm00300a046

Otabe, A., Fujieda, T., Masuyama, T., Ubukata, K., & Lee, C. (2011). Advantame - An overview of the toxicity data. *Food and Chemical Toxicology*, *49*(SUPPL. 1), S2–S7. https://doi.org/10.1016/j.fct.2011.06.046

Angelini, L. G., Martini, A., Passera, B., & Tavarini, S. (2018). Cultivation of Stevia rebaudiana Bertoni and Associated Challenges. In *Reference Series in Phytochemistry*. https://doi.org/10.1007/978-3-319-27027-2_8

Prakash, I., & Chaturvedula, V. S. P. (2018). Steviol Glycosides: Natural Noncaloric Sweeteners. *Reference Series in Phytochemistry*

Word, K. (1976). *Institute of Pharmaceutical Sciences, Hiroshima University School of Medicine, Kasumi l-2-3, Hiroshima-shi, Japan*. *15*, 981–983 .

Fitch, C., & Keim, K. S. (2012). Position of the Academy of Nutrition and Dietetics: Use of Nutritive and Nonnutritive Sweeteners. *Journal of the Academy of Nutrition and Dietetics*, *112*(5), 739–758. https://doi.org/10.1016/j.jand.2012.03.009

Graebin, C. S. (2018). The Pharmacological Activities of Glycyrrhizinic Acid ("Glycyrrhizin") and Glycyrrhetinic Acid. *Reference Series in Phytochemistry*, 245–261. https://doi.org/10.1007/978-3-319-27027-2_15

Sharma, V., Katiyar, A., & Agrawal, R. C. (2018). Glycyrrhiza glabra: Chemistry and Pharmacological Activity. *Reference Series in Phytochemistry*, 87–100. https://doi.org/10.1007/978-3-319-27027-2_21

Masuda. T (2018). Sweet-Tasting Protein Thaumatin: Physical and Chemical Properties.

Ezura, H., & Hiwasa-Tanase, K. (2018). Mass Production of the Taste-Modifying Protein Miraculin in Transgenic Plants. *Reference Series in Phytochemistry*, 167–184. https://doi.org/10.1007/978-3-319-27027-2_17

Ming, D., & Hellekant, G. (1994). Brazzein, a new high-potency thermostable sweet protein from Pentadiplandra brazzeana B. *FEBS Letters*, *355*(1), 106–108. https://doi.org/10.1016/0014-5793(94)01184-2

Poirier, N., Roudnitzky, N., Brockhoff, A., Belloir, C., Maison, M., Thomas-Danguin, T., … Briand, L. (2012). Efficient production and characterization of the sweet-tasting brazzein secreted by the yeast pichia pastoris. *Journal of Agricultural and Food Chemistry*, *60*(39), 9807–9814. https://doi.org/10.1021/jf301600m

Hung, L. W., Kohmura, M., Ariyoshi, Y., & Kim, S. H. (1999). Structural differences in D and L-monellin in the crystals of racemic mixture. *Journal of Molecular Biology*, *285*(1), 311–321. https://doi.org/10.1006/jmbi.1998.2308

Aghera, N., & Udgaonkar, J. B. (2011). Heterologous expression, purification and characterization of heterodimeric monellin. *Protein Expression and Purification*, *76*(2), 248–253. https://doi.org/10.1016/j.pep.2010.11.002

LIU, X., MAEDA, S., HU, Z., AIUCHI, T., NAKAYA, K., & KURIHARA, Y. (1993). Purification, complete amino acid sequence and structural characterization of the heat-stable sweet protein, mabinlin II. *European Journal of Biochemistry*, *211*(1–2), 281–287. https://doi.org/10.1111/j.1432-1033.1993.tb19896.x

NIRASAWA, S., NISHINO, T., KATAHIRA, M., UESUGI, S., HU, Z., & KURIHARA, Y. (1994). Structures of heat-stable and unstable homologues of the sweet protein mabinlin. The difference in the heat stability is due to replacement of a single amino acid residue. *European Journal of Biochemistry*, *223*(3), 989–995. https://doi.org/10.1111/j.1432-1033.1994.tb19077.x

Faus, I. (2000). Recent developments in the characterization and biotechnological production of sweet-tasting proteins. *Applied Microbiology and Biotechnology*, *53*(2), 145–151. https://doi.org/10.1007/s002530050001

Sun, S. S. M., Zuo, W., Tu, H. M., & Xiong, L. (1996). Plant proteins: Engineering for improved quality. *Annals of the New York Academy of Sciences*, *792*(December), 37–42. https://doi.org/10.1111/j.1749-6632.1996.tb32488.x

Moore, D. G., & Buffington, E. C. (1968). Taste-Modifying Protein. 1241–1243.

Theerasilp, S., & Kurihara, Y. (1988). Complete purification and characterization of the taste-modifying protein, miraculin, from miracle fruit. *The Journal of Biological Chemistry*, *263*(23), 11536–11539. https://doi.org/10.1016/s0021-9258(18)37991-2

Suzuki, M., Kurimoto, E., Nirasawa, S., Masuda, Y., Hori, K., Kurihara, Y., … Kato, K. (2004). Recombinant curculin heterodimer exhibits taste-modifying and sweet-tasting activities. *FEBS Letters*, *573*(1–3), 135–138. https://doi.org/10.1016/j.febslet.2004.07.073

Yamashita, H., Theerasilp, S., Aiuchi, T., Nakaya, K., Nakamura, Y., & Kurihara, Y. (1990). Purification and complete amino acid sequence of a new type of sweet protein with taste-modifying activity, curculin. *Journal of Biological Chemistry*, *265*(26), 15770–15775. https://doi.org/10.1016/s0021-9258(18)55464-8

Neiers, F., Canivenc-Lavier, M. C., & Briand, L. (2016). What Does Diabetes "Taste" Like? *Current Diabetes Reports*, *16*(6). https://doi.org/10.1007/s11892-016-0746-2

Heidari-Beni, M., & Kelishadi, R. (2018). The Role of Dietary Sugars and Sweeteners in Metabolic Disorders and Diabetes. *Reference Series in Phytochemistry*, 225–243. https://doi.org/10.1007/978-3-319-27027-2_31

Li, X., Staszewski, L., Xu, H., Durick, K., Zoller, M., & Adler, E. (2002). Human receptors for sweet and umami taste. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(7), 4692–4696. https://doi.org/10.1073/pnas.072090199

Mace, O. J., Affleck, J., Patel, N., & Kellett, G. L. (2007). Sweet taste receptors in rat small intestine stimulate glucose absorption through apical GLUT2. *Journal of Physiology*, *582*(1), 379–392. https://doi.org/10.1113/jphysiol.2007.130906

Kniazeff, J., Prézeau, L., Rondard, P., Pin, J. P., & Goudet, C. (2011). Dimers and beyond: The functional puzzles of class C GPCRs. *Pharmacology and Therapeutics*, *130*(1), 9–25. https://doi.org/10.1016/j.pharmthera.2011.01.006

Robertson, C. W. (1936). The metamorphosis of Drosophila melanogaster, including an accurately timed account of the principal morphological changes. *Journal of Morphology*, *59*(2), 351–399. https://doi.org/10.1002/jmor.1050590207

Affleck, J. G., & Walker, V. K. (2019). Drosophila as a model for developmental toxicology: Using and extending the drosophotoxicology model. *Methods in Molecular Biology*, *1965*, 139–153. https://doi.org/10.1007/978-1-4939-9182-2_10

Malik, V. S., Popkin, B. M., Bray, G. A., Després, J. P., & Hu, F. B. (2010). Sugar-sweetened beverages, obesity, type 2 diabetes mellitus, and cardiovascular disease risk. *Circulation*, *121*(11), 1356–1364. https://doi.org/10.1161/CIRCULATIONAHA.109.876185

Malik, V. S., Popkin, B. M., Bray, G. A., Després, J. P., Willett, W. C., & Hu, F. B. (2010). Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes: A meta-analysis. *Diabetes Care*, *33*(11), 2477–2483. https://doi.org/10.2337/dc10-1079

Bloomgarden, Z. T. (2004). Type 2 Diabetes in the Young: The evolving epidemic. *Diabetes Care*, *27*(4), 998–1010. https://doi.org/10.2337/diacare.27.4.998

G. Roglic. (2014). *The Burden of Mortality Attributable to Diabetes*. *28*(9), 2130–2135.

Lobstein, T., Baur, L., & Uauy, R. (2004). Obesity in children and young people: A crisis in public health. *Obesity Reviews, Supplement*, *5*(1), 4–104. https://doi.org/10.1111/j.1467-789x.2004.00133.x