

DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

**IDENTIFICATION OF BIOACTIVE AND
ALLERGEN PROTEINS BY *IN SILICO* TOOLS**



by
Ceylan TÜRKMEN

July, 2019
İZMİR

IDENTIFICATION OF BIOACTIVE AND ALLERGEN PROTEINS BY *IN SILICO* TOOLS

**A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University
In Partial Fulfilment of the Requirements for the Degree of Master of Science
Department of Chemistry, Biochemistry Program**

**by
Ceylan TRKMEN**

**July, 2019
İZMİR**

M. Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled “**IDENTIFICATION OF BIOACTIVE AND ALLERGEN PROTEINS BY *IN SILICO* TOOLS**” completed by **CEYLAN TRKMEN** under supervision of **PROF.DR LEVENT AVAŞ** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Levent AVAŞ

(Jury Member)

(Jury Member)

Prof. Dr. Kadriye ERTEKİN

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

I would like to express my heartfelt thanks to my advisor Prof. Dr. Levent ÇAVAŞ, who has always supported me since I started my thesis work and believes in what I can do, and shares his knowledge and experience with me. I sincerely thank you to my colleagues Yeşim YILMAZ ABEŞKA and Gudrat SAFAROV for their support and valuable friendship.

I would like to present my thanks to my dear brother Yusuf TÜRKMEN, who has always supported me throughout my life and whose presence has always given me strength. My sincere thanks to my younger brother Ömer Ramazan TÜRKMEN who has always given importance to my advice and is a model to everyone with his conscience besides being a strong person. I would like to thank my mother Belgin YALIN, who believes that I can whatever do want, supports me on everything and makes me come to these days with her eternal love. I would like to thank my friend Ferhat GÜNDÜZALP with all my sincerity and love, who was always there for me whenever I needed help.

In addition, I owe a debt of gratitude to all my friends who have been with me throughout my thesis work.

Ceylan TÜRKMEN

IDENTIFICATION OF BIOACTIVE AND ALLERGEN PROTEINS BY *IN SILICO* TOOLS

ABSTRACT

Bioinformatics tools (*in silico* tools) can be used to identify allergens from different sources and to find the compounds that are associated with food allergens. Identification of allergen proteins by using wet-lab technology is a time-consuming and also costly process. Thanks to the development in the field of bioinformatics, it is now possible to estimate the allergen proteins. In the present thesis, four different computer-based comparison tools were used to investigate the allergic effects of potential allergen proteins in different plants and previously unknown to be allergens. For multiple sequence alignment Clustal W, for homology modelling SWISS-MODEL, for comparison of protein parameters ProtParam and for allergen estimation AlgPred were used. In addition, BIOPEP tool is used for *in silico* hydrolysis. Considering the results, seven proteins *Corchorus olitorius*, *Cucumis sativus*, *Capsicum chinense*, *Carica papaya*, *Morus notabilis*, *Theobroma cacao* and *Jatropha curcas* were defined as potential allergens in different species. At the same time, potential allergen proteins were analysed by BIOPEP. At the end of the analysis, the bioactive peptides in the potential allergen proteins and their functions were explained. Non-allergenic forms of potential allergen proteins were tried to be obtained with the results of *in silico* hydrolysis. In conclusion, *in silico* tools developed under the field of bioinformatics can provide a big contribution to the estimation of unknown allergen proteins in different fruits. Based on the *in silico* results, physicians can suggest people who have allergenicity to kiwellin not to consume the fruits that contain kiwellin-like proteins.

Keywords: Allergen, *in silico* tools, kiwellin, kiwi, bioinformatics

BİOAKTİF VE ALLERJEN PROTEİNLERİN *IN SILICO* ARAÇLARLA TANIMLANMASI

ÖZ

İnternet bağlantılı veri tabanları ve onlarla birlikte kullanılan biyoinformatik analiz araçları (*in silico*) kullanılarak gıda bileşenlerinin alerjenitesi ile ilişkili biyoinformatik analizler yapılmaktadır. Alerjen proteinlerinin klasik laboratuvar teknolojisi kullanılarak tanımlanması zaman alıcı ve aynı zamanda maliyetli bir işlemdir. Son yıllarda, biyoinformatik alanındaki gelişim sayesinde, alerjen proteinlerini tahmin etmek artık mümkündür. Bu çalışmada biyoinformatik analiz araçları kullanılarak alerjen olduğu bilinen proteinlere benzer potansiyel alerjen proteinler araştırılmıştır. Sunulan tezde farklı bitkilerde bulunan ve alerjenitesi daha önce bilinmeyen potansiyel alerjen proteinlerin alerjik etkisini araştırmak için bilgisayar tabanlı, dört farklı karşılaştırma aracı kullanılmış olup potansiyel alerjenler farklı parametrelerde incelenmiştir. Çoklu sekans karşılaştırması için Clustal W, homoloji modelleme için SWİSS-MODEL, protein parametrelerini karşılaştırma için ProtParam ve alerjenite tahmini için AlgPred kullanılmıştır. Sonuçlarının ışığı altında yedi potansiyel alerjen *Corchorus olitorius*, *Cucumis sativus*, *Capsicum chinense*, *Carica papaya*, *Morus notabilis*, *Theobroma cacao* ve *Jatropha curcas* türlerinde tanımlanmıştır. Araştırılan potansiyel alerjen proteinler aynı zamanda BIOPEP aracılığıyla analiz edilmiştir. Sonuçlar göz önüne alındığında, farklı türlerde yedi protein potansiyel alerjenler olarak tanımlandı. Analiz sonunda potansiyel alerjen proteinlerde bulunan biyoaktif peptitler ve bu biyoaktif peptitlerin işlevleri açıklanmıştır. Potansiyel alerjen proteinlerin *in silico* hidrolizi sonucu yorumlanarak potansiyel alerjen proteinlerin alerjen olmayan formu elde edilmeye çalışılmıştır. Sonuç olarak, biyoinformatik alanında kapsamında geliştirilen *in silico* araçlar bilinmeyen alerjen proteinlerin tahmini için büyük bir katkı sağlayabilir. Analiz sonuçlarına dayanarak, doktorlar kiwellin proteinine alerjisi olan kişilerin kiwellin-benzeri proteinleri içeren meyveleri tüketmemesi için tavsiyede bulunabilirler.

Anahtar kelimeler: Alerji, *in silico* araçlar, kiwellin, kivi, biyoinformatik

CONTENTS

	Page
M. Sc. THESIS EXAMINATION RESULT FORM.....	ii
ACKNOWLEDGMENTS	iii
ABSTRACT.....	iv
ÖZ	v
LIST OF FIGURES	ix
LIST OF TABLES	xii
CHAPTER ONE - INTRODUCTION	1
1.1 Bioinformatics and its development	1
1.1.1 Human Genome Project	2
1.2 Application Areas of Bioinformatics.....	3
1.2.1 Omics technologies	3
1.2.2 Clinical Bioinformatics	3
1.2.3 Phylogenetic Researches	4
1.2.4 Bioinformatics in Food Science	4
1.3 Bioinformatics Database and <i>in silico</i> tools	5
1.3.1 ExPASy	6
1.3.2 Entrez	7
1.3.3 UniProt	8
1.3.4 DDJB.....	9
1.3.5 GenBank.....	10
1.3.6 PDB	10
1.3.7 NCBI Protein Databases	11
1.3.8 EMBL-EBI.....	12

1.3.9 Allergen Nomenclature	12
1.3.10 BLAST	13
1.3.11 FASTA	14
1.3.12 Clustal ω	15
1.3.13 AlgPred	17
1.3.14 SWISS-MODEL	18
1.3.15 ProtParam	19
1.3.16 BIOPEP	20
1.4 Protein and Protein Structure	21
1.5 Bioactive Peptides	23
1.5.1 Bioactive Peptides Investigated in Thesis	25
1.5.1.1 ACE-inhibitor Bioactive Peptides	25
1.5.1.2 Antioxidative Peptides	26
1.5.1.3 Antithrombotic Peptides	27
1.5.1.4 Antiamnesic Peptides	27
1.5.1.5 Dipeptidyl Endopeptidase – IV (DPP – IV) Inhibitor Peptides	28
1.5.1.6 Ubiquitin–Proteasome System (UPS) Activator Peptides	28
1.6 Food Allergy and Allergen Protein	28
1.6.1 Kiwellin	29
1.7 The Aim of Thesis	30
CHAPTER TWO - MATERIAL AND METHOD	31
2.1. Retrieval of Allergen Protein Sequences	31
2.2. Searching Similar Proteins with BLASTp	32
2.3. Multiple Sequence Alignment	32

2.4. Creation of Phylogenetic Tree	32
2.5. ProtParam Analysis	33
2.6. Homology Modelling with SWISS-MODEL	33
2.7. AlgPred Analysis	33
2.8. <i>In silico</i> Hydrolysis with BIOPEP	33
CHAPTER THREE - RESULTS.....	35
3.1. The Result of Multiple Sequence Alignment and Phylogenetic Tree	36
3.2. The Results From ProtParam.....	41
3.3. The Homology Modelling	45
3.4. AlgPred Results	47
3.5. BIOPEP Results.....	48
3.5.1 Properties of bioactive peptides in <i>Cucumis sativus</i>	49
3.5.2 Properties of bioactive peptides in <i>Capsicum chinense</i>	59
3.5.3 Properties of bioactive peptides in <i>Vitis vinifera</i>	70
CHAPTER FOUR - CONCLUSION	81
REFERENCES.....	84

LIST OF FIGURES

	Page
Figure 1.1 Screen shot of ExPASy Home page (retrieval date: 25.03.2019).....	6
Figure 1.2 Screen shot of ExPASy search on the database with cross-source interface (retrieval date: 25.03.2019)	7
Figure 1.3 Screen shot of NCBI Entrez Home Page (retrieval date: 25.03.2019)	8
Figure 1.4 Screen shot of UniProt database (retrieval date: 25.03.2019)	9
Figure 1.5 Screen shot of DDJB database home page (retrieval date: 25.03.2019)	9
Figure 1.6 Example of searching data in the GenBank database (retrieval date: 25.03.2019)	10
Figure 1.7 Screen shot of Protein Data Bank home page (retrieval date: 25.03.2019)	11
Figure 1.8 Screen shot of NCBI protein database home page (retrieval date: 25.03.2019)	11
Figure 1.9 Screen shot of EMBL-EBI database home page (retrieval date: 25.03.2019)	12
Figure 1.10 Screen shot of Allergen Nomenclature database home page (retrieval date: 25.03.2019)	13
Figure 1.11 Screen shot of BLAST program home page (retrieval date: 25.03.2019)	14
Figure 1.12 Screen shot of FASTA program home page on the EMBL-EBI site (retrieval date: 25.03.2019)	15
Figure 1.13 Screen shot of Clustal ω home page (retrieval date: 25.03.2019)	16
Figure 1.14 Screen shot of AlgPred Home Page (retrieval date: 25.03.2019).....	17
Figure 1.15 Screen shot of SWISS-MODEL home Page (retrieval date: 25.03.2019)	19
Figure 1.16 Screen shot of proteolytic enzyme list in BIOPEP (retrieval date: 25.03.2019)	21
Figure 1.17 The view of the protein structures	23
Figure 1.18 Effects of bioactive peptides on the body systems.	24
Figure 1.19 The Renin-angiotensin system.....	26
Figure 2.1 Screenshot of the enzymes contained in the BIOPEP tool	34

Figure 3.1 Comparison of 10 potential allergen proteins (The NCBI codes of the proteins can be found in Table 3.1).....	36
Figure 3.2 Phylogenetic tree of the 10 potential allergen protein sequences.....	41
Figure 3.3 Pairwise sequence alignment of P85261 and P84527	43
Figure 3.4 Dimensional visualization of each protein by SWISS-MODEL a) <i>Actinidia deliciosa</i> , b) <i>Capsicum chinense</i> , c) <i>Carica papaya</i> , d) <i>Corchorus olitorius</i> , e) <i>Cucumis sativus</i> , f) <i>Jatropha curcas</i> , g) <i>Actinidia chinensis</i> h) <i>Morus notabilis</i> , i) <i>Prunus persica</i> , j) <i>Prunus persica</i> , j) <i>Punica granatum</i> , k) <i>Theobroma cacao</i> l) <i>Vitis vinifera</i>	46



LIST OF TABLES

	Page
Table 1.1 The meaning of Clustal ω abbreviations.....	16
Table 1.2 The meaning of colours in Clustal ω	17
Table 1.3 Abbreviation list of amino acids	22
Table 2.1 Kiwi allergens on allergen.org (Retrieved 27.03.19).....	31
Table 3.1 Table of potential allergen proteins that is formed according to results obtained with BLASTp.....	35
Table 3.2 Protein parameters of potential allergen protein	42
Table 3.3 Amino acid number and percentages in potential allergen proteins from different plants, part 1.....	44
Table 3.4 Amino acid number and percentages in potential allergen proteins from different fruits, part 2.....	45
Table 3.5 AlgPred analysis of potential allergens, part 1	47
Table 3.6 AlgPred analysis of potential allergens, part 2	48
Table 3.7 DH _t , AE, W, BE and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with chymotrypsin.	49
Table 3.8 DH _t , AE, W, BE and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with trypsin.	49
Table 3.9 DH _t , AE, W, BE and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with pepsin.....	50
Table 3.10 DH _t , AE, W, BE and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with proteinase K.....	50
Table 3.11 DH _t , AE, W, BE and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with elastase.....	51
Table 3.12 DH _t , AE, W, BE and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with V-protease.....	51
Table 3.13 DH _t , AE, W, BE and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with thermolysin.	51
Table 3.14 DH _t , AE, W, BE and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with chymotrypsin.	52

Table 3.15 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with plasmin.....	52
Table 3.16 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with cathepsin.	52
Table 3.17 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with chymase.	53
Table 3.18 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with papain.....	53
Table 3.19 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with ficin.	54
Table 3.20 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with leukocyte elastase.	54
Table 3.21 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with metridin.....	55
Table 3.22 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with pancreatic elastase II.....	55
Table 3.23 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with bromelain.	55
Table 3.24 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with oligopeptidase B.	56
Table 3.25 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with calpain 2.....	55
Table 3.26 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with glycyl endopeptidase.	56
Table 3.27 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with oligopeptidase F.....	57
Table 3.28 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with proteinase P1 (lactocepin).	57
Table 3.29 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with pepsin (pH>2).	58
Table 3.30 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with cocolysin enzyme.	58

Table 3.31 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with subtilisin.....	58
Table 3.32 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. sativus</i> obtained through <i>in silico</i> hydrolysis with V-8 protease (Glutamyl endopeptidase).	59
Table 3.33 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with chymotrypsin.	59
Table 3.34 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with trypsin.	60
Table 3.35 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with pepsin.	60
Table 3.36 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with proteinase K.	60
Table 3.37 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with pancreatic elastase.....	61
Table 3.38 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with prolyl oligopeptidase.....	61
Table 3.39 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with prolyl thermolysin.....	62
Table 3.40 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with chymotrypsin.	62
Table 3.41 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with plasmin.....	63
Table 3.42 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with cathepsin.	63
Table 3.43 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with clostripain.....	63
Table 3.44 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with chymase.	64
Table 3.45 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with papain.....	64

Table 3.46 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with ficin.	65
Table 3.47 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with leukocyte elastase.	65
Table 3.48 DH _t , A _E , W, B _E and V values o in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with metridin.	65
Table 3.49 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with pancreatic elastase II.	66
Table 3.50 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with stem bromelain.	66
Table 3.51 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with oligopeptidase B.	67
Table 3.52 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with calpain 2.	67
Table 3.53 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with glycyl endopeptidase.	68
Table 3.54 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with oligopeptidase F.	68
Table 3.55 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with proteinase P1 (lactocepin).	68
Table 3.56 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with pepsin (pH>2).	69
Table 3.57 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with cocolysin.	69
Table 3.58 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>C. chinense</i> obtained through <i>in silico</i> hydrolysis with subtilisin.	70
Table 3.59 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with chymotripsin.	70
Table 3.60 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with trypsin.	71
Table 3.61 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with pepsin.	71

Table 3.62 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with proteinase K.....	71
Table 3.63 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with elastase.....	72
Table 3.64 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with V-protease.....	72
Table 3.65 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with thermolysin.	72
Table 3.66 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with chymotrypsin.	73
Table 3.67 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with plasmin.....	73
Table 3.68 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with cathepsin.	74
Table 3.69 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with chymase.	74
Table 3.70 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with papain.....	74
Table 3.71 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with ficin.	75
Table 3.72 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with leukocyte elastase.	75
Table 3.73 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with metridin.....	75
Table 3.74 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with pancreatic elastase II.....	76
Table 3.75 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with stem bromelain.	76
Table 3.76 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with oligopeptidase B.	77
Table 3.77 DH _t , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with calpain 2.....	77

Table 3.78 DH _i , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with glycyl endopeptidase.	77
Table 3.79 DH _i , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with oligopeptidase F.....	78
Table 3.80 DH _i , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with proteinase P1 (lactocepine).	78
Table 3.81 DH _i , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with pepsin (pH>2).	78
Table 3.82 DH _i , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with coccolysin.	79
Table 3.83 DH _i , A _E , W, B _E and V values obtained through <i>in silico</i> hydrolysis with subtilisin.....	79
Table 3.84 DH _i , A _E , W, B _E and V values in kiwellin-like protein in <i>V. vinifera</i> obtained through <i>in silico</i> hydrolysis with V-8 protease (Glutamyl endopeptidase).	79

CHAPTER ONE

INTRODUCTION

1.1 Bioinformatics and its development

Bioinformatics was defined in early of 1980 to compare proteins and DNA data. However, this term was initially defined as informatics within biological systems. Paulien Hogeweg and Ben Hesper used first bioinformatics in 1970 as “informatics work processes in biotic systems” (Hogeweg, 2011).

Bioinformatics is an area created by the integration of fields such as biology, biochemistry, computer science and statistics in order to make biological data more comprehensible. Bioinformatics has three main objectives. The first aim is to bring the biological data that has been created so far to be accessible to the researchers. The second objective is to develop tools (*in silico* tools) for the fast, efficient and reliable evaluation of biological data. The third is the evaluation of new biological data with the result of *in silico* tools (Luscombe, Greenbaum, & Gerstein, 2001).

The basis of bioinformatics has begun with the development of protein sequencing method. Frederick Sanger and his team analysed the amino acid sequence of bovine insulin hormone in Cambridge University 1945-1955. Thus, they performed the sequencing of a whole protein at the first time ever (Sanger and Tuppy, 1951; Hagen, 2000). Sanger's insulin hormone sequencing study revealed that proteins have a unique primer structure and Sanger won the Nobel Prize in Chemistry in 1958 with his work on protein structure (Hagen, 2000). John Kendrew and his team with the help of a computer carried out the first study describing the three-dimensional structure of myoglobin protein (Attwood, Gisel, & Eriksson, Bongcam-Rudloff, 2011; Kendrew et al., 1958). The use of computers in order to research protein chemistry began in the 1960s; FORTRAN, a high-level programming language, was developed by International Business Machines (IBM) and it provides an opportunity an advancement of computerized biology (Hagen, 2000). With the support of the National Biomedical Research Foundation (NBRF), Margaret O. Dayhoff used mathematical

and computational methods to analyse amino acid sequence data in biochemistry and she wrote algorithms for extracting amino acid sequences with FORTRAN program. In 1965, she compiled the 65 protein sequences obtained until that time in the publication Atlas of Protein Sequence and Structure, thus she created the first database. This publication was later transformed into an online database called Protein Information Resource (PIR), founded in 1983 (Attwood et al., 2011; Hagen, 2000).

1.1.1 Human Genome Project

A high-resolution image of the crystallized deoxyribonucleic acid (DNA) fibres was obtained by x-ray crystallography. In this way, the first information about the structure of DNA was revealed (Attwood et al., 2011). Then the structure of the DNA was elucidated that it consists of a double helical structure and the outer branches of the helical structure are formed of sugar and phosphate (Watson & Crick, 1953; Attwood et al., 2011). Between 1961 and 1966, Nirenberg and Matthaei elucidated the genetic coding for protein synthesis. In addition, it has now been discovered that phenylalanine amino acid is synthesized from the messenger ribonucleic acid (mRNA) strands containing only the UUU nucleotide sequence (Trifonov, 2000). Ray Wu performed the first nucleic acid sequence analysis of DNA with the development of protein sequencing in 1972. Sanger developed the method of Wu in 1975 (Lesk, 2005). Thus, Sanger method emerged. Gilbert and Maxam discovered the method of DNA sequencing by chemical degradation in the same year (Lesk, 2005). The main purpose of the Human Genome Project can be defined as creating the human gene map and establishing the location of the genes with unknown function. In 1986, the project was introduced at the scientific meeting organized by the US Department of Energy (DOE) and in 1988; the National Research Commission prepared a report for the project. In 1990, the US National Ministry of Health (NIH) and DOE together with the UK, France, Germany, China, and Japan and John Watson started the Human Genome Project (Lander et al., 2001). In The Human Genome Project, approximately 20,500 human genes and 3 billion nucleotides were identified. The Human Genome Project has enabled the development of the bioinformatics field and databases such as the Genome Database (GDB) and *in silico* tools have been created to store the large amounts of genomic mapping data obtained with the project (Chial, 2008).

1.2 Application Areas of Bioinformatics

1.2.1 Omics technologies

In general, technologies that analyse the entire class of biomolecules are called omics technologies (Derks, Hoeijmakers, & Pothof, 2014). In addition, it aims to elucidate the regulatory mechanisms and interactions of related biomolecules such as the gene, ligand, and protein (Ballereau et al., 2013). Thanks to this approach, a complex system is better understood when taken as a whole (Horgan and Kenny, 2011). Omics technology, as it can be used better understand normal physiological processes also; it can be applied to help us understand the molecular aetiology of the disease and (Horgan and Kenny, 2011). The omics technologies are divided into many branches. In general, there are four major branches: genomics, proteomics, transcriptomic, metabolomics (Derks, Hoeijmakers & Pothof, 2014). Genome can be defined as whole DNA in a cell or organism and investigation of all genes is called genomics (Horgan & Kenny, 2011). The definition of proteome refers to all proteins encoded in a cell or organism. (Horgan & Kenny, 2011). Investigation of the protein-protein interaction, modifications of the functional properties, and the structure of proteins in the organism or cell is called proteomics (Ballereau et al., 2013). Metabolites such as low molecular weight lipids, amino acids, nucleic acids, peptides in a cell, tissue or organism involved in metabolic reactions are defined as metabolome (Horgan & Kenny, 2011; Ballereau et al., 2013). Metabolomics refers to the determination of metabolites in a particular sample, in cells or tissues, and identification of known target metabolites, the biochemical changes in this cell or tissue (Shulaev, 2006; Valdes, Glass & Spector, 2013).

1.2.2 Clinical Bioinformatics

Omics technologies are used for cancer research, drug discovery, and development, biomarker development, gene therapy, and personal drug development in many clinical applications (Wang & Liotta, 2011; Wu & Wang, 2015).

The *in silico* studies of protein-ligand interactions allow the identification of novel precursor compounds for synthetic drug development. Genome sequences are used to

identify potential drug targets, protein expression, and pattern data (Lesk, 2005). In addition, thanks to *in silico* modelling methods such as homological modelling is discovered depending on the structure of the protein the structure of the component to be bound to the target protein (Chen et al., 2012). Protein, peptide or physicochemical-based biomarkers are revealed by data mining methods and omics technologies, (Baumgartner, Osl, Netzer, & Baumgartner, 2011). Oncogenomics-based studies with the identification of one of the biomarker tumor-forming genes for liver cancer indicate the importance of biomarkers for cancer research (Sawey et al., 2011). Cancer research with *in silico* is based on data mining and the identification of hidden patterns (Mokhararak, Al Khalaf, & Altman, 2012). Some databases are established for data showing alterations such as DNA mutation, gene expression profile (Charoentong et al., 2012). In addition, studies are carried out with bioinformatics tools to calculate the probability of cancer in studies (Mokhararak et al., 2012).

1.2.3 Phylogenetic Researches

Phylogenetic studies are the researches that analyse the evolutionary relationship between groups of organisms using sequence comparison methods (Lesk, 2005). This relationship is called phylogeny. The term phylogenetic is of Greek origin. Taxonomy, the classification and nomenclature of organisms, is greatly influenced by phylogenetic but it is methodically and logically different (Edwards, 1964).

1.2.4 Bioinformatics in Food Science

Bioinformatics applications in the field of food have been based in the field of genomics and drug development (Holton, Vijayakumar, & Khaldi, 2013). Changing the genetics of organisms is used in the agricultural field to produce food resistant to diseases and climatic conditions and pests. Omics technologies are used characterization of genetically modified organisms and the detection of possible unknown effects and potential risk analysis (Valdés, Simó, Ibáñez & García-Cañas, 2013). The linkages between genomic sequences of pathogenic microorganisms in food poisoning and the epidemic are examined concerning food safety (Cooper et al., 2014a; Cooper et al., 2014b). In addition, Next Generation Sequencing methods

provide sequence analysis of new or developing pathogens that are not characterized by current methods (Bergholz, Switt & Wiedmann, 2014). New biomarkers for the rapid detection of stereotypes of pathogenic microorganisms such as *Salmonella* are being investigated. For this purpose, the information of the stereotype-determining gene is extracted by sequencing method (Zou, 2014). Omics technologies are used to characterization of food pathogenic microorganisms and determine the resistance of pathogens to stress and antimicrobial use (Bergholz et al., 2014). Bioinformatics is also used in taste and aroma studies. The chemical properties of taste and aroma components are stored in databases such as BitterDB, SupperSweet and this data are used in taste research (Ahmed et al., 2011; Wiener, Shudler, Levit & Niv, 2012). The relationship between the bacteria used in fermented foods and the aroma is clarified by ascertaining the sequences of bacteria (Holton et al., 2013). Another area of application of bioinformatics is to investigate the allergen potential of the component with *in silico* tools such as SDAP (Ivanciuc, Schein & Braun, 2003), Allermatch (Fiers, Kleter, Nijland, Peijnenburg, Nap & Van Ham, 2004) and AllerTop (Dimitrov, Flower & Doytchinova, 2013), and to create databases related to these allergen compounds.

In silico approaches are used in the discovery of novel bioactive peptides that exhibit therapeutic properties. Therefore, *in silico* approaches have begun to replace traditional methods, which are quite time-consuming, costly and require many attempts (Udenigwe & Mohan, 2014). Another tool such as (Quantitative Structure-Activity Relationship, QSAR) is used to predict new bioactive peptides. This tool uses the regression analysis equation to establish a relationship between the structure and activity of the compounds (Gu, Majumder & Wu, 2011).

The presence of cryptic bioactive peptides in the protein sequence is analysed using the sequence mapping method *in silico* tools (Lacroix & Li-Chan, 2012; Dziuba, Dziuba, & Iwaniak, 2009; Udenigwe, Gong & Wu, 2013).

1.3 Bioinformatics Database and *in silico* tools

The fundamental goal of this field is to create new perspectives in biology, as well as the discovery of new biological insights. These areas could be grouped as follows (NCBI, 2019):

- New algorithms and statistics are developed for evaluate the relationships between big dataset
- The analysis and interpretation of diverse data types inclusive of protein domains, protein structures, nucleotide, and amino acid sequences
- Developing tools that provide an opportunity for effective access

1.3.1 ExPASy

ExPASy (Expert Protein Analysis System) is one of the most known bioinformatics sources for proteomics in the world (EXPASY, 2019). The Swiss Institute of Bioinformatics (SIB) has developed ExPASy. The databases and *in silico* tools related each category are existed in this portal.

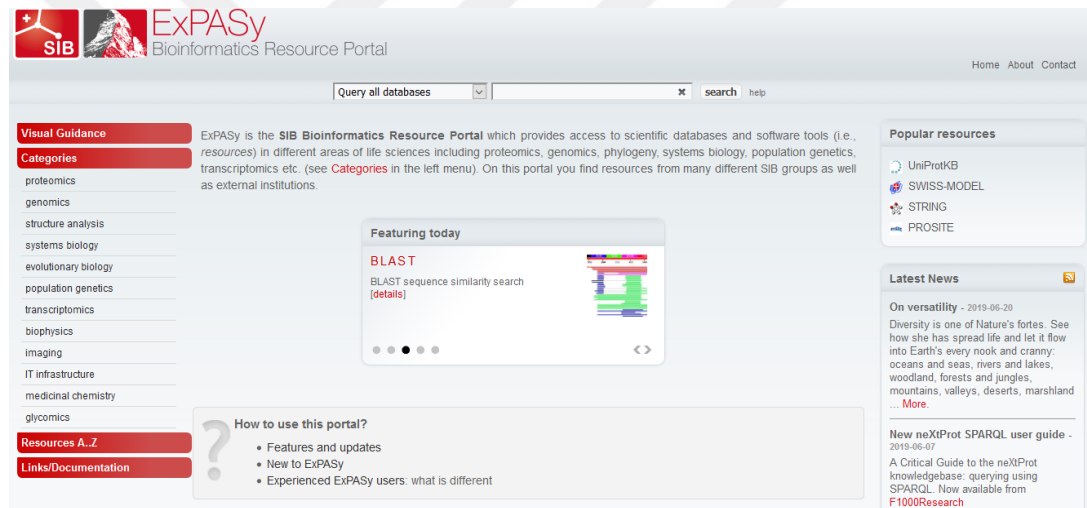


Figure 1.1 Screen shot of ExPASy Home page (Retrieval date: 25.03.2019)

The portal designed for both specialist users and people who are not familiar with a specific area in the life sciences. Thanks to the cross-source search interface of the portal, the search button on the home page can extract all the databases in the ExPASy with the keyword, code or name entered.

The screenshot shows the ExPASy Bioinformatics Resource Portal interface. At the top, there's a search bar with 'chymotrypsin' entered and a 'search' button. Below the search bar, a dropdown menu shows 'Query all databases'. The main content area displays search results for 'chymotrypsin'. On the left, there's a sidebar with 'Categories' and 'Resources A..Z'. The main table lists various databases and their corresponding hit counts and categories.

Resource	Hits	Category	Comment
EPD		ge	No valid response from server.
neXtProt	169 hits	pr	Entries matching the query chymotrypsin in neXtProt
OMA	1758 hits	ev, ge	
OrthoDB	10739 hits	ev, ge	
PROSITE	16 hits	pr	Protein domains, families and functional sites, and signatures to identify them
STRING	3 hits	pr	STRING interaction neighborhoods found in 2 species.
SWISS-MODEL Repository	3881 hits	bl, pr, st	Number of SWISS-MODEL Homology models matching chymotrypsin
UniProtKB	97326 hits	pr	Protein sequence and functional information from the UniProt Knowledgebase
ViralZone	0 hits	pr	General molecular and epidemiological information on viral genus and families
CellSaurus	0 hits	sy	Cell lines
ENZYME	10 hits	pr	Enzyme nomenclature
HAMAP	2 hits	pr	Manually created annotation rules for protein families
MetaNetX	1 hit	pr, sy	search results for MetaNetX
miROrtho	0 hits	ev, ge	orthologous group(s) returned by your request
MyHits	477 hits	pr	Protein found: 476, Motif found: 1, Other node found: 0;

Figure 1.2 Screen shot of ExPASy search on the database with cross-source interface (retrieval date: 25.03.2019)

1.3.2 Entrez

The National Center for Biotechnology Information (NCBI) is an organization, which was created in 1988 in order to develop information systems (NCBI, 2019). Entrez is a database that founded by NCBI, which provides access to 39 databases and lots of *in silico* tools such as BLAST (Entrez, 2019). The database consists of subtitles such as literature, protein, BLAST sequence analysis, gene and expression, genome, genetic and drug, chemical and biological determination, domain and structure, taxonomy, DNA and RNA (Madden, 2013).

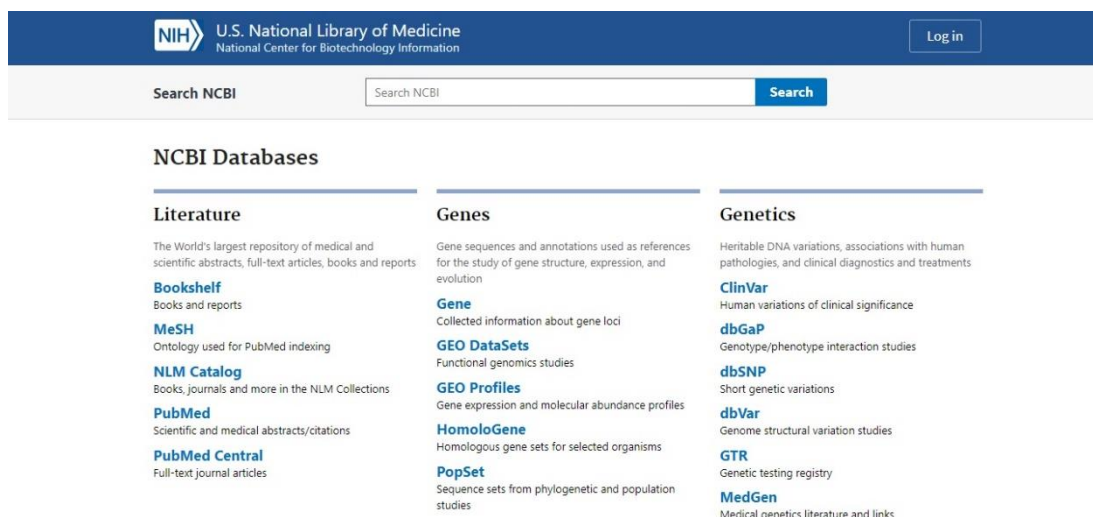


Figure 1.3 Screen shot of NCBI Entrez Home Page (Retrieval date: 25.03.2019)

1.3.3 UniProt

UniProt is one of the significant database in which protein sequences and annotations reaching 80 million sequences are recorded (UNIPROT, 2019). UniProt is a protein and gene sequence database established in 2002 and it was established by the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI), Protein Information Resource (PIR) and Swiss Bioinformatics Institute (SIB) (Figure 1.4). The database provides the function, taxonomy, structure, cross-source, related publication and similar protein information in relation to protein, and provides access to FASTA format information showing primary structure. Selected sequences can be analysed using comparison programs available in the UniProt database.

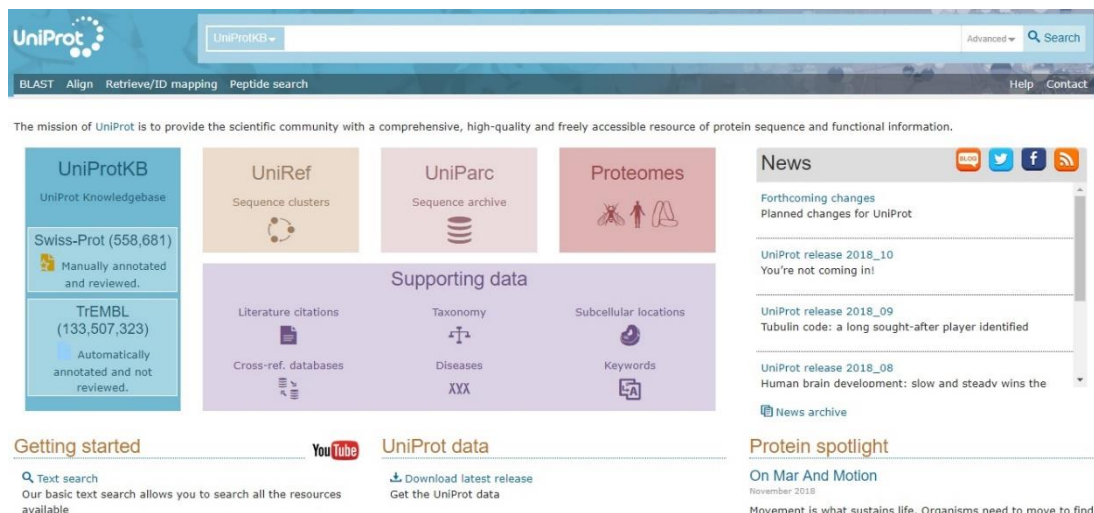


Figure 1.4 Screen shot of UniProt database (Retrieval date: 25.03.2019)

1.3.4 DDJB

The National Genetics Institute established the DNA Data Bank of Japan (DDBJ) in 1984 as a DNA sequence database (DDBJ, 2019). DDBJ, together with EMBL Bank and GenBank, forms the International Nucleotide Sequence Database (INSDC), which records updates in these databases every 24 hours (INSDC, 2019; Tateno et al., 2002).

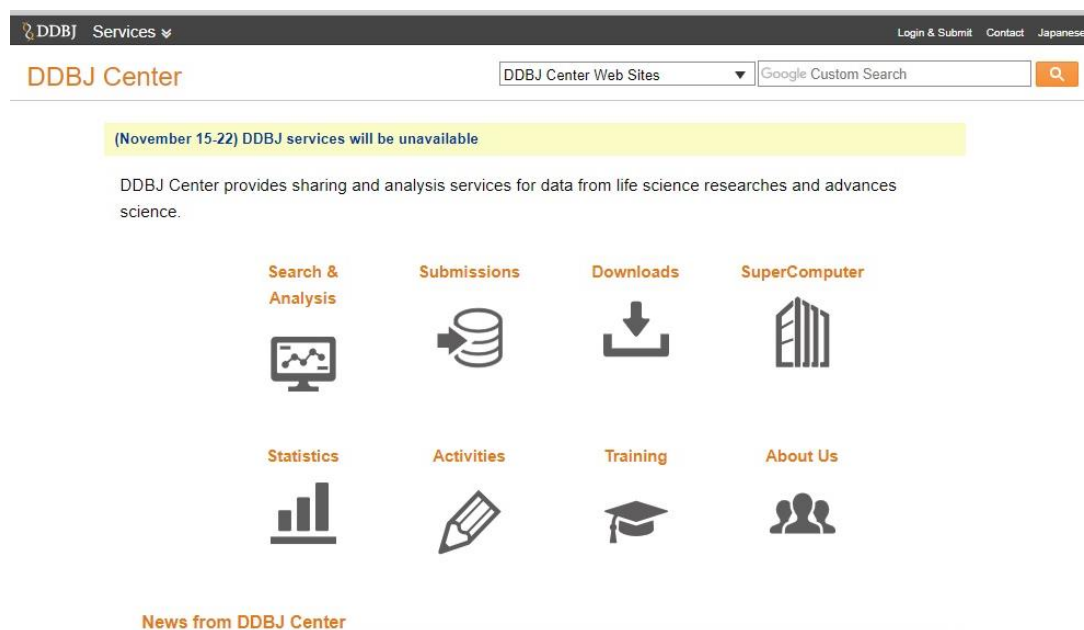


Figure 1.5 Screen shot of DDJB database home page (Retrieval date: 25.03.2019)

1.3.5 GenBank

GenBank (Genbank, 2019) was founded by NCBI in 1982 and it is the nucleotide database accessible via the Entrez access system (Benson et al., 2012) (Figure 1.6). GenBank is a part of the International Nucleotide Database Association. Each data has its own unique access number, and these numbers apply to DDJB and EMBL databases. It is open to the public users and it allows users to enter data and use *in silico* tools. The majority of the data on the site consists of the data obtained from the “expressed sequence tag”, “genome survey sequence” and “whole-genome shotgun” studies. GenBank provides use of BLAST programs for sequence similarity studies (Benson et al., 2012).

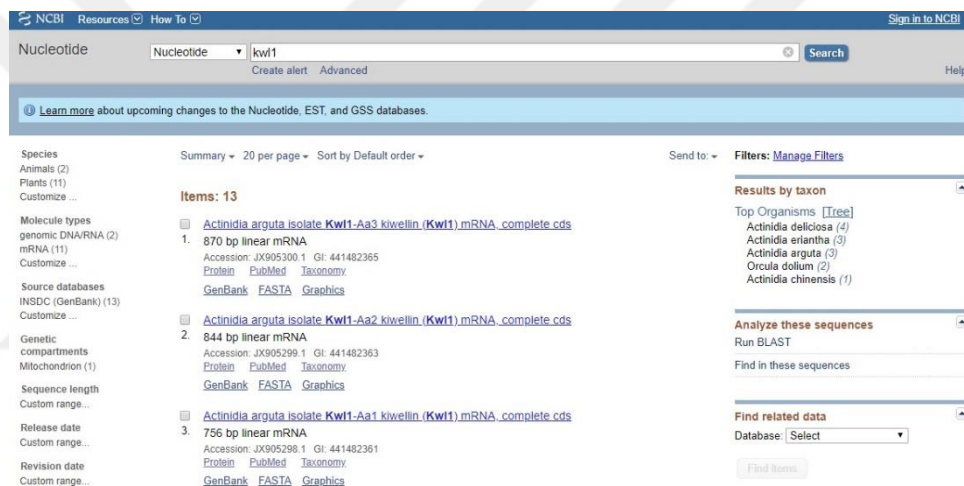


Figure 1.6 Example of searching data in the GenBank database (Retrieval date: 25.03.2019)

1.3.6 Protein Data Bank

Protein Data Bank (PDB) is a database containing “crystalline structures of biological macromolecules” (Protein data bank, 2019) (Berman et al., 2000). It contains macro molecular structures such as DNA, RNA, glycopeptide, hormone, enzyme, and protein provides molecular imaging with Jmol. The PDB was founded in 1971 and it is now managed by RCSB (Research Collaboratory for Structural Bioinformatics). The database provides storage of protein structure information obtained by studies such as X-ray crystallography and Nuclear Magnetic Resonance

(NMR). The PDB database provides the essential information about the 3-dimensional structure of the protein studied, the function of the protein, and the relevant source reference, and gives access to data such as the FASTA sequence. Furthermore, these data are used to determine the 3-dimensional structure of the protein together with different *in silico* tools (Berman et al., 2000).

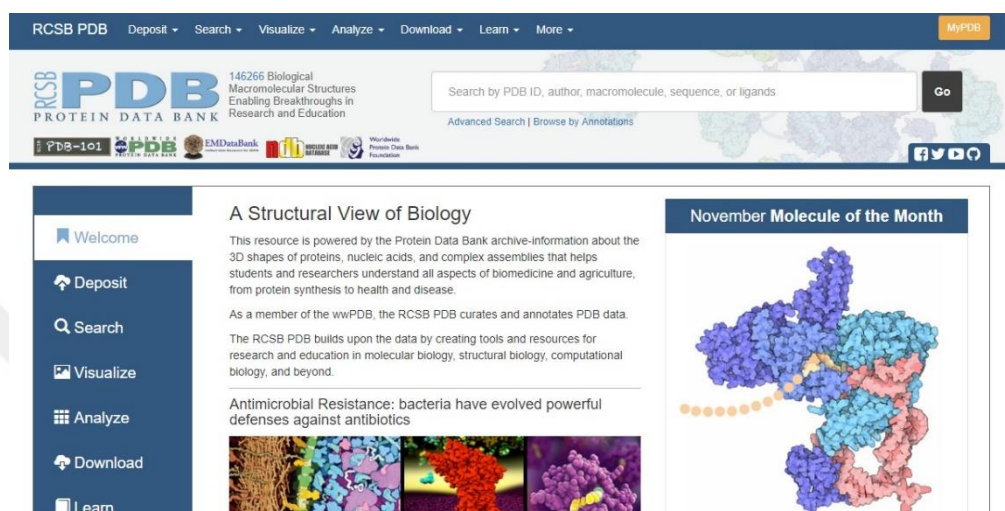


Figure 1.7 Screen shot of Protein Data Bank home page (Retrieval date: 25.03.2019)

1.3.7 NCBI Protein Databases

The NCBI database is a protein database established by NCBI and accessible via the Entrez access system (NCBI, 2019). The database also includes data from databases such as PDB, PIR, GenBank (Madden, 2013).

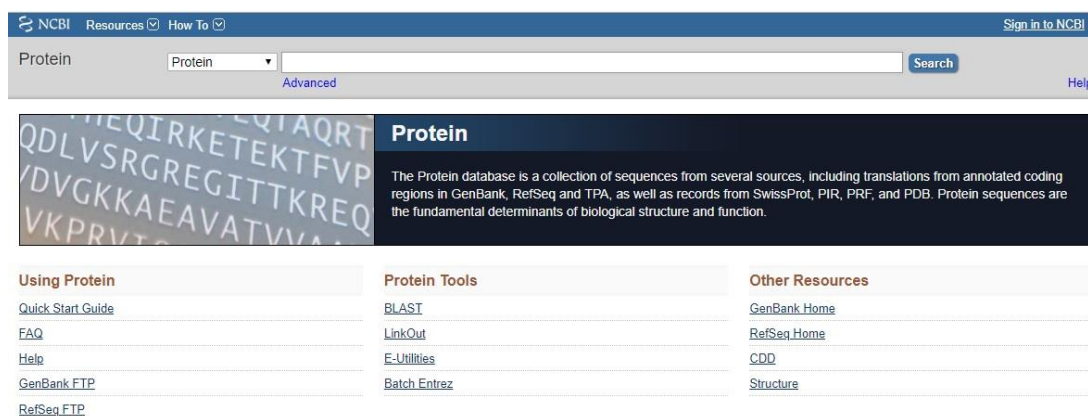


Figure 1.8 Screen shot of NCBI protein database home page (Retrieval date: 25.03.2019)

1.3.8 EMBL-EBI

In 1980, the European Molecular Biology Laboratory (EMBL) in Heidelberg founded the EMBL data library. This is the first nucleotide database to be established. The aim of this data source was also to promote standardization and free data exchange. As in the PDB, transfer the data to the researchers and import their data was important. Therefore, the UK research groups planned to be collected and transferred periodically nucleic acid data. In addition, nucleotide and protein databases provides access to several *in silico* tools such as sequence similarity research tools (FASTA, BLAST), multiple sequence analysis tools (CLUSTAL, MUSCLE) since 2004 (EMBL, 2019) (Figure 1.9) (McWilliam et al., 2013).

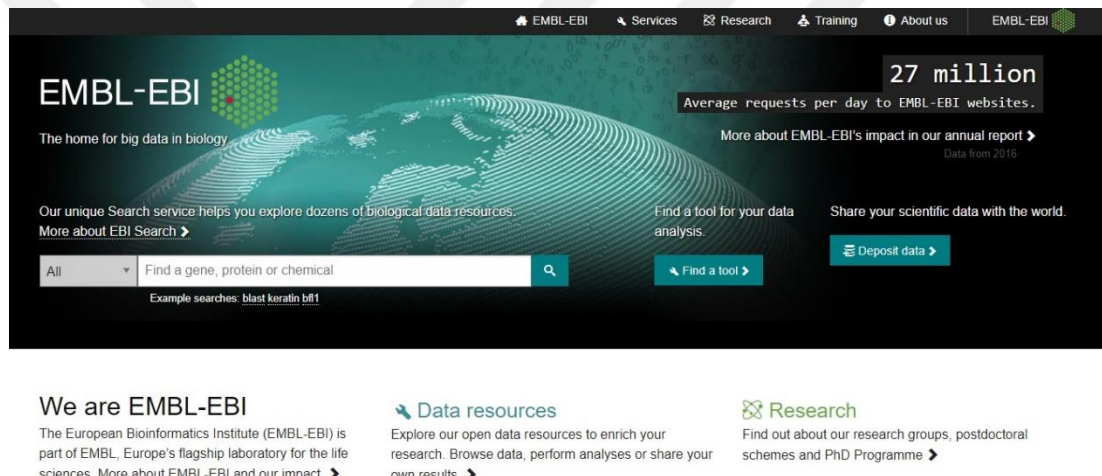


Figure 1.9 Screen shot of EMBL-EBI database home page (Retrieval date: 25.03.2019)

1.3.9 Allergen Nomenclature

The World Health Organization and the International Union of Immunological Societies (WHO / IUIS) have constituted a site for the systematic allergen nomenclature approved by the Allergy Naming Subcommittee. The systematic nomenclature of the allergen is based on the Linnaean system and applied to all allergens. The committee created an allergen database containing approved and formally recognized allergens. The database can be accessed at allergen.org web page.

Research can be carried out according to the name of the allergens or the source of the allergen in this database. The results can be filtered by allergen type (food allergens, contact allergens, airborne allergens etc.). This site also gives some details about the selected allergen. Information such as the source of the allergen, the species to which the allergen belongs and the reference belonging to the allergen. GenBank and UniProt codes of the selected allergen are available. With these codes, the sequence of the selected protein can be reached.



Figure 1.10 Screen shot of Allergen Nomenclature database home page (Retrieval date: 25.03.2019)

1.3.10 Basic Local Alignment Search Tool

BLAST (Basic Local Alignment Search Tool) that accessible via the NCBI Entrez system is the most common local sequence comparison tool available for protein and nucleotide sequences (BLAST, 2019).

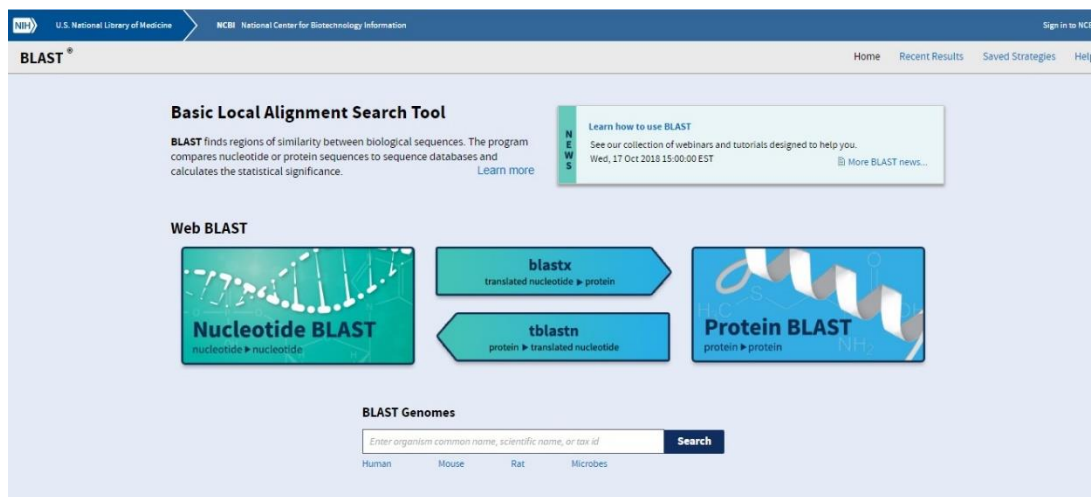


Figure 1.11 Screen shot of BLAST program home page (Retrieval date: 25.03.2019)

It reveals the similarity between the studied sequence fragment (query) and the protein or nucleotide sequence in the database (Madden, 2013). The BLAST tool obtains results as quickly as possible and calculates a value that indicates the reliability of the tool. As the name suggests, the BLAST "local" performs alignments. There are many different types of BLAST searches. Nowadays, BLAST searches have taken interests of scientist while mega BLASTN is used for nucleotide-nucleotide search, BLASTP is used for protein-protein matching, BLASTX is used for protein databases to nucleotide matching TBLASTN explores a protein sequence against the nucleotide database and instantly translates the database (Madden, 2013).

1.3.11 FASTA

FASTA is an *in silico* tool used for local and global sequence similarity research. Pearson and Lipman developed it in 1988. It uses Smith-Waterman algorithm to access FASTA databases and to perform local and global comparison between protein and nucleotide sequences. FASTA program package contains many programs and it can be accessed to program from various internet sources (FASTA, 2019). In addition to the FASTA format developed for the program, other formats can also be used (McWilliam et al., 2013; Pearson, 2013).

The FASTA format is a text-based format for use in FASTA programs. The first line ">" begins with a greater than sign. The abbreviation followed by the sign

indicates the database in which the sequence is located and refers to the ID number. For example, >tr|A9ZLU7 sequence is in the UniProtKB / TrEMBL database and shows the ID number in this database. Then it includes the name of the protein or nucleotide and the name of the organism to which the sequence belongs and the name of the gene that makes the synthesis. As of the bottom line, the protein and gene sequences are indicated by the single-letter coding system for amino acids or nucleotides (Figure 1.12). The FASTA format is also used in other *in silico* tools such as BLAST and Clustal ω (Lesk, 2005; Pearson, 2013).

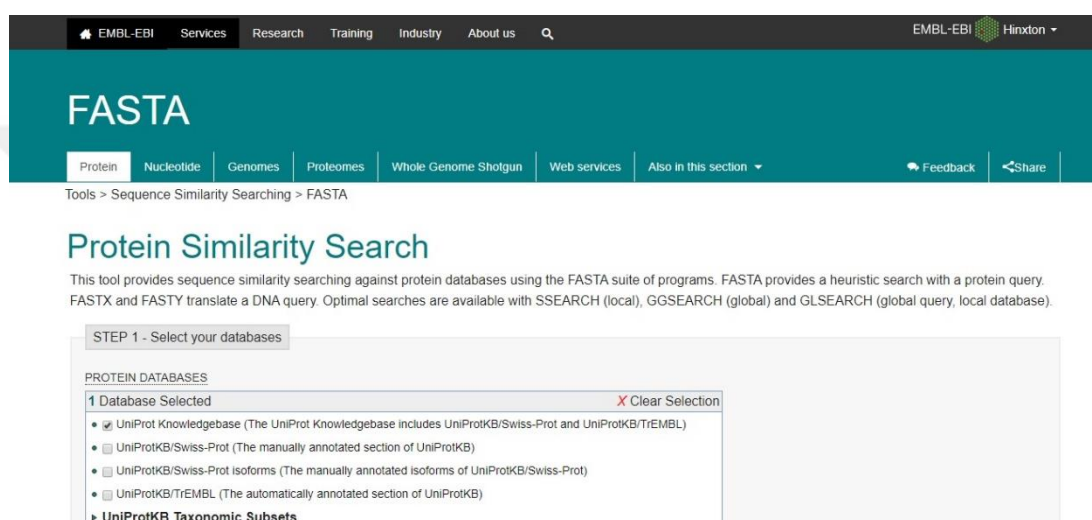


Figure 1.12 Screen shot of FASTA program home page on the EMBL-EBI site (Retrieval date: 25.03.2019)

1.3.12 Clustal ω

Clustal ω was developed in 1994. Clustal ω is a tool used to compare different protein sequences. Also, Cladograms or Phylograms programs help to understand the evolutionary relationships between different species. It is widely used for multiple comparisons of protein and DNA sequences and it has performed at least three or more sequence comparisons (Clustal w, 2019).

Figure 1.13 Screen shot of Clustal ω home page (Retrieval date: 25.03.2019)

In order to understand the consequences of programs need to know certain abbreviations. These abbreviations are shown in Table 1.3. The symbols described in the table are shown for all comparisons, so * (asterisk) symbols have the same meaning for all comparisons, while other symbols for DNA / RNA alignments should not be considered (Larkin et al., 2007).

Table 1.1 The meaning of Clustal ω abbreviations (Larkin et al., 2007)

Symbol	Definition	Meaning
*	asterisk	positions that have a single and fully conserved residue
:	colon	conservation between groups of strongly similar properties with a score greater than .5 on the PAM 250 matrix
.	point	conservation between groups of weakly similar properties with a score less than or equal to .5 on the PAM 250 matrix

Clustal ω allows results to be displayed in colour. Thus, results are clearer for researchers. The meaning of colours is given in the table below:

Table 1.2 The meaning of colours in Clustal ω (Larkin et al., 2007)

Residue	Colour	Property
AVFPMILW	RED	Small (small+ hydrophobic (incl.aromatic -Y))
DE	BLUE	Acidic
RK	MAGENTA	Basic – H
STYHCNGQ	GREEN	Hydroxyl + sulfhydryl + amine + G
Others	Grey	Unusual amino/imino acids etc

1.3.13 AlgPred

AlgPred (Prediction of allergenic proteins and mapping of IgE epitopes) is a systematic work that integrates various approaches. It was established for accurately predict allergic proteins. Detailed information about this tool is available on the website (Algpred, 2019).

AlgPred: Prediction of Allergenic Proteins and Mapping of IgE Epitopes

Introduction [Mirror site at UAMS](#)

The prediction of allergenic proteins is becoming very important in present time due to use of modified proteins in foods (genetically modified foods), therapeutics, bio-pharmaceuticals etc. World Health Organization (WHO) and Food and Agriculture Organization (FAO) realize the importance of prediction and proposed guidelines to assess the potential allergenicity of proteins. In past, number of approaches and methods has been developed to predict allergens; each has their own merits and demerits. In AlgPred a systematic attempt has been made to integrate various approaches in order to predict allergenic proteins with high accuracy.

The salient features of AlgPred server are,

- AlgPred allows prediction of allergens based on similarity of known epitope with any region of protein.
- The mapping of IgE epitope(s) feature of server allows user to locate the position of epitope in their protein.
- Server search MEME/MAST allergen motifs using MAST and assign a protein allergen if it have any motif.
- Allows to predict allergens based on SVM modules using amino acid or dipeptide composition.
- It facilitates BLAST search against 2890 allergen-representative peptides (ARPs) obtained from Bjorklund et al 2005 and assign a protein allergen if it have a BLAST hit.
- Hybrid option of server allows to predict allergen using combined approach (SVMc + IgE epitope + ARPs BLAST + MAST).

World Health Organization (WHO) and Food and Agriculture Organization (FAO) proposed guidelines to assess the potential allergenicity of protein are available from <http://www.fao.org/es/FSN/food/pdf/allergygm.pdf>.

Please cite following paper if you are using AlgPred:
 Saha, S. and Raghava, G.P.S. (2006) AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. Nucleic Acids Research, Volume 34, W202-W209.

Abstract:
http://nar.oxfordjournals.org/cgi/content/abstract/34/suppl_2/W202?ijkey=FvK5uzLHi3Xygdq&keytype=ref

Figure 1.14 Screen shot of AlgPred Home Page (Retrieval date: 25.03.2019)

1.3.14 SWISS-MODEL

The SWISS-MODEL is another tool available in the ExPASy database. Homology modelling is an important technique that facilitates understanding of structural biology. This tool can be used to estimate the modelling with high accuracy using FASTA formats. This user-friendly tool simplifies the process and makes it easier homology modelling. It also allows users who do not have a specific calculation expertise to produce reliable protein models and obtain modelling results, visualize and interpret them. The general structure of the complex, starting from the amino acid sequences of proteins loaded into the system, is understood by homology modelling. The three-dimensional structures of proteins allow the understanding of their functions at the molecular level.

Swiss model reveals 3D structures of proteins by comparative modelling. The following steps are followed when creating a 3D structure in the SWISS-MODEL:

1. Input data: the sequence of the target protein can be entered as the FASTA form or the UniProt access code can be written. If the target protein consists of different subunits, the sequences of each subunit should be specified.
2. Template search: The data provided in the first step is evaluated. The SWISS-MODEL template searches for the template of the protein that is evolutionarily associated with the target protein. These searches are performed using two methods. The first is BLAST, which has fast and sufficient accuracy for templates closely related to the target template. The second is the HHblits method, which gives more precise results than other methods in the case of remote homology with the target template.
3. Template selection: In order to model a sequence, first of all, a search on similar template is carried out. As Biasini et al. (2013) mentioned that the search must full fil some criteria as predicted by the Global Model Quality Estimation (GMQE) and Quaternary Structure Quality Estimation (QSQE). (Bertoni, Kiefer, Biasini, Bordoli, & Schwede, 2017). The templates in the top row are compared.
4. Modelling: In order to create a 3D protein model, the atom coordinates must be transferred in a suitable orientation based template model found in 3rd step of

the strategy. For creation of the model, side chain of amino acids is of great importance. SWISS-MODEL creates OpenStructure calculations within the framework of structural biology (Schwede et al., 2013) and it is based on the ProMod3 modelling engine.

5. Model quality estimation: The QMEAN scoring function, which measures modelling errors and provides estimates of expected model accuracy, is used in SWISS-MODEL, so that the quality of the selected model is estimated (Benkert, Biasini, & Schwede, 2011). In order to produce quality estimates, QMEAN uses the statistical potentials of the average force.

Figure 1.15 Screen shot of SWISS-MODEL home Page (Retrieval date: 25.03.2019)

1.3.15 ProtParam

ProtParam is one of the tools on the ExPASy web site. ProtParam (Protparam, 2019) calculates the various physical and chemical properties that can be extracted from a protein sequence. In protein identification and research, protein mass spectrometry and 2-dimensional gel research have central importance. The results received from experimental studies and the matching of the recorded results in the database are used for protein identification.

Additional information about the proteins examined in ProtParam tool is not necessary. In this tool, the protein to be examined can be identified as the Swiss-Prot

/ TrEMBL accession number or ID or the raw sequence, and the white spaces and numbers within the sequence are ignored.

The parameters such as molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, indecision index, aliphatic index and large hydropathicity (GRAVY) are calculated by ProtParam. The details calculators can be found in ExPASy web site (Gill & Von Hippel 1989; Ciechanover & Schwartz, 1989; Guruprasad, Reddy & Pandit, 1990; Kyte & Doolittle, 1982; Ikai, 1980).

1.3.16 BIOPEP

BIOPEP-UWM (UWM, 2019) was developed by the Department of Food Biochemistry at the University of Warmia and Mazury in Olsztyn, Poland. BIOPEP includes the following databases:

- Sequence databases (proteins, bioactive peptides, allergic proteins and epitopes, sensor peptides)
- The tools that also comprise proteolytic processes for analysing bioactive peptides within proteins

Proteins contained in BIOPEP-UWM data banks are stored in coded form FASTA. However, there are special codes in the storage process. For example, the ~ symbol indicates that the peptide sequence has an amide group at the C-terminus. This kind of peptides are situated in BIOPEP-UWM twice as a precursor and mature protein. All sequences in the BIOPEP-UWM database are based on the literature. The proteins are with Swiss-Prot access number in the system. At the same time, the EC₅₀ values for the bioactivity of the protein of interest are also presented to the users if present in the literature. The proteolytic enzyme list of BIOPEP (Figure 1.16) is shown below.

Databases	Current number
Proteins	740
Bioactive peptides	3681
Allergenic proteins with their epitopes	135
Sensory peptides and amino acids	480

Submit new peptide sequence

Login

Please cite the following paper if you are using BIOPEP-UWM database:
Minkiewicz P., Dziuba J., Iwaniak A., Dziuba M., Darewicz M., BIOPEP database and other programs for processing bioactive peptide sequences.
Journal of AOAC International, 91, 2008, 965-980.

Figure 1.16 Screen shot of proteolytic enzyme list in BIOPEP (Retrieval date: 25.03.2019)

BIOPEP-UWM values of bioactive peptides can be obtained from their above-mentioned web sites.

The BIOPEP-UWM system also contains important information about *in silico* proteolytic degrading enzymes. In the system, a FASTA-formatted protein sequence is cleaved by selected proteolytic enzymes / enzymes. The resulting fragments are compared with defined bioactive fragments by the database in this ways the scores are obtained.

BIOPEP values based equations can be found from following papers (Dziuba et al., 2009; Kandemir-Cavas, Pérez-Sanchez, Mert-Ozupek & Cavas, 2019; Minkiewicz et al., 2011; Minkiewicz et al., 2008; Minkiewicz et al., 2011).

1.4 Protein and Protein Structure

Proteins are important building blocks of living organisms. They are consisted of 20 different amino acids. The proteins can be called as biopolymers since amino acids from their monomers. All amino acids are composed of the amine group (NH_2) attached to α -carbon atom in the chiral center, the carboxyl group (COOH), hydrogen and a functional R group in the side chains. In addition, all amino acids are differed according to the R group (Lehninger, 2004). There are 20 different standard amino

acids, each amino acid is abbreviated with three, and single letter symbols (Table 1.3) (Lehninger, 2004). The structures of proteins have covalent bonds and non-covalent bonds. Covalent bonds in the structure of proteins are peptide bonds and disulphide bonds. Non-covalent bonds are hydrogen bonds, ion bonds and hydrophobic bonds (apolar bond). The two amino acid binds with each other via a condensation reaction. A covalent bond is formed between the carboxyl group of an amino acid and the amine group of the other amino acid at the same time water is released. The dipeptide is formed by this reaction and the bond between the two amino acids is called the dipeptide bond. Polypeptides and proteins are formed by repeating the described condensation reaction (Lehninger, 2004).

The protein structures can be categorised under 4 groups; primary, secondary, tertiary and quaternary (Lehninger, 2004) (Figure 1.17).

Table 1.3 Abbreviation list of amino acids

List of Amino Acids					
Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Asparagine	Asn	N	Methionine	Met	M
Aspartic Acid	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamine	Gln	Q	Serine	Ser	S
Glutamic acid	Glu	E	Threonine	Thr	T
Glycine	Gly	G	Tryptophan	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

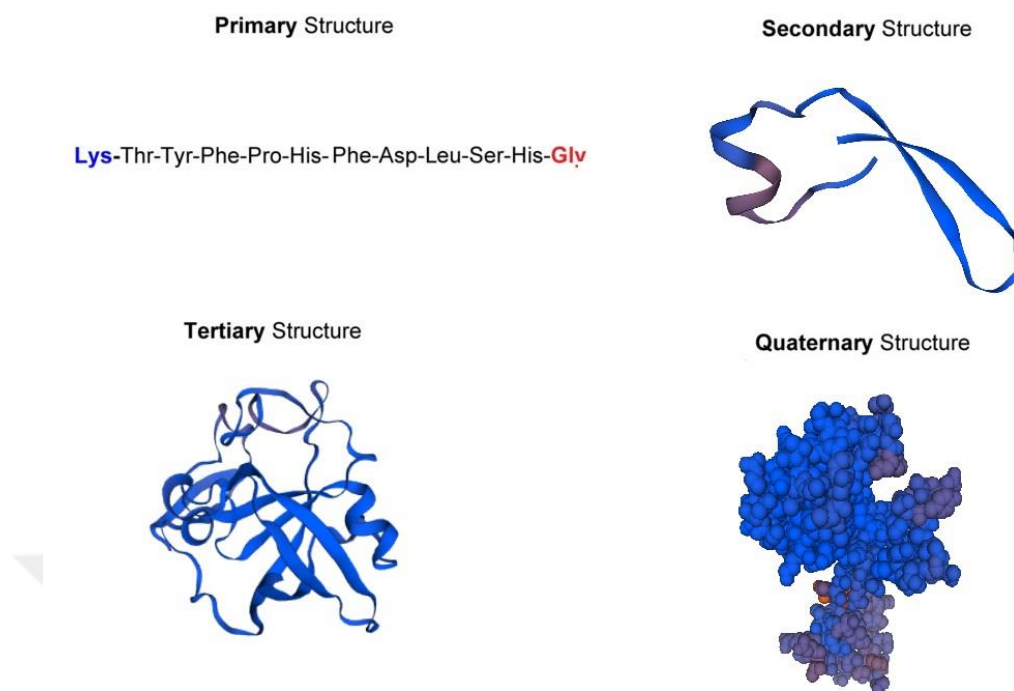


Figure 1.17 The view of the protein structures

1.5 Bioactive Peptides

Bioactive peptides can be defined as short fragments having a positive effect on body (Kitts and Weiler, 2003). Many of the foods taken daily are known to be rich sources of bioactive peptides. Bioactive peptides are found inactive form in the protein sequence and have become active form when proteins hydrolysed. This hydrolysis reaction can be occurring in three ways: the first one, protein can be hydrolysed by digestive enzymes. The second one, it can be hydrolysed by proteolytic enzymes from plant or microorganism. The third one, it can be hydrolysed with the action of microorganisms or various proteolytic enzymes. In addition, bioactive peptides are released during gastrointestinal digestion and during fermentation of foodstuffs with lactic acid bacteria from various food proteins (Korhonen & Pihlanto, 2006). Bioactive peptides, depending on amino acid sequences, can affect the cardiovascular system, the digestive system, the immune system and the nervous system (Figure 1.17). Therefore, it is recommended to take different peptide sequences that improve human health by dieting, by reducing the risk of chronic disease or by strengthening the

natural immune system. Agyei and Danquah (2012) classify bioactive peptide types and the systems they act in the body as follows:

Peptides acting on the cardiovascular system:

- Antithrombotic peptides which prevent blood clotting
- Antihypertensive peptides with blood pressure-lowering effect as a result of their various mechanisms
- Antioxidative peptides, peptides showing hypolipidemic activity

Peptides acting on gastrointestinal systems:

- Mineral Chelation
- Satiety Control
- Antimicrobial
- Opioid activities
- Anticarcinogenicity
- Antigastric

Peptides acting on the immune system:

- Immunomodulatory peptides,
- Antimicrobial peptides
- Cytomodulator peptides

The calcium-binding peptides act on the musculoskeletal system.

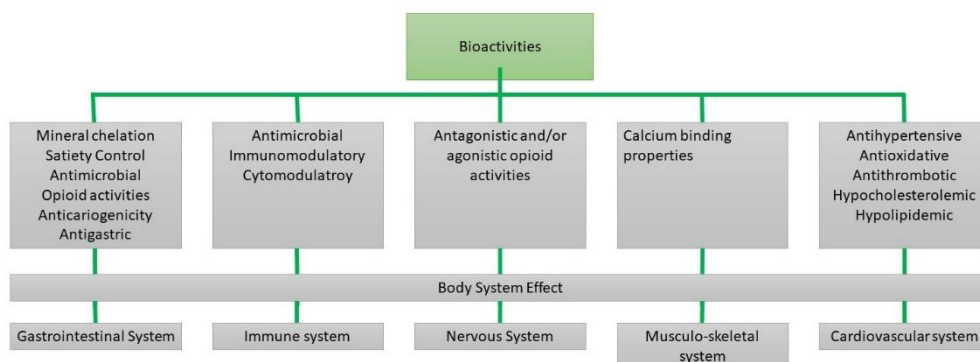


Figure 1.18 Effects of bioactive peptides on the body systems (Inspired from Agyei, & Danquah, 2012)

In addition, there are many bioactive peptides known to be multifunctional, and the size of these bioactive peptides ranges from two to twenty amino acid residues (Meisel & FitzGerald, 2003). Many bioactive peptides such as angiotensin converting enzyme (ACE) inhibitory peptides and calcium binding phosphopeptides (CPPs) are produced by gastrointestinal enzymes such as pepsin and trypsin (Korhonen & Pihlanto, 2006). Most of the bioactive peptides from milk and other food sources show antihypertensive activity, especially angiotensin-1-converting enzyme (ACE) (Korhonen & Pihlanto, 2006).

1.5.1 Bioactive Peptides Investigated in Thesis

1.5.1.1 ACE-inhibitor Bioactive Peptides

Angiotensin-I converting enzyme (ACE, EC 3.4.15.1), which is a member of the zinc metalloprotease family of M2 globulin and was first described by the conversion of angiotensin-I (AngI) to angiotensin-II (AngII), serves as a dipeptidyl carboxypeptidase. In addition, this enzyme catalyses the hydrolysis of carboxyl-terminal dipeptides from a wide variety of functional peptides (Anthony et al., 2012).

The renin-angiotensin system (RAS), which plays an important role in blood pressure regulation, produces a series of bioactive peptides that control events such as fluid and electrolyte homeostasis, hormone release. ACE inhibitors, an important component of the medical treatment of hypertension, are known to play an important role in RAS. Renin plays a key role in RAS regulation. This enzyme is produced in the juxtaglomerular complex. Juxtaglomerular complex formed by the differentiation of smooth muscle cells in the wall of the arteriole, which is into the Bowman capsule, and these cells have flat nuclei, and their cytoplasm is filled with RAS-positive granules. ACE that hydrolyses peptides by the removal of a dipeptide from the C-terminus, in the same manner; it transforms the angiotensin I, which is inactive deca peptide to the angiotensin II that is octa peptide via removing the dipeptide His-Leu (Coates, 2013). Angiotensin II, a potent vasoconstrictor dependent on substrate concentration, binds to type 1 angiotensin II receptor (AT1) (Zhang et al., 2000) (Figure 1.18). ACE also causes the degradation of the bradykinin hormone, which has vasodilation properties (Erdmann, Cheung, & Schröder, 2008).

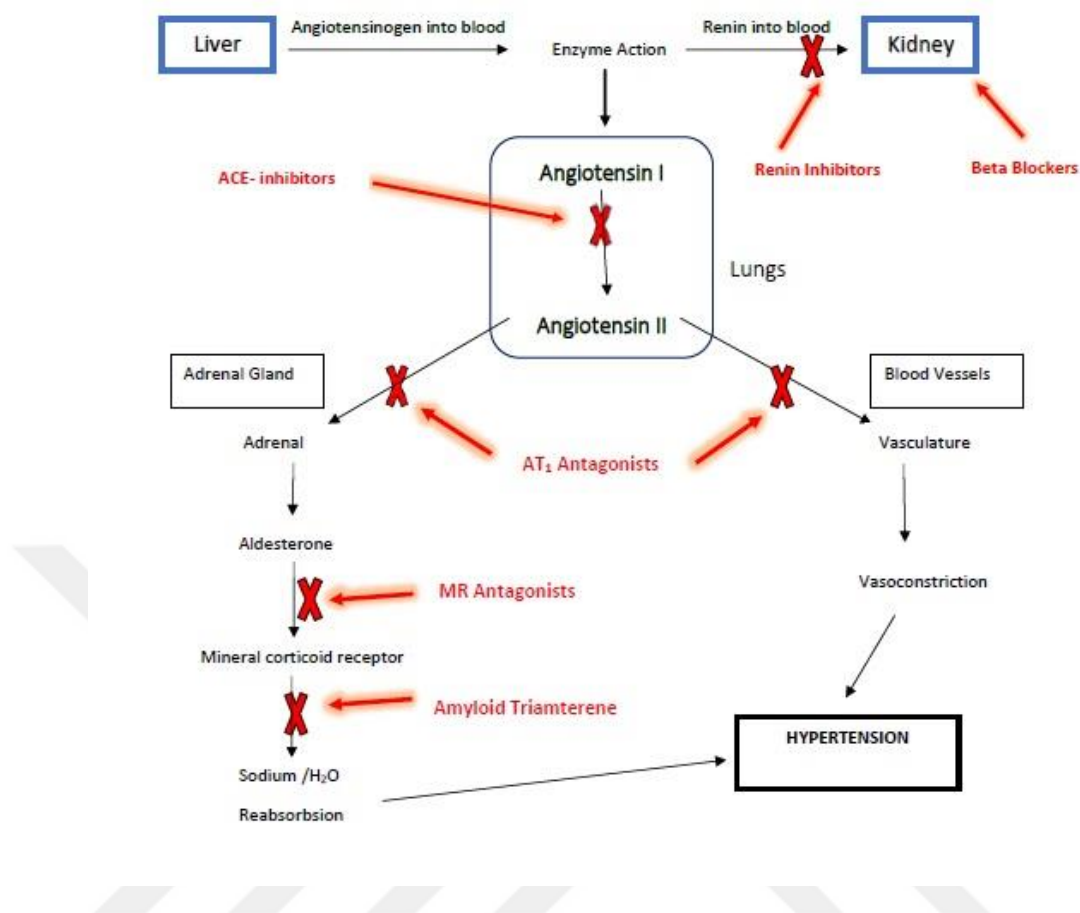


Figure 1.19 The Renin-angiotensin system

In general, the drugs used in the treatment of hypertension contain ACE inhibitors (Captopril, enalapril, fosinopril) (Iwaniak et al., 2014). Research is being conducted to use food-borne ACE inhibitor peptides instead of synthetic drugs with such potential side effects and to examine their efficacy compared to drugs (Iwaniak et al., 2014).

1.5.1.2 Antioxidative Peptides

Oxidative stress is known to play a role in the development of cardiovascular diseases such as arteriosclerosis, Alzheimer's disease, Parkinson, cancer and diabetes (Reuter, Gupta, Chaturvedi & Aggarwal, 2010). Vitamin E, polyphenol compounds, carotenoids as well as bioactive peptides are also involved in antioxidative defence against oxidative stress caused by free radicals such as superoxide anion ($O_2^{\bullet-}$), hydrogen peroxide (H_2O_2), hydroxyl radical ($OH^{\bullet}$), and free radicals (Fang, Yang &

Wu, 2002; Erdmann et al., 2008). Bioactive peptides show antioxidative properties with radical scavenging or extinction, prevention of fatty acid peroxidation, metal ion chelate formation mechanisms (Power, Jakeman & FitzGerald, 2013). Factors such as amino acid composition, hydrophobic structures, molecular size, and structural properties of peptides affect the antioxidative activity of peptides (Udenigwe & Aluko, 2012).

1.5.1.3 Antithrombotic Peptides

Another important cardiovascular disease is thrombosis that is caused by blood clotting because of abnormalities in blood clotting. Thrombosis is triggered with increase of thrombocyte activity and increased of proteins such as fibrinogen, inexact of fibrinolysis, and increased of blood density (Erdmann et al., 2008). The thrombin enzyme, a blood coagulant enzyme, coagulates blood by providing degradation of fibrinogen to fibrin and fibrino peptides. Thrombin-based coagulation is inhibited by antithrombotic peptides. These peptides inhibit both fibrinogen binding and prevent aggregate composition (Phelan & Kerins, 2011). About the antithrombotic peptides, studies have been reported on milk-derived proteins such as casein, lactoferrin and other proteins. (Phelan & Kerins, 2011; Rojas-Ronquillo et al., 2012; Yang, Wang & Xu, 2007).

1.5.1.4 Antiamnesic Peptides

Prolyl endopeptidase (PEP) (EC 3.4.21.26) plays a role in the generation of neurodegenerative diseases such as Alzheimer's, Parkinson's, and Huntington's disease. There is also a high rate of PEP activity in the case of mental disorders such as post-traumatic stress disorder, bipolar disorder, eating disorders and memory loss (Garcia-Horsman et al., 2007; Wilson et al., 2011). PEP inhibition shows antiamnesic and memory-enhancing effect (Garcia-Horsman et al., 2007). Many antiamnesic drug groups are produced according to the PEP inhibitor mechanism. PEP inhibitor peptides found in natural origin are red wine green tea, fish and cheese (Sørensen, Kildal, Stepaniak, Pripp & Sørhaug, 2004; Wilson et al., 2011; Yanai, Suzuki & Sato, 2003; Kim, Kim, & Song, 2001).

1.5.1.5 Dipeptidyl Endopeptidase – IV (DPP – IV) Inhibitor Peptides

According to the WHO (2006) in the 2000 data, in 2030 the number of patients who are predicted to be 171 million to 366 million people have type-2 diabetes. One of the treatment modalities for type-2 diabetes is the inhibition of Dipeptidyl Endopeptidase inhibitor IV (DPP 3.4 IV) (EC 3.4.14.5) (Nongonierma & FitzGerald, 2013). DPP-IV performs proteolysis from the N-terminal region of the dipeptides that have Xaa-Pro and Xaa-Ala structure. Glucose-dependent insulintropic peptide (GIP) and glucagon-like peptide (GLP-1) leading to insulin secretion are inactivated by the incretin hormones DPP-IV. As a result, it reduces insulin secretion (Nongonierma & FitzGerald, 2013). Food-borne DPP-IV inhibitor peptides are found in different food sources such as milk proteins, whey proteins, Gouda type cheese (Nongonierma & FitzGerald, 2013; Uenishi, Kabuki, Seto, Serizawa & Nakajima, 2012).

1.5.1.6 Ubiquitin–Proteasome System (UPS) Activator Peptides

UPS-mediated catabolism is necessary to protect amino acid pools in acute fasting and to contribute to the degradation of defective proteins. This system is associated with neurodegenerative diseases such as Alzheimer's and Parkinson's disease. The failure of protein degradation leads to decrease in proteasome activity. This system also enables the regulation of basic cellular processes such as immune system reactions, cell cycle progression, and degradation of proteins for the cell (Löw, 2011).

1.6 Food Allergy and Allergen Protein

Food allergy (FA), significant public health problem affecting many people especially children, can cause severe allergic reactions and even risk of death. Today there is no effective treatment for FA, just prevention or treatment of symptoms may be solution for food allergy. Furthermore, many problems arise when confused with non-allergic food reactions, such as food intolerance, with FAs, so it is clear that the diagnosis of FA is important. According to National Institute of Allergy and Infectious Diseases (NIAID) guidelines, food allergy, food allergens are defined as follows:

A negative health impact caused by specific immune response that occurs reproducibly when exposed to a particular food is defined as food allergy (Boyce et al., 2010).

Food allergens are defined as components that are recognized by immune cells and cause specific symptoms (Boyce et al., 2010; Sampson et al., 2018).

Many allergens cause allergic reactions, primarily when eaten raw. However, most food allergens may cause the reaction to be cooked food or even after digestion in the stomach and intestine. Non-immune reactions involving metabolic, toxic, pharmacological and undefined mechanisms caused by a particular food are defined as food intolerance (Boyce et al., 2010). Comprehensive data indicate that food allergies are common (approximately 10% of the population), the prevalence rate has increased in the last 2 to 3 years; it has disproportionately affected the people in the industrialized/western regions and is more common in children. If a short list of foods is encountered where adults are exposed to serious allergies, this list creates peanuts, fish, shellfish, eggs, milk, wheat, soya and seeds (Osborne et al., 2017). As in all chronic diseases, a number of factors, including genetic, environmental and epigenetic effects, influences the expression of food allergy. A number of risk factors have been defined to increase food allergy or susceptibility, including invariant risks such as gender (children with male sex), race/ethnicity (The possibility of the appearance of allergies among Asian and black children is higher than white children) and genetic risks (Osborne et al., 2016; Sicherer & Sampson, 2018). When the risks of allergenicity of food allergens are considered how important it is understood that the identification. Food allergy occurs in two ways. The first is non-IgE-mediated food allergy is a reaction involving other components of the immune system. In this type of allergy reactions occur after ingestion of food, and usually cause to symptom such as vomiting, bloating, and diarrhoea. The second one is IgE-mediated allergy occurs when antibodies to allergen. It usually occurs within minutes and it caused many reactions such as redness of the skin, vomiting and anaphylaxis in more severe reactions.

1.6.1 Kiwellin

Kiwi plant grows in temperate conditions and is not frost resistant. The pulp of the kiwifruit is sweet and nutritious. *Actinidia deliciosa* (green kiwi) species are coloured

due to chlorophyll, while the pulp of *Actinidia chinensis* (gold kiwi) is yellow in colour. The general nutritional composition of two kiwi species (gold and green kiwifruit), however, differs in the levels of vitamin C, chlorophyll, lutein, b-carotene and other carotenoids and anthocyanin. The amounts of the substances in the kiwi species were obtained by previous experimental studies (Lim, 2012). There is a lot of research about the content of kiwi fruit that causes allergies, although it is a nutritious food. Green kiwi allergy began to increase after the spread of green kiwi to North America and Europe about thirty years ago. Gold kiwifruit and green kiwi fruit have been found to have in vitro immunoglobulin E (IgE) cross-reactivity (Chen et al., 2006; Lucas et al., 2005). Kiwi fruit is a common IgE-mediated mechanism that causes local mucosal, systemic, or both types of symptoms (Pastorello et al., 1998). Studies have shown that Act c 1, the main allergy of the kiwi fruit, is actinidin, a proteolytic enzyme of the thiol-protease class. In addition, two other 24 and 28 kDa allergens were identified. The amount of kiwellin, which is an allergic protein previously isolated from green kiwi, is higher in gold kiwi (Tuppo et al., 2008). Kiwellin is a modular protein composed of two parts. When Kiwellin performs proteolytic processing in vivo, KiTH and Kissper fragments occur.

1.7 The Aim of Thesis

The aim of this thesis is to identify novel allergen proteins that have not been previously identified by using *in silico* tools and to determine the bioactive peptide in unique defined potentials allergen proteins. Kiwellin protein in kiwi fruit was chosen as the subject of study. The selection of the protein in this study is associated with increased consumption of kiwi in Turkey because of increasing planting kiwi along the northern coastline of Turkey. Allergenicity of kiwellin is known to be high. Identifying allergen proteins similar to kiwellin in different plants will reduce the risk of allergenicity. Although it is known that allergens harm human health and significantly reduce the quality of life, the *in silico* scientific investigations have not been in the objective adequate level.

CHAPTER TWO

MATERIAL AND METHOD

2.1 Retrieval of Allergen Protein Sequences

WHO/IUIS Allergen Nomenclature Sub-Committee was used for finding allergen proteins (Larsen, 2006; Chapman et al., 2007). The sequence of kiwellin in kiwi fruit was retrieved from Allergen.org. Kiwellin protein was selected from the list of allergens existed in Allergen.org (Table 2.1). Kiwellin can be called an important protein in kiwi because of the existing of abundant kiwellin protein in kiwi. The shorter version of kiwellin sequences was chosen from the database because it was thought that it could be more advantageous to use shorter sequence for finding similar proteins in databases. Isoallergens of kiwi fruit are also defined in allergen.org as Act c 5.0101 and Act c 5.0102.

Table 2.1 Kiwi allergens on allergen.org (Retrieved 27.03.19)

<i>Actinidia chinensis (Gold Kiwi Fruit)</i>	
Allergen	Biochemical Name
Act c 1	-
Act c 5	Kiwellin
Act c 8	Pathogenesis-related protein, PR-10, Bet v 1 family member
Act c 10	nsLTP1
<i>Actinidia deliciosa (Green Kiwi fruit)</i>	
Allergen	Biochemical Name
Act d 1	Cysteine protease (actinidin)
Act d 2	Thaumatin-like protein
Act d 3	-
Act d 4	Phytocystatin
Act d 5	Kiwellin
Act d 6	Pectin methylesterase inhibitor

Act d 7	Pectin methylesterase
Act d 8	Pathogenesis-related protein, PR-10, Bet v 1 family member
Act d 9	Profilin
Act d 10	nsLTP1
Act d 11	Major latex protein/ripening-related protein (MLP/RRP), Bet v 1 family member
Act d 12	Cupin, 11S globulin
Act d 13	2S albumin

2.2 Searching Similar Proteins with BLASTp

Basic Local Alignment Search Tool for proteins (BLASTp) search was carried out by BLASTp tool of NCBI (Johnson et al., 2007; Boratyn et al., 2013). Then kiwellin-like protein sequences were screened using BLASTp in NCBI website. BLASTp tool gives percentage identity and E-value. These values represent the similarity between sequences.

2.3 Multiple Sequence Alignment

The CLUSTALW program on the EMBL-EBI site was used to perform multiple sequence analyses of kiwellin-like protein sequences and to create phylogenetic trees (McWilliam et al., 2013). The type of data set to be analysed was determined as "protein". The sequences in FASTA format were uploaded to the site. The analysis was made under default settings.

2.4 Creation of Phylogenetic Tree

Phylogenetic trees were created to understand the evolutionary relationship among the selected 10 potential allergen proteins. Clustal omega was used for this process. After comparing the selected kiwellin-like proteins with Clustal omega, the phylogenetic tree was created from the "phylogenetic tree" section on the site. The website also creates a matrix showing the similarities of each of the 10 different proteins. In the section of result summary, each property calculated by the site can be viewed.

2.5 ProtParam Analysis

Physicochemical properties of the sequences such as individual amino acid number and percentage, molecular weight, theoretical pI values, total number of negatively and positively charged residues and instability index were calculated by ProtParam tool of ExPASy. The analysis of selected 10 potential allergen proteins was performed using this programme. FASTA form of the protein was employed, and the results given in the table.

2.6 Homology Modelling with SWISS-MODEL

SWISS-MODEL was used for homology modelling (Biasini et al., 2014). FASTA forms of proteins were used in this tool. The results can be obtained by clicking on the section on the website "build models". With the 3D structure of the protein on the website, it is possible to know the location and position of each amino acid to be investigated.

2.7 AlgPred Analysis

AlgPred is used for *in silico* allergenicity assessment of the proteins. FASTA form of proteins is sufficient for analysis in AlgPred tool. Simply click on the desired algorithm from the five different algorithms on the home page. Three algorithms have been chosen in this study. Because of the proteins with high similarity to Kiwellin were expected to be allergens, the proteins were evaluated according to amino acids and dipeptide composition. The SVM algorithm was also chosen to measure the allergenicity.

2.8 *In silico* Hydrolysis with BIOPEP

The presence of bioactive peptides was investigated by hydrolysing potential allergen proteins with BIOPEP. The "Bioactive peptides" button was clicked on the home page. "Analyses" section was clicked on the page that was opened and then "enzyme (s) action" was selected. Then "for your sequence" was clicked and the sequence of kiwellin-like proteins was entered into the opened box. All of the enzymes in the following part were selected separately and hydrolysis was performed. The

presence of bioactive peptides was investigated in three protein in *Vitis vinifera*, *Cucumis sativus* *Capsicum chinense*.

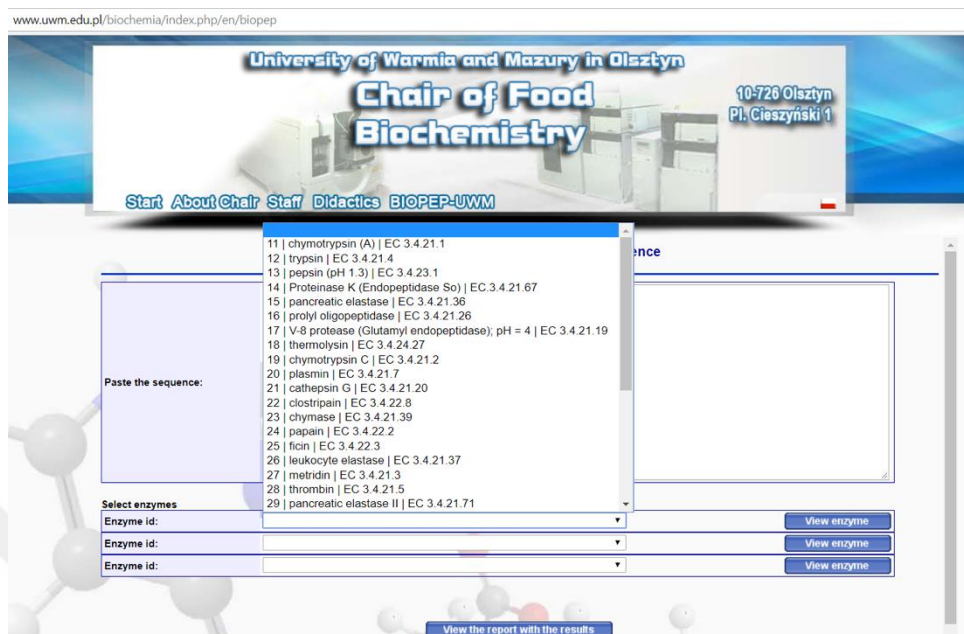


Figure 2.1 Screenshot of the enzymes contained in the BIOPEP tool

CHAPTER THREE

RESULTS

Kiwellin is one of the well-defined allergen proteins in *A. chinensis*. However, kiwellin-like proteins in other fruit have not been characterized yet. In our study, we searched kiwellin-like protein in NCBI database and we found 10 potential proteins, which could be considered as candidate allergen proteins (Table 3.1). Maximum percent identity was found in Barwin-related protein as 95%. On the other hand, minimum identities were observed in the kiwellin like proteins of *C.sativus* and *P.persica* as 80 percentage. E-values were maximum when percentage identity is high and they were minimum when percentage identity is high. These similar proteins were selected from 10 different and commonly used plants.

Table 3.1 Table of potential allergen proteins that is formed according to results obtained with BLASTp

Species	Name of protein (UniProt)	NCBI Reference Sequence	E-value for BLAST T	% identity
<i>Corchorus olitorius</i>	Barwin-related endoglucanase	OMO91533.1	5e-04	95
<i>Vitis vinifera</i>	unnamed protein product, partial (BLASTp)	CBI16343.3	0.007	85
<i>Cucumis sativus</i>	Uncharacterized protein	KGX46853.1	0.16	80
<i>Punica granatum</i>	hypothetical protein CDL15_Pgr026889	OWM73785.1	0.012	90
<i>Capsicum chinense</i>	Ripening-related protein grip22	PHU03889.1	0.012	85
<i>Prunus persica</i>	Receptor-like protein kinase (BLASTp)	XP_020413324. 1	0.15	80
<i>Carica papaya</i>	Kiwellin-like (BLASTp)	XP_021896278. 1	0.059	85
<i>Theobroma cacao</i>	Kiwellin	EOY34420.1	0.013	90
<i>Morus notabilis</i>	Uncharacterized protein	XP_010109489. 1	0.063	85
<i>Jatropha curcas</i>	Uncharacterized protein	KDP44018.1	0.064	85

3.1 The Result of Multiple Sequence Alignment and Phylogenetic Tree

We used Clustal omega to analyse the similarity of kiwellin-like proteins in this study. The results of the similarity of the ten potential allergen proteins are as follows:

CLUSTAL O(1.2.4) multiple sequence alignment

```

C.chinense      MNPDALVKTWNISKFQLISGAREEVRLNMEKIML -----VPPPQNGHYTSSNAIS  51
A.chinensis     -----IS  2
P.persica       -----MVSLLFMFVYLLSLILLINIYLPQPSLALS  31
P.granatum      -----MERVWARTM--ASLSFLLALASLPSPVAIS  29
J.curcas        -----MENT---LL--FLSLLFNI--LLFASTSNAIS  25
V.vinifera      -----MPLLL--SLCLITYT--ISFFPFLAFAIS  24
M.notabilis     -----MARFAFLDV--SLSLIFNILISQSLPSHAIS  29
C.olitorius     -----MAKLIL-FA--SLSFFFVTMISSPLLSTAIS  28
T.cacao         -----MAKTVF-LA--SLSFFFI-LISSPWSNAIS  27
C.sativus       -----MENL---GL--QLFSVILSTIMFFHLCEALS  26
C.papaya        -----MAH-----I--SAFLLFNIMFIFPLANSLS  24
               :*

C.chinense      QCNGPCKSLNDCDGLICIRGKCNDDPDVGTSGGGTS-----  90
A.chinensis     SCNGPCRDLNDCDGLICGTTTSH-----QPGGCKPS-----  34
P.persica       SCNGACATLNDGGLICINRKCNDDPDVGTTHICKKPPP---PPPMTAGGCPNPGTVHC  87
P.granatum      SCNGPCTTLNDCDGLICISGKCTDSDVGTTHICKKSSP---SPTQSSGSCSPSGTLVC  85
J.curcas        SCNGPCQTENDCDGLICIHGSCNDDPDIGTNTCSGSGGGGGGSPSDDSCQPSGSLTC  85
V.vinifera      SCNGPCRFTNDCDGLICINGKCNDDPELGTHVCEGVP----SPPPSGYCHSSGSLRC  79
M.notabilis     SCYGSCRDNLNDCDGLICINGKCNDDTDLGTRICRRT-----PAPGGGCKQYGT LTC  81
C.olitorius     QCNGPCRDLNDCDGLICINGKCNDDPD LGTQICPKSP----T-PTPSGSCQSGTLRC  82
T.cacao         QCNGPCRDLNDCDGLICINGKCNDDPD LGTRICGTPS----S-PSPGTNCNPSGSLRC  81
C.sativus       SCNGPCQTLNDCDGLICINGKCLDDPNLGTHICSGSGGGGGG--GGGHGCHSIGNLNC  84
C.papaya        SCNGPCR TLNDCDGLICINGKCNDDPDVGTTHICSSGGGG-----GGGECKPNGNMF  77
               * * *   * * *   *

C.chinense      -----PSVPPSPSSTGSTPGILTLNDFSEGGDGGGPSECDQYHDNNRNVVALTTRWYAG  145
A.chinensis     -----  34
P.persica       NNVSYPPTYTCSPPVTSSTIAKLDISEFNKGGDGGGAECDEQYHENTERIVGLSTGWYNG  147
P.granatum      KGKSYPKYRCSPPVTPSTRATLTENDFSQGGDGGGPSECDNKYHSNSEHVALSTGC---  142
J.curcas        QGQSYPTYTCSPTVTSSTNAKLTNDFSEGGDGGGAQSECDGNYHDKTELVVALSTGWYAG  145
V.vinifera      KGQYYPPTYTCSPTVTSSTSAILTNDSEGGDGGGPSECDGKYHNNSDPIVALSTGWYNG  139
M.notabilis     DGNTYPTYRCSPPVTSSTQAKLTNDFSEGGDGGGPSECDRYHSNSERIVALSTGWYDG  141
C.olitorius     GGQTYPTYECSPVTSSTKAKLTNDFSEGGDGGGASCDGQYHSNSERVVALSTGNFSN  142
T.cacao        NGQTYPTYICSPPITSSTKARLTNDFSEGGDGGAPSECDRYHSNSERVVALSTGWYAG  141
C.sativus       KGKTFPQFKCSPPVTSSTRAVLTNDFSKGGSGGDPSECDGKFHDNSDPIVALSTGWYNG  144
C.papaya        QGSSHPTFQCSPPVKMANEALLTNNFEPPGGDGGDPSECDGQFHKNSELIVALSTGWYNG  137

```

Figure 3.1 Comparison of 10 potential allergen proteins (The NCBI codes of the proteins can be found in Table 3.1)

Figure 3.1 continues

C.chinense	GSMCGKLIRIRANNNGKSVTAKVWDECDTKDGC-----KNNVVDGSIATWKAALGL	195
A.chinensis	-----	34
P.persica	GKMCGRMIITAG-NGRSVTAKVWDECDSRVGCDAEHAGQPPCENNIVVASEAVWKAALQL	206
P.granatum	-----VAAKVWDECDSMNGCDSEHAGQPPCRNNIVDASSVWHAALGL	184
J.curcas	GTTCGKMIRISAS-NGKSVEAKVWDECDSRNGCDSEHAGQPPCRNNIVDGSDAVWKAALGL	204
V.vinifera	GFRCGKMIKITAD-NGRSVLAKVWDECDSLRGCDSEHSDQPPCKNNIVNGSDAVWKAALGL	198
M.notabilis	GSRCGKMIRITA--NGRSVAKVWDECDSMHGCDAEHAGQPPCKNNIVDGSDAVWKAALGL	199
C.olitorius	GSRCGKMIRIKAS-NGRSVTAKVWDECDSMQGCDEEHAYQPPCKNNIVDGSDAVWKAALGL	201
T.cacao	GSRCGKMITITAS-NGRSVTAKVWDECDSMRGCDSEHAYQPPCKNNIVDGSDAVWKAALGL	200
C.sativus	GSRCGQMIQITAS-NGRSVLAKVWDECDSEINGCDKAHAGLPPCHNNIVDGSDAVWHAALGL	203
C.papaya	GSRCGKIRIRAR-NGKSVAKVWDECDSMNGCDKEHAGLPPCGNNIVDGSNAVWNTLGL	196
C.chinense	DTNLGRVPITWSMA-----	209
A.chinensis	-----	34
P.persica	DVQGEVGVQVWLSLHSCSEAINCGGIGGLSYPFWGVNRS--YCGLPGEVACLDNVPVI	264
P.granatum	DINVGEENITWSMA-----	198
J.curcas	NKDDGVVDVTWSMA-----	218
V.vinifera	DTNLGEVDVTWSEVTDVP----CDSNNYIAFLPWVLLQWQIKFYSCNPFTFS-----	246
M.notabilis	NKDLGVVDVTWSMA-----	213
C.olitorius	NKDVGIQVDTWSMA-----	215
T.cacao	DKNVGIQVDTWSTA-----	214
C.sativus	DIDVGEIPVTWSDA-----	217
C.papaya	NVDDGIQVDTWSMA-----	210
C.chinense	-----	209
A.chinensis	-----	34
P.persica	NMSNINYRILKTNSSRAPPSTVARQDYWKTCPPSFVNTSINSSMFHYTSSGLTNLTLY	324
P.granatum	-----	198
J.curcas	-----	218
V.vinifera	-----SVSSPFHFFHWLCPLVMIHVKHLTT-----VKVSSY	277
M.notabilis	-----	213
C.olitorius	-----	215
T.cacao	-----	214
C.sativus	-----	217
C.papaya	-----	210

Figure 3.1 continues

C.chinense	-----	209
A.chinensis	-----	34
P.persica	YGCNTSTSTTTTPELTSTTQVCSRRNSGNITVSYV--TP---STRVDPVAPGACKNRV	378
P.granatum	-----	198
J.curcas	-----	218
V.vinifera	A---STVSATTIPILGLT---YAEVQGLLPQLMTVSPLAVLSVTGNPILPTHAPLRL	330
M.notabilis	-----	213
C.olitorius	-----	215
T.cacao	-----	214
C.sativus	-----	217
C.papaya	-----	210

C.chinense	-----	209
A.chinensis	-----	34
P.persica	IVPVSTTAAAALETNRTAIQDAVDGGF--EELQIHTDQCNTCTQSGGKCGLDASGGFS	436
P.granatum	-----	198
J.curcas	-----	218
V.vinifera	HPPPALFLLRTIL-VREGMVEAHPVMTNTMKIQTLSLPCQLVGIMVDPCAAR-----	382
M.notabilis	-----	213
C.olitorius	-----	215
T.cacao	-----	214
C.sativus	-----	217
C.papaya	-----	210

C.chinense	-----	209
A.chinensis	-----	34
P.persica	CFQQDNRAYATRCNITTNISNPPPPRPKGSNSKAKYEIGISAGVSGIFIIIFCIVCISRK	496
P.granatum	-----	198
J.curcas	-----	218
V.vinifera	-----	382
M.notabilis	-----	213
C.olitorius	-----	215
T.cacao	-----	214
C.sativus	-----	217
C.papaya	-----	210

Figure 3.1 continues

C.chinense	-----	209
A.chinensis	-----	34
P.persica	RISGFFKKDIRDEFDVEAFIRMYGSLTPKRYSYADVKKLTDSFKDKIGKGGFGTVYKGRL	556
P.granatum	-----	198
J.curcas	-----	218
V.vinifera	-----	382
M.notabilis	-----	213
C.olitorius	-----	215
T.cacao	-----	214
C.sativus	-----	217
C.papaya	-----	210

C.chinense	-----	209
A.chinensis	-----	34
P.persica	PDGLVVAVKVLSESNAGNAEDFINEVASIGRTSHVNIIVTLSGFCYERDKRALIYEFMPNGS	616
P.granatum	-----	198
J.curcas	-----	218
V.vinifera	-----	382
M.notabilis	-----	213
C.olitorius	-----	215
T.cacao	-----	214
C.sativus	-----	217
C.papaya	-----	210

C.chinense	-----	209
A.chinensis	-----	34
P.persica	LDKFIQKQGSEAKCRLEWKTLSIEIAVGIARGLEYLHRCNTRILHLDIKPQNILLDKDFC	676
P.granatum	-----	198
J.curcas	-----	218
V.vinifera	-----	382
M.notabilis	-----	213
C.olitorius	-----	215
T.cacao	-----	214
C.sativus	-----	217
C.papaya	-----	210

Figure 3.1 continues

C.chinense	-----	209
A.chinensis	-----	34
P.persica	PKIADFGLAKLCKPKESIVSMMGTRGTAGYIAPEVFNRFGGVSHKSDVYSYGMVLVLEMV	736
P.granatum	-----	198
J.curcas	-----	218
V.vinifera	-----	382
M.notabilis	-----	213
C.olitorius	-----	215
T.cacao	-----	214
C.sativus	-----	217
C.papaya	-----	210

C.chinense	-----	209
A.chinensis	-----	34
P.persica	GARKNLDSEASHTSELFPHYVYKOLELDNDENVFGAITEEEKIARKMVIISMWCIQTNP	796
P.granatum	-----	198
J.curcas	-----	218
V.vinifera	-----	382
M.notabilis	-----	213
C.olitorius	-----	215
T.cacao	-----	214
C.sativus	-----	217
C.papaya	-----	210

C.chinense	-----	209
A.chinensis	-----	34
P.persica	SNRPSMSKVVEMLEGPLHTLKIAPQPFLFSPAQGSMTTSQPAEIDTEAILFHEQISGN	854
P.granatum	-----	198
J.curcas	-----	218
V.vinifera	-----	382
M.notabilis	-----	213
C.olitorius	-----	215
T.cacao	-----	214
C.sativus	-----	217
C.papaya	-----	210

First amino acid of kiwellin isoleucine (I) was same in all candidate proteins except for *Prunus persica*, *Cucumis sativus*, and *Carica papaya*. In these species, Isoleucine was substituted with Leucine. It was very interesting to note that the amino acids at the position of 2, 4, 6, 8, 12-14, 16, 18-20 were same in all studied sequences. It could be said that these regions must have been conserved. Third amino acid serine (S) of kiwellin was same in all candidate proteins except for *C. chinense*, *C. olitorius*, and *Theobroma cacao*. Serine is substituted with Glutamine (Q). 17th amino acid glutamine (Q) of kiwellin is Glutamic acid (E) in *V. vinifera*. The amino acid at the position of 25 is serine in kiwellin of *A. chinensis*; however, it is cysteine (C) in other studied sequences. Similarly, the amino acid at the position of 25 is Serine in kiwellin of *A. chinensis* it is cysteine(C) in all studied samples. According to Figure 3.2, a phylogenetic tree was constructed based on multiple sequence alignment by Clustal

omega. Phylogenetic tree reveals that the sequences can be classified under three clusters. Kiwellin of *A. chinensis* took place in the first cluster with receptor-like protein kinase of *P. persica*, hypothetical protein CDL 15_Pgr026889 of *P. granatum*, uncharacterized protein of *Jatropha curcas*, and ripening-related protein grip22 of *C. chinense*. From Figure 3.2, it could be said that the proteins in *P. persica*, *J. curcas*, *P. granatum* and *C. chinense* are more close to *A. chinensis*. These four proteins have not been mentioned in allergen.org yet.

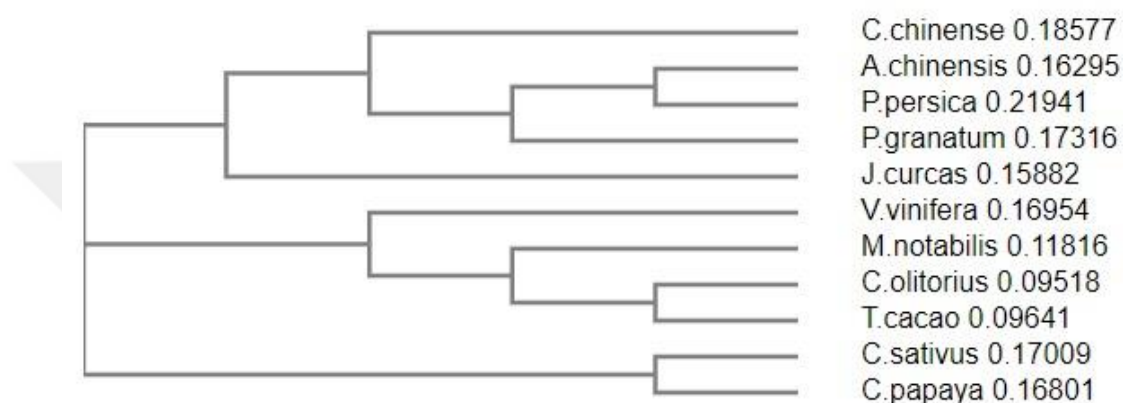


Figure 3.2 Phylogenetic tree of the 10 potential allergen protein sequences (The NCBI codes of the proteins can be found in Table 3.1)

3.2 The Results from ProtParam

ProtParam tool is used to characterize the physicochemical properties of proteins. The ProtParam tool is available under expasy.ch developed by Swiss Bioinformatics Institute. By using this tool one can obtain the parameters such as amino acid length, molecular weight, theoretical pI values, negatively and positively charged residues, net charges and instability index (Table 3.2), number and percentage of amino acids (Table 3.3) in the studied samples. Kiwellin (P85261 (UniProt), Act c 5 (Allergen.org)) in *Actinidia chinensis* (Gold Kiwi Fruit) was selected as a model allergen protein in this study. There is also one more kiwellin (P84527 (UniProt), Act d 5 (allergen.org)) in *Actinidia deliciosa* in allergen.org. We selected P85261 instead of P84527 to find more matched candidate proteins in BLASTp search. Because the length of P85261 is shorter than that of P84527. Therefore, the ProtParam parameters of P85261 in Table

3.2 and table 3.3 are quite different compared to other studied proteins because of its length. On the other hand, there is a problem regarding sequence of P85261. Although it is two separated fragments, it seems like it is consisted of just one fragment (Figure 3.3).

Table 3.2 Protein parameters of potential allergen protein (The NCBI codes of the proteins can be found in Table 3.1)

Species	#aa	Mw (KDa)	Theoretical pI	# Negatively Charged Residues (Asp + Glu)	# Positively Charged Residues (Arg + Lys)	Net Charge	Instability Index
<i>Actinidia chinensis</i>	34	3501.86	5.98	3	2	-1	34.23
<i>Corchorus olitorius</i>	215	22690.36	4.82	22	16	-8	39.49
<i>Vitis vinifera</i>	382	41391.41	5.18	33	20	-13	44.58
<i>Theobroma cacao</i>	214	22729.20	4.72	23	16	-7	38.59
<i>Punica granatum</i>	198	20615.93	4.95	21	13	-8	62.58
<i>Capsicum chinense</i>	209	22148.87	5.18	23	19	-4	36.94
<i>Prunus persica</i>	854	92694.39	6.25	84	78	-6	79.79
<i>Carica papaya</i>	210	21995.70	5.06	20	13	-7	38.89
<i>Jatropha curcas</i>	218	22461.42	4.12	28	10	-18	47.05
<i>Morus notabilis</i>	213	22664.24	4.98	25	18	-7	48.58
<i>Cucumis sativus</i>	217	22343.93	5.00	21	11	-10	73.23

According to Table 3.2, the maximum and minimum number of amino acids were found in *P.persica* and *A.chinensis*, respectively. Theoretical pI value of kiwellin in *A.chinensis* is 5.98 and it was found as 5.83 (data not shown). From this comparison, it could be said that theoretical pI values cannot be affected by sequence length. The maximum pI value was found in *P.persica* as 6.25 and minimum value was found from *J.curcas* as 4.12. In Table 3.2, the net charges were calculated from the subtraction of negatively charged residues from positively charged residues. The net charges of P85261 and P84527 were found as -1 and -4, respectively, which show that kiwellin is a negatively charged protein. The highest and lowest instability index were found in

the proteins of *P.persica* and *A.chinensis*. The correlation test was carried out for the data in Table 3.2. However, no meaningful correlation was found among the data. When the amino acid numbers and percentages were compared, it was found that glycine and serine were dominant amino acids in the sequences apart from the protein of *V.vinifera* in Table 3.3.

```

CLUSTAL O(1.2.4) multiple sequence alignment

sp|P84527|KIWEL_ACTDE      MAQLALLLLSLFLTLISLAPPGASISSCNGPCRDLNDCDGLICIKGKCNDPQVGTHIC 60
sp|P85261|KIWEL_ACTCH      -----ISSCNGPCRDLNDCDGLICG----- 21
                             *****

sp|P84527|KIWEL_ACTDE      RGTTPSPQPGGCKPSGTLTCRGKSYPTYDCSPPVTSSTPAKLTNDFSEGGDDGGPSECD 120
sp|P85261|KIWEL_ACTCH      --TTHSHQPGGCKPS----- 34
                             ** *

sp|P84527|KIWEL_ACTDE      ESYHNNNERIVALSTGWYNGGSRGCKMIRITASNGKSVSAKVWDECDSRHGCDKEHAGQP 180
sp|P85261|KIWEL_ACTCH      ----- 34

sp|P84527|KIWEL_ACTDE      PCRNINVDGSNAVWSALGLDKNVGVVDITWSMA 213
sp|P85261|KIWEL_ACTCH      ----- 34

```

Figure 3.3 Pairwise sequence alignment of P85261 and P84527

Table 3.3 Amino acid number and percentages in potential allergen proteins from different plants, part 1 (The NCBI codes of the proteins can be found in Table 3.1)

	<i>Actinidia chinensis</i>		<i>Corchorus olitorius</i>		<i>Vitis vinifera</i>		<i>Cucumis sativus</i>		<i>Punica granatum</i>		<i>Capsicum chinense</i>	
	#	%	#	%	#	%	#	%	#	%	#	%
Ala (A)	0	0.0	12	5.6	18	4.7	10	4.6	13	6.6	10	4.8
Arg (R)	1	2.9	6	2.8	9	2.4	3	1.4	5	2.5	8	3.8
Asn (N)	2	5.9	12	5.6	19	5.0	15	6.9	10	5.1	17	8.1
Asp (D)	3	8.8	16	7.4	19	5.0	16	7.4	13	6.6	15	7.2
Cys (C)	5	14.7	14	6.5	20	5.2	15	6.9	14	7.1	10	4.8
Gln (Q)	2	5.9	9	4.2	9	2.4	7	3.2	4	2.0	5	2.4
Glu (E)	0	0.0	6	2.8	14	3.7	5	2.3	8	4.0	8	3.8
Gly (G)	5	14.7	22	10.2	28	7.3	34	15.7	17	8.6	24	11.5
His (H)	2	5.9	2	0.9	11	2.9	8	3.7	5	2.5	2	1.0
Ile (I)	2	5.9	10	4.7	20	5.2	13	6.0	7	3.5	13	6.2
Leu (L)	2	5.9	13	6.0	40	10.5	17	7.8	13	6.6	13	6.2
Lys (K)	1	2.9	10	4.7	11	2.9	8	3.7	8	4.0	11	5.3
Met(M)	0	0.0	5	2.3	8	2.1	3	1.4	4	2.0	5	2.4
Phe (F)	0	0.0	6	2.8	15	3.9	7	3.2	2	1.0	2	1.0
Pro (P)	3	8.8	12	5.6	32	8.4	10	4.6	13	6.6	13	6.2
Ser (S)	4	11.8	26	12.1	37	9.7	21	9.7	30	15.2	19	9.1
Thr (T)	2	5.9	15	7.0	27	7.1	10	4.6	13	6.6	13	6.2
Trp (W)	0	0.0	3	1.4	6	1.6	3	1.4	2	1.0	4	1.9
Tyr (Y)	0	0.0	4	1.9	11	2.9	1	0.5	4	2.0	3	1.4
Val (V)	0	0.0	12	5.6	28	7.3	11	5.1	13	6.6	14	6.7
Pyl (O)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0

Table 3.4 Amino acid number and percentages in potential allergen proteins from different fruits, part 2 (The NCBI codes of the proteins can be found in Table 3.1)

	<i>Prunus persica</i>		<i>Carica papaya</i>		<i>Theobroma cacao</i>		<i>Morus notabilis</i>		<i>Jatropha curcas</i>	
	#	%	#	%	#	%	#	%	#	%
Ala (A)	53	6.2	11	5.2	13	6.1	13	6.1	12	5.5
Arg (R)	35	4.1	5	2.4	10	4.7	11	5.2	4	1.8
Asn (N)	52	6.1	21	10.0	15	7.0	12	5.6	15	6.9
Asp (D)	40	4.7	13	6.2	17	7.9	19	8.9	19	8.7
Cys (C)	36	4.2	14	6.7	14	6.5	14	6.6	14	6.4
Gln (Q)	22	2.6	4	1.9	5	2.3	5	2.3	7	3.2
Glu (E)	44	5.2	7	3.3	6	2.8	6	2.8	9	4.1
Gly (G)	72	8.4	30	14.3	21	9.8	24	11.3	29	13.3
His (H)	15	1.8	5	2.4	2	0.9	4	1.9	3	1.4
Ile (I)	60	7.0	11	5.2	12	5.6	11	5.2	8	3.7
Leu (L)	58	6.8	13	6.2	12	5.6	14	6.6	15	6.9
Lys (K)	43	5.0	8	3.8	6	2.8	7	3.3	6	2.8
Met(M)	17	2.0	6	2.9	3	1.4	4	1.9	3	1.4
Phe (F)	31	3.6	10	4.8	5	2.3	4	1.9	4	1.8
Pro (P)	47	5.5	12	5.7	12	5.6	9	4.2	9	4.1
Ser (S)	76	8.9	17	8.1	25	11.7	22	10.3	27	12.4
Thr (T)	60	7.0	7	3.3	16	7.5	13	6.1	16	7.3
Trp (W)	7	0.8	3	1.4	4	1.9	3	1.4	3	1.4
Tyr (Y)	28	3.3	1	0.5	5	2.3	6	2.8	4	1.8
Val (V)	58	6.8	12	5.7	11	5.1	12	5.6	11	5.0
Pyl (O)	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0

3.3 The Homology Modelling

The homology modelling was carried by using SWISS-MODEL. 3-D structures of proteins are essential and important for understanding biological systems. 3-D homology models also reveal that the kiwellin-like proteins in these species may exhibit similar allergenicity in human. The created 3-D structures were given in Figure 3.9. From these homology-modelling results, one can estimate the allergenicity of the proteins by comparing the 3-D structures. Because there is a direct relationship between structure and function in proteins. It is important to note that the structure of kiwellin belongs to P84527 in Figure 3.4 due to the sequence length of P85261 is shorter than P84527.

Swiss-MODEL tool calculates GMQE and QSQE values when it creates 3D structures. The maximum GMQE value was found to be 0.92 for *A. deliciosa* and the minimum GMQE value was found to be 0.19 for *P. persica*.

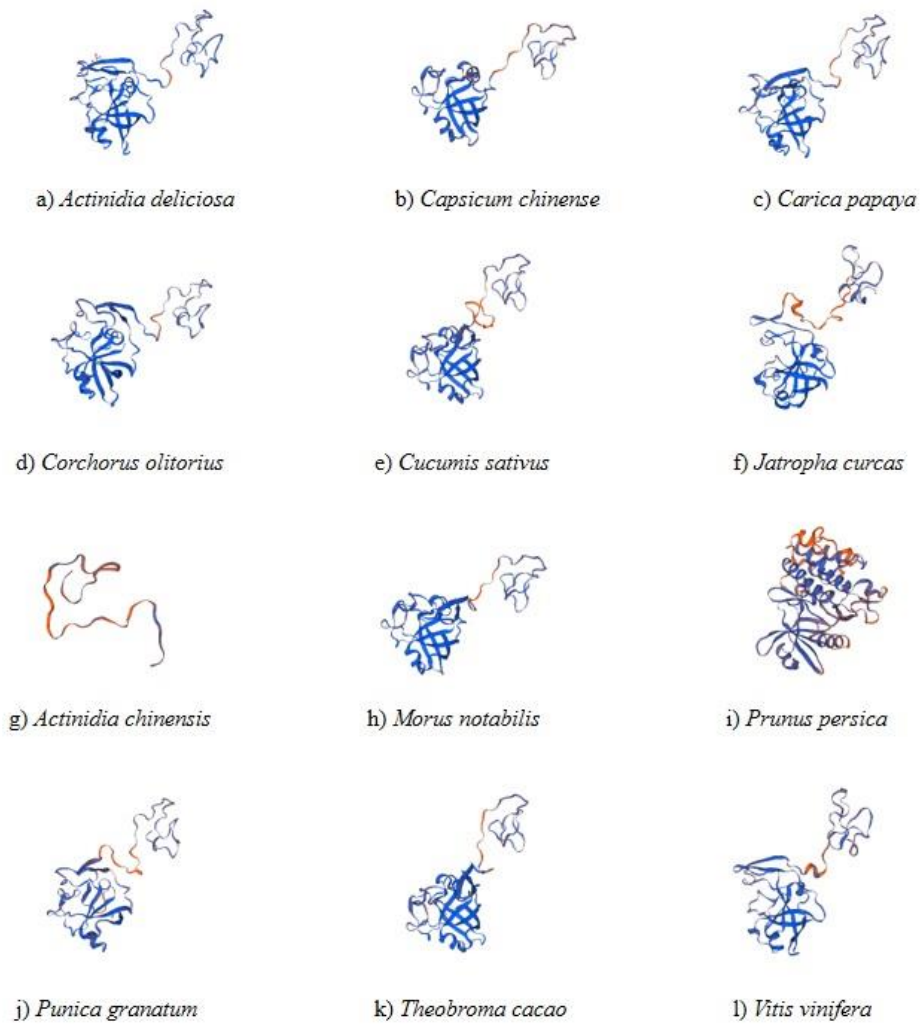


Figure 3.4 Dimensional visualization of each protein by SWISS-MODEL a) *Actinidia deliciosa*, b) *Capsicum chinense*, c) *Carica papaya*, d) *Corchorus olitorius*, e) *Cucumis sativus*, f) *Jatropha curcas*, g) *Actinidia chinensis* h) *Morus notabilis*, i) *Prunus persica*, j) *Prunus persica*, j) *Punica granatum*, k) *Theobroma cacao* l) *Vitis vinifera* (The NCBI codes of the proteins can be found in Table 3.1)

3.4 AlgPred Results

AlgPred are used for allergen prediction. Three types prediction algorithms are used to estimate allergenicity:

- i) Mapping of IgE epitopes and PID
- ii) Support vector machine module based on amino acid composition
- iii) Support vector machine module based on dipeptide composition

The results of these algorithms are given in Table 3.6 and 3.7. The mapping of IgE epitopes and PID estimated that none of 10 different proteins was identified as potential allergen. According to SVM based on amino acid results all protein except *V.vinifera* are showed allergen. Considering the SVM based on dipeptide composition results, seven protein *C.olitorius*, *C.sativus*, *C.chinense*, *C.papaya*, *M.notabilis*, *T.cacao* and *J.curcas* were identified allergen. There are seven proteins (*C.olitorius*, *C.sativus*, *C.chinense*, *C.papaya*, *M.notabilis*, *T.cacao* and *J.curcas*) in which the two SVM algorithms give positive results.

Table 3.5 AlgPred analysis of potential allergens, part 1 (The NCBI codes of the proteins can be found in Table 3.1)

AlgPred Parameters	Mapping of IgE epitopes and PID	SVM module based on amino acid composition			
Species		Allergen prediction	Score	Positive predictive value (%)	Negative predictive value (%)
<i>Corchorus olitorius</i>	X	√	0.43526052	81.83	74.03
<i>Vitis vinifera</i>	X	X	-0.74099846	22.82	92.94
<i>Cucumis sativus</i>	X	√	0.91145594	85.64	67.96
<i>Punica granatum</i>	X	√	-0.14027725	64.55	86.61
<i>Capsicum chinense</i>	X	√	0.83191078	85.64	67.96
<i>Prunus persica</i>	X	√	0.37866622	74.81	76.94
<i>Carica papaya</i>	X	√	0.77608681	87.05	71.53
<i>Theobroma cacao</i>	X	√	0.40183874	81.83	74.03
<i>Morus notabilis</i>	X	√	0.25337214	74.81	76.94
<i>Jatropha curcas</i>	X	√	0.78994643	87.05	71.53

Table 3.6 AlgPred analysis of potential allergens, part 2 (The NCBI codes of the proteins can be found in Table 3.1)

AlgPred Parameters	Mapping of IgE epitopes and PID	SVM module based on dipeptide composition			
Species		Allergen prediction	Score	Positive predictive value (%)	Negative predictive value (%)
<i>Corchorus olitorius</i>	X	√	-0.048627	63.10	85.56
<i>Vitis vinifera</i>	X	X	-0.676397	13.26	74.19
<i>Cucumis sativus</i>	X	√	0.004279	74.14	79.04
<i>Punica granatum</i>	X	X	-0.643383	13.26	74.19
<i>Capsicum chinense</i>	X	√	-0.034468	63.10	85.56
<i>Prunus persica</i>	X	X	-0.652229	13.26	74.19
<i>Carica papaya</i>	X	√	0.320983	85.88	72.01
<i>Theobroma cacao</i>	X	√	-0.1140998	63.10	85.56
<i>Morus notabilis</i>	X	√	-0.028054	63.10	85.56
<i>Jatropha curcas</i>	X	√	0.304532	85.88	72.01

It should be noted that percentage identity might not be used as important criteria to evaluate if a protein is allergen or not. When we compare all results, we see that only seven of ten potential allergens are classified as an allergen. When BLASTp results were examined, the proteins in *C.olitorius* (95%) and *C.sativus* (%80) were classified as allergen proteins. However, even if the protein in *P.granatum* has high similarity (90%), it was not identified as an allergen protein. From these results, it could be said that the similarity index cannot be used as criteria. When SWISS-MODEL results are considered, 3-D structures are different. AlgPred contains different algorithms to evaluate the submitted protein sequences (Saha & Raghava, 2006). Three of them are mapping of IgE epitopes and PID, support vector machine module based on amino acid and dipeptide compositions. So the methods are based on different algorithms. Seven proteins in this study were evaluated as allergen proteins inasmuch as they were identified by support vector machine module based on amino acid and dipeptide compositions (Table 3.5 and Table 3.6).

3.5 BIOPEP Results

Three of the potential allergen proteins (*Vitis vinifera*, *Cucumis sativus*, and *Capsicum chinense*) were separately hydrolysed with each of the enzymes in the BIOPEP tool. Many bioactive peptides are released when potential allergen proteins are hydrolysed. The majority of bioactive peptides obtained are ACE inhibitor and

dipeptidyl peptidase IV inhibitor peptides. The results obtained using the BIOPEP tool are given in the tables below.

3.5.1 Properties of bioactive peptides in *Cucumis sativus*

As the seen in the Table 3.7 when kiwellin-like protein in *C. sativus* (KGN46853.1) was hydrolysed by chymotrypsin, nine bioactive peptides were obtained and these peptides have just two properties according to BIOPEP results (data not shown). When the hydrolysed three proteins were compared, it was observed that the A_E value of DPP-IV inhibitor peptide was the highest in *V. vinifera* species in Table 3.59.

Table 3.7 DH_t , A_E , W , BE and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with chymotrypsin

DH_t [%]: 23.6842					
No	Activity	A_E	W	BE	V
1	ACE inhibitor	0.0131	0.0357	1.55199 E-5	0.00131
2	dipeptidyl peptidase IV inhibitor	0.0262	0.0522	8.29008 E-6	0.02211

Kiwellin like protein in *C. sativus* was hydrolysed by trypsin and it is revealed that this enzyme is not suitable for this application because the number of bioactive peptides and the value of DH_t was found very low. The results were shown in Table 3.8.

Table 3.8 DH_t , A_E , W , BE and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with trypsin

DH_t [%]:4.8246					
No	Activity	A_E	W	BE	V
1	ACE inhibitor	0.0044	0.0120	8.08669 E-7	6.84340 E-5

After enzymatic reaction, eight bioactive peptides were obtained and were categorized based of their mode of action. These bioactive fragments were obtained by an enzymatic reaction between kiwellin like protein in *C. sativus* and pepsin. Dipeptidyl peptidase IV inhibitor has the highest observed A_E value. DH_t value of pepsin is 10.5263 and rest results can be seen in Table 3.9.

Table 3.9 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with pepsin

DH _t [%]: 10.5263					
No	Activity	A _E	W	B _E	V
1	antioxidative	0.0044	0.1679	0	
2	ACE inhibitor	0.0131	0.0357	4.85808 E-6	0.00041
3	dipeptidyl peptidase IV inhibitor	0.0175	0.0348	3.38364 E-5	0.09023

Table 3.10 shows bioactive fragment parameters after hydrolytic reaction of proteinase K with kiwellin like protein in *C. sativus*. Sixteen bioactive peptides were obtained but were not revealed in Table 3.10. According to these results, four different activities were observed and shown in the Table 3.10

Table 3.10 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with proteinase K

DH _t [%]: 28.5088					
No	Activity	A _E	W	B _E	V
1	antioxidative	0.0087	0.3321	0	
2	ACE inhibitor	0.0131	0.0357	4.85808 E-6	0.00041
3	dipeptidyl peptidase IV inhibitor	0.0437	0.0870	9.07726 E-5	0.24206
4	hypotensive	0.0044	0.5057	0	0

Elastase is a suitable enzyme for this type of enzymatic reaction because of its DH_t value. When DH_t value is high, it has a chance to produce even more bioactive peptides compared to other enzymes with low DH_t value. When *in silico* hydrolysis was performed, 21 bioactive peptides were obtained and they were classified into four groups, which are shown in Table 3.10. When we compare three proteins produced after elastase hydrolysis, the DPP-IV inhibitor was obtained from the hydrolysis of kiwellin-like protein in *C.chinense*. It has the highest A_E value and the results are shown in Table 3.37.

Table 3.11 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with elastase

DH _t [%]: 50.8772					
No	Activity	A _E	W	B _E	V
1	dipeptidyl peptidase IV inhibitor	0.0699	0.1392	6.78603 E-6	0.01810
2	activating ubiquitin-mediated proteolysis	0.0044	0.5057	0	
3	ACE inhibitor	0.0480	0.1309	0.00037	0.03134
4	antioxidative	0.0044	0.1679	0	

V-protease is not a suitable enzyme for hydrolysis of kiwellin like protein. Because DH_t value cleavage of protein was found to be 2.1930. Only two bioactive fragments are released and they are seen in Table 3.12.

Table 3.12 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with V-protease

DH _t [%]: 2.1930					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0044	0.0120	0	0
2	dipeptidyl peptidase IV inhibitor	0.0044	0.0088	0	0

Table 3.13 reveals the results related to kiwellin-like protein in *C. sativus* hydrolysis by thermolysin. DH_t value was found to be 25.8772 and the max A_E value was found in ACE inhibitor.

Table 3.13 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with thermolysin

DH _t [%]: 25.8772					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0218	0.0594	2.73918 E-6	0.00023
2	antioxidative	0.0044	0.1679	0	
3	dipeptidyl peptidase IV inhibitor	0.0175	0.0348	0	0

When the protein was hydrolysed with the enzyme chymotrypsin, sixteen bioactive fragments were released. Although the DH_t value was 26.7544%, the resulting bioactive peptide diversity was high. As can be seen in Table 3.14, bioactive peptides

have seven different properties. Enzyme-linked bioactive fragment frequency (A_E) value was highest in the DPP-IV inhibitor peptide.

Table 3.14 DH_i , A_E , W , B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with chymotrypsin

DH_i [%]: 26.7544					
No	Activity	A_E	W	B_E	V
1	regulating	0.0044	1.0000	0	
2	neuropeptide	0.0044	0.5057	0	
3	dipeptidyl peptidase IV inhibitor	0.0306	0.0609	8.74073 E-6	0.02331
4	antithrombotic	0.0044	1.0000	0	
5	antiamnestic	0.0044	1.0000	0	
6	ACE inhibitor	0.0175	0.0477	2.14027E-5	0.00181
7	stimulating	0.0044	0.1438	0	

According to the results obtained after hydrolysis of protein by plasmin enzyme, DH_i value was found to be quite low. Because the DH_i value was 4.8246, one bioactive fragment was released (Table 3.15). In addition, A_E was found to be very low.

Table 3.15 DH_i , A_E , W , B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with plasmin

DH_i [%]: 4.8246					
No	Activity	A_E	W	B_E	V
1	ACE inhibitor	0.0044	0.0120	8.08669 E-7	6.84340 E-5

Table 3.16 shows the bioactive protein parameters after cathepsin hydrolysis. According to results, six bioactive peptides were obtained and these bioactive peptides were associated with two different activities. These are ACE inhibitor and DPP-IV inhibitor.

Table 3.16 DH_i , A_E , W , B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with cathepsin

DH_i [%]: 15.7895					
No	Activity	A_E	W	B_E	V
1	ACE inhibitor	0.0087	0.0237	3.49345 E-6	0.00030
2	dipeptidyl peptidase IV inhibitor	0.0175	0.0348	8.29008 E-6	0.02211

As can be seen from the Table 3.17, after chymase *in silico* hydrolysis, eight bioactive peptides were obtained and these bioactive peptides have three different activities.

Table 3.17 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with chymase

DH _t [%]: 12.2807					
No	Activity	A _E	W	B _E	V
1	antioxidative	0.0044	0.1679	0	
2	ACE inhibitor	0.0131	0.0357	4.85808 E-6	0.00041
3	dipeptidyl peptidase IV inhibitor	0.0175	0.0348	3.38364 E-5	0.09023

As can be seen from Table 3.18 when the protein was hydrolysed by papain, the degree of hydrolysis value was found high, thirty-one bioactive peptides that have five properties were obtained. When the three proteins were hydrolysed by the same enzyme, it was found that DPP-IV inhibitor peptide obtained from the hydrolysis of kiwellin-like protein in *C.chinense* had the highest A_E value. These values can be seen in Table 3.45

Table 3.18 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with papain

DH _t [%]: 35.9649					
No	Activity	A _E	W	B _E	V
1	antioxidative	0.0131	0.5000	0	
2	ACE inhibitor	0.0480	0.1309	0.00294	0.24911
3	hypotensive	0.0087	1.0000	0.00024	1
4	inhibitor	0.0044	1.0000	0	
5	dipeptidyl peptidase IV inhibitor	0.0611	0.1217	4.53475 E-5	0.12092

As can be seen from Table 3.19, the value of hydrolysis was found to be 43.4211. Because the enzymatic cleavage value was high, twenty four bioactive fragments were obtained. These bioactive peptides show four different activities and max A_E value was found in the ACE inhibitor. When the hydrolysed three proteins were compared,

it was observed that the A_E value of DPP-IV inhibitor peptide was the highest in *C.chinense* in Table 3.46.

Table 3.19 DH_t , A_E , W , B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with ficin

DH_t [%]: 43.4211					
No	Activity	A_E	W	B_E	V
1	ACE inhibitor	0.0393	0.1071	0.00087	0.07338
2	antioxidative	0.0044	0.1679	0	
3	stimulating	0.0044	0.1438	0	
4	dipeptidyl peptidase IV inhibitor	0.0568	0.1131	0.00015	0.40821

After the kiwellin like protein hydrolysed by leukocyte elastase, nineteen bioactive peptides were obtained. As can be seen from Table 3.20, these bioactive peptides have three activities. Three potential allergen proteins were hydrolysed by elastase enzyme and when the A_E values of bioactive peptides obtained from these three proteins were compared, the A_E value of DPP-IV inhibitor peptide was found to be highest in *C.chinense* species in Table 3.47.

Table 3.20 DH_t , A_E , W , B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with leukocyte elastase

DH_t [%]: 35.5263					
No	Activity	A_E	W	B_E	V
1	dipeptidyl peptidase IV inhibitor	0.0437	0.0870	1.17957 E-5	0.03145
2	activating ubiquitin-mediated proteolysis	0.0044	0.5057	0	
3	ACE inhibitor	0.0349	0.0951	1.89468 E-5	0.00160

As it can be understood from the results of *in silico* hydrolysis by metridin enzyme, DH_t value was found to be 12.2807. Eight bioactive fragments showed three different activities were obtained (Table 3.21).

Table 3.21 DH_t, A_E, W, B_E and V in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with metridin

DH _t [%]: 12.2807					
No	Activity	A _E	W	B _E	V
1	antioxidative	0.0044	0.1679	0	
2	ACE inhibitor	0.0131	0.0357	4.85808 E-6	0.00041
3	dipeptidyl peptidase IV inhibitor	0.0175	0.0348	3.38364 E-5	0.09023

Table 3.22 shows the bioactive protein parameters after pancreatic elastase II hydrolysis. According to results, the degree of hydrolysis value was found to be 11.8421 and eight bioactive peptides had three properties.

Table 3.22 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with pancreatic elastase II

DH _t [%]: 11.8421					
No	Activity	A _E	W	B _E	V
1	antioxidative	0.0044	0.1679	0	
2	ACE inhibitor	0.0131	0.0357	4.85808 E-6	0.00041
3	dipeptidyl peptidase IV inhibitor	0.0175	0.0348	3.38364 E-5	0.09023

According to Table 3.23 bromelain is a suitable enzyme for hydrolysis of potential allergen protein. We observed significant increased number of bioactive peptides. Twenty-seven bioactive peptides were found and these peptides showed six properties.

Table 3.23 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with bromelain

DH _t [%]: 49.1228					
No	Activity	A _E	W	B _E	V
1	dipeptidyl peptidase IV inhibitor	0.0524	0.1043	3.72827 E-5	0.09942
2	antioxidative	0.0087	0.3321	0	
3	ACE inhibitor	0.0437	0.1191	0.00060	0.05065
4	hypotensive	0.0044	0.5057	0.00024	1
5	inhibitor	0.0044	1.0000	0	
6	stimulating	0.0044	0.1438	0	

A low DH_t value indicates that the oligopeptidase B enzyme is not suitable for hydrolysis. From the Table 3.24, we can see that only one bioactive peptide was released and the activity of ACE inhibitor peptide was 0.0044.

Table 3.24 DH_t , A_E , W , B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with oligopeptidase B

DH_t [%]: 4.8246					
No	Activity	A_E	W	B_E	V
1	ACE inhibitor	0.0044	0.0120	8.08669 E-7	6.84340 E-5

The result of *in silico* hydrolysis using calpain 2 is shown in the Table 3.25. Nineteen bioactive peptides revealed three activities. Although ACE inhibitor peptide and DPP-IV inhibitor peptide had similar A_E values, ACE inhibitor peptide activity of fragments (B_E) value was highest.

Table 3.25 DH_t , A_E , W , B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with calpain 2

DH_t [%]: 42.5439					
No	Activity	A_E	W	B_E	V
1	ACE inhibitor	0.0437	0.1191	0.00387	0.32763
2	antioxidative	0.0044	0.1679	0	
3	dipeptidyl peptidase IV inhibitor	0.0349	0.0695	1.48509 E-5	0.03960

Table 3.26 reveals the results related to hydrolysis by glycyl endopeptidase. According to Table 3.26, the degree of hydrolysis value was found to be 14.9123 and three bioactive peptides revealed ACE inhibitor properties.

Table 3.26 DH_t , A_E , W , B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with glycyl endopeptidase

DH_t [%]: 14.9123					
No	Activity	A_E	W	B_E	V
1	ACE inhibitor	0.0131	0.0357	1.72063 E-6	0.00015

The results of hydrolysis of potential allergen protein in *C. sativus* by oligopeptidase F enzyme are shown in Table 3.27. The degree of hydrolysis value was found to be 10.5263 and eight bioactive peptides have three properties.

Table 3.27 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with oligopeptidase F

DH _t [%]: 10.5263					
No	Activity	A _E	W	B _E	V
1	antioxidative	0.0044	0.1679	0	
2	ACE inhibitor	0.0131	0.0357	4.85808 E-6	0.00041
3	dipeptidyl peptidase IV inhibitor	0.0175	0.0348	3.38364 E-5	0.09023

Table 3.28 shows that 38.1579 DH_t value was found after proteinase P1 (lactocepin) cleavage. The results also shows that fragments are ACE inhibitor, stimulating, antioxidative, dipeptidyl peptidase IV inhibitor.

Table 3.28 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with proteinase P1 (lactocepin)

DH _t [%]: 38.1579					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0087	0.0237	0.00036	0.03009
2	stimulating	0.0044	0.1438	0	
3	antioxidative	0.0044	0.1679	0	
4	dipeptidyl peptidase IV inhibitor	0.0175	0.0348	5.19859 E-5	0.13863

When the protein was hydrolysed by pepsin (pH>2) the value of the degree of hydrolysis was found to be higher than to be other results. As can be seen from Table 3.29 thirty-six bioactive peptides with five different properties were obtained. When the hydrolysed three proteins were compared, it was observed that the A_E value of DPP-IV inhibitor peptide was same in *C. chinense* and *V. vinifera*. These results can be seen in Table 3.56 and Table 3.81. In addition, the variety of bioactive peptides kiwellin-like protein in *C. chinense* were found higher than other results.

Table 3.29 DH_t, A_E, W, BE and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with pepsin (pH>2)

DH _t [%]: 63.5965					
No	Activity	A _E	W	B _E	V
1	dipeptidyl peptidase IV inhibitor	0.0873	0.1738	0.00017	0.46526
2	antioxidative	0.0044	0.1679	0	
3	activating ubiquitin-mediated proteolysis	0.0044	0.5057	0	
4	ACE inhibitor	0.0480	0.1309	0.00227	0.19183
5	stimulating	0.0131	0.4281	0	

As can be seen from Table 3.30 when the protein was hydrolysed by cocolysin enzyme, thirteen bioactive peptides with four different properties were obtained. These are ACE inhibitor, stimulating, antioxidative, dipeptidyl peptidase IV inhibitor.

Table 3.30 DH_t, A_E, W, BE and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with cocolysin enzyme

DH _t [%]: 21.0526					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0262	0.0714	7.30564 E-6	0.00062
2	antioxidative	0.0044	0.1679	0	
3	stimulating	0.0044	0.1438	0	
4	dipeptidyl peptidase IV inhibitor	0.0218	0.0434	0	0

Twenty-one bioactive peptides with five different properties were observed after subtilisin *in silico* hydrolysis (Table 3.31).

Table 3.31 DH_t, A_E, W, BE and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with subtilisin

DH _t [%]: 25.4386					
No	Activity	A _E	W	B _E	V
1	dipeptidyl peptidase IV inhibitor	0.0349	0.0695	0.00015	0.40694
2	antioxidative	0.0087	0.3321	0	
3	ACE inhibitor	0.0349	0.0951	0.00313	0.26466
4	stimulating	0.0087	0.2843	0	
5	alpha-glucosidase inhibitor	0.0044	1.0000	0.00019	1

Table 3.32 reveals the results related to kiwellin-like protein in *C.sativus* hydrolysis by V-8 protease (Glutamyl endopeptidase). The value of the degree of hydrolysis was found to be 9.2105 and eight bioactive peptides with two different activities were observed these are ACE inhibitor and DPP-IV inhibitor.

Table 3.32 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. sativus* obtained through *in silico* hydrolysis with V-8 protease (Glutamyl endopeptidase)

DH _t [%]: 9.2105					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0044	0.0120	0	0
2	dipeptidyl peptidase IV inhibitor	0.0044	0.0088	0	0

In silico enzymatic cleavage was done by using other enzymes in BIOPEP such as clostripain, thrombin etc. However, we observed no bioactive peptides from these enzymes.

3.5.2 Properties of bioactive peptides in *Capsicum chinense*

As can be seen from the Table 3.33 when kiwellin-like protein in *C. chinense* (Ripening-related protein grip22 PHU03889.1) was hydrolysed by chymotrypsin, eight bioactive peptides were obtained. In addition, these peptides have just two properties according to BIOPEP results (data not shown).

Table 3.33 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with chymotrypsin

DH _t [%]: 21.4953					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0140	0.0367	1.61704 E-5	0.00134
2	dipeptidyl peptidase IV inhibitor	0.0233	0.0398	1.77863 E-6	0.00377

Kiwellin like protein in *C. chinense* was hydrolysed by trypsin and it was revealed that this enzyme is not suitable for this application because the number of bioactive peptides and the value of DH_t was found to be 8.8785. The results were shown in Table 3.34.

Table 3.34 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with trypsin

DH _t [%]: 8.8785					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0093	0.0244	7.55365 E-6	0.00063
2	antioxidative	0.0047	0.1263	0	
3	hypotensive	0.0047	0.2527	0.00051	0.39750
4	inhibitor	0.0047	0.2527	0	
5	dipeptidyl peptidase IV inhibitor	0.0047	0.0080	0	0

After enzymatic reaction, four bioactive peptides were obtained and were categorized based of their mode of action. These bioactive fragments were obtained by an enzymatic reaction between kiwellin like protein in *C. chinense* and pepsin. Dipeptidyl peptidase IV inhibitor has the highest observed A_E value. DH_t value of pepsin is 7.0093 and rest of the results can be seen in Table 3.35.

Table 3.35 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with pepsin

DH _t [%]: 7.0093					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0047	0.0123	1.86047 E-6	0.00015
2	dipeptidyl peptidase IV inhibitor	0.0140	0.0239	1.77863 E-6	0.00377

Table 3.36 shows the bioactive fragment parameters after hydrolytic reaction of proteinase K with kiwellin like protein in *C. chinense*. Fifteen bioactive peptides were obtained but were not revealed in Table 3.36. According to these results, four different activities were shown in the Table 3.36.

Table 3.36 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with proteinase K

DH _t [%]: 31.3084					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0140	0.0367	0.00137	0.11338
2	antioxidative	0.0047	0.1263	0	
3	dipeptidyl peptidase IV inhibitor	0.0512	0.0874	6.32000 E-5	0.13390

Pancreatic elastase is a suitable enzyme for this type of enzymatic reaction because of its DH_t value. When DH_t value is high, it has a chance to produce even more bioactive peptides compared to other enzymes with low DH_t value. When *in silico* hydrolysis was performed, twenty-nine bioactive peptides were obtained and they were classified into four groups which are shown in Table 3.37.

Table 3.37 DH_t , A_E , W , B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with pancreatic elastase

DH_t [%]: 50.9346					
No	Activity	A_E	W	B_E	V
1	regulating	0.0047	0.3357	0	
2	dipeptidyl peptidase IV inhibitor	0.0837	0.1428	5.83396 E-5	0.12360
3	antithrombotic	0.0047	0.3357	0	
4	antiamnestic	0.0047	0.3357	0	
5	ACE inhibitor	0.0326	0.0855	0.00104	0.08592
6	activating ubiquitin-mediated proteolysis	0.0047	1.0000	0	

Prolyl oligopeptidase is not a suitable enzyme for hydrolysis of kiwellin like protein. Because it has 6.0748 DH_t value cleavage of protein. Only two bioactive fragments were released and it can be seen in Table 3.38.

Table 3.38 DH_t , A_E , W , B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with prolyl oligopeptidase

DH_t [%]: 6.0748					
No	Activity	A_E	W	B_E	V
1	ACE inhibitor	0.0047	0.0123	6.98373 E-5	0.00578
2	dipeptidyl peptidase IV inhibitor	0.0047	0.0080	7.77786 E-7	0.00165

Table 3.39 reveals the results related to kiwellin-like protein in *C. chinense* hydrolysis by thermolysin. DH_t value was found to be 25.7009 and the max A_E value was found as DPP-IV inhibitors.

Table 3.39 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with prolyl thermolysin

DH _t [%]: 25.7009					
No	Activity	A _E	W	B _E	V
1	dipeptidyl peptidase IV inhibitor	0.0372	0.0635	1.09157 E-5	0.02313
2	ACE inhibitor	0.0279	0.0732	0.00020	0.01687
3	antioxidative	0.0093	0.2500	0	
4	hypotensive	0.0093	0.5000	0.00101	0.79501
5	inhibitor	0.0093	0.5000	0	

When the protein was hydrolysed with the enzyme chymotrypsin, twelve bioactive fragments were released. Although the DH_t value was 31.7757%, the resulting bioactive peptide diversity was high. As can be seen in Table 3.40, bioactive peptides have six different properties. Enzyme-linked bioactive fragment frequency (A_E) value was highest in the DPP-IV inhibitor peptide.

Table 3.40 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with chymotrypsin

DH _t [%]: 31.7757					
No	Activity	A _E	W	B _E	V
1	regulating	0.0047	0.3357	0	
2	dipeptidyl peptidase IV inhibitor	0.0233	0.0398	8.32182 E-6	0.01763
3	antithrombotic	0.0047	0.3357	0	
4	antiamnestic	0.0047	0.3357	0	
5	ACE inhibitor	0.0140	0.0367	3.13456 E-5	0.00259
6	stimulating	0.0047	0.1011	0	

According to the results obtained after hydrolysis of protein by plasmin enzyme, DH_t value was found to be quite low. Because the DH_t value was found 8.8785, six bioactive fragments were released (Table 3.41).

Table 3.41 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with plasmin

DH _t [%]: 8.8785					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0093	0.0244	7.55365 E-6	0.00062
2	antioxidative	0.0047	0.1263	0	
3	hypotensive	0.0047	0.2527	0.00051	0.39750
4	inhibitor	0.0047	0.2527	0	
5	dipeptidyl peptidase IV inhibitor	0.0047	0.0080	0	0

Table 3.42 shows the bioactive protein parameters after cathepsin hydrolysis. According to results, five bioactive peptides were obtained and these bioactive peptides were associated with two different activities. These were ACE inhibitor and DPP-IV inhibitor activity.

Table 3.42 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with cathepsin

DH _t [%]: 11.6822					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0047	0.0123	1.86046 E-6	0.00015
2	dipeptidyl peptidase IV inhibitor	0.0186	0.0317	1.77863 E-6	0.00377

Table 3.43 indicates *in silico* hydrolysis of the protein by clostripain. Five bioactive peptides were obtained and these bioactive peptides had five different activities.

Table 3.43 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with clostripain

DH _t [%]: 3.7383					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0047	0.0123	6.69232 E-6	0.00055
2	antioxidative	0.0047	0.1263	0	
3	hypotensive	0.0047	0.2527	0.00051	0.39750
4	inhibitor	0.0047	0.2527	0	
5	dipeptidyl peptidase IV inhibitor	0.0047	0.0080	0	0

As can be seen from Table 3.44, after *in silico* hydrolysis by chymase, and. four bioactive peptides were obtained and these bioactive peptides had two different activities

Table 3.44 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* with chymase

DH _t [%]: 10.2804					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0047	0.0123	1.86046 E-6	0.00015
2	dipeptidyl peptidase IV inhibitor	0.0140	0.0239	1.77863 E-6	0.00377

When the protein was hydrolysed by papain, the degree of hydrolysis value was found as 32.7103 and thirty-six bioactive peptides that have nine properties were obtained. Results can be seen in Table 3.45.

Table 3.45 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* with papain

DH _t [%]: 32.7103					
No	Activity	A _E	W	B _E	V
1	regulating	0.0047	0.3357	0	
2	dipeptidyl peptidase IV inhibitor	0.0698	0.1191	9.52237 E-5	0.20175
3	ACE inhibitor	0.0419	0.1099	0.00177	0.14683
4	antithrombotic	0.0047	0.3357	0	
5	antiamnestic	0.0047	0.3357	0	
6	antioxidative	0.0140	0.3763	0	
7	hypotensive	0.0093	0.5000	0.00101	0.79501
8	inhibitor	0.0093	0.5000	0	
9	stimulating	0.0093	0.2000	0	

As can be seen from the Table 3.46, the value of hydrolysis was found 38.3178. Because the enzymatic cleavage value was high, thirty-eight bioactive fragments were obtained. These bioactive peptides shown six different activities and max A_E value was found in the ACE inhibitor.

Table 3.46 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with ficin

DH _t [%]: 38.3178					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0465	0.1219	0.00126	0.10460
2	antioxidative	0.0140	0.3763	0	
3	hypotensive	0.0093	0.5000	0.00101	0.79501
4	inhibitor	0.0093	0.5000	0	
5	stimulating	0.0093	0.2000	0	
6	dipeptidyl peptidase IV inhibitor	0.0884	0.1509	8.46784 E-5	0.17941

After the kiwellin like protein hydrolysed by leukocyte elastase, twenty bioactive peptides were obtained. As can be seen from Table 3.47, these bioactive peptides have three activities.

Table 3.47 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with leukocyte elastase

DH _t [%]: 38.3178					
No	Activity	A _E	W	B _E	V
1	dipeptidyl peptidase IV inhibitor	0.0651	0.1111	6.01182 E-5	0.12737
2	activating ubiquitin-mediated proteolysis	0.0047	1.0000	0	
3	ACE inhibitor	0.0233	0.0611	1.63372 E-5	0.00135

As it can be understood from the results of *in silico* hydrolysis by metridin enzyme DH_t value was found to be 10.2804. Four bioactive fragments showed two different activities (Table 3.48).

Table 3.48 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with metridin

DH _t [%]: 10.2804					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0047	0.0123	1.86046 E-6	0.00015
2	dipeptidyl peptidase IV inhibitor	0.0140	0.0239	1.77863 E-6	0.00377

Table 3.49 shows the bioactive protein parameters after pancreatic elastase II hydrolysis. According to results, the degree of hydrolysis value was found to be 9.3458 and five bioactive peptides that have two properties were obtained.

Table 3.49 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with pancreatic elastase II

DH _t [%]: 9.3458					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0047	0.0123	1.86046 E-6	0.00015
2	dipeptidyl peptidase IV inhibitor	0.0186	0.0317	1.77863 E-6	0.00377

According to Table 3.50, bromelain is a suitable enzyme for hydrolysis of a potential allergen proteins. We observed significantly increased number of bioactive peptides. Forty-one bioactive peptides were found and these peptides showed nine properties.

Table 3.50 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with stem bromelain

DH _t [%]: 48.1308					
No	Activity	A _E	W	B _E	V
1	regulating	0.0047	0.3357	0	
2	dipeptidyl peptidase IV inhibitor	0.0884	0.1509	7.22791 E-6	0.01531
3	ACE inhibitor	0.0465	0.1219	0.00114	0.09457
4	antithrombotic	0.0047	0.3357	0	
5	antiemetic	0.0047	0.3357	0	
6	antioxidative	0.0093	0.2500	0	
7	hypotensive	0.0140	0.7527	0.00127	1
8	inhibitor	0.0140	0.7527	0	
9	stimulating	0.0047	0.1011	0	

A low DH_t value indicates that the oligopeptidase B enzyme is not suitable for hydrolysis. From Table 3.51, six bioactive peptides were released and we can see five different activity of bioactive peptides.

Table 3.51 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with oligopeptidase B

DH _t [%]: 8.8785					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0093	0.0244	7.55365 E-6	0.00062
2	antioxidative	0.0047	0.1263	0	
3	hypotensive	0.0047	0.2527	0.00051	0.39750
4	inhibitor	0.0047	0.2527	0	
5	dipeptidyl peptidase IV inhibitor	0.0047	0.0080	0	0

The result of *in silico* hydrolysis using calpain 2 is shown in Table 3.52. Forty-seven bioactive peptides have nine activities. Although ACE inhibitor peptide and DPP-IV inhibitor peptide had similar A_E values, ACE inhibitor peptide activity of fragments (B_E) value was highest.

Table 3.52 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with calpain 2

DH _t [%]: 42.5234					
No	Activity	A _E	W	B _E	V
1	regulating	0.0047	0.3357	0	
2	dipeptidyl peptidase IV inhibitor	0.0930	0.1587	9.86270 E-5	0.20895
3	ACE inhibitor	0.0698	0.1830	0.00503	0.41612
4	antithrombotic	0.0047	0.3357	0	
5	antiemetic	0.0047	0.3357	0	
6	antioxidative	0.0140	0.3763	0	
7	hypotensive	0.0093	0.5000	0.00101	0.79501
8	inhibitor	0.0093	0.5000	0	
9	stimulating	0.0093	0.2000	0	

Table 3.53 reveals the results related to hydrolysis by glycyl endopeptidase. According to Table 3.53, DH_t value was found to be 11.2150 and bioactive peptide had ACE inhibitor properties.

Table 3.53 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with glycyl endopeptidase

DH _t [%]: 11.2150					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0047	0.0123	0.00038	0.03130

The results of hydrolysis of potential allergen protein in *C. chinense* by oligopeptidase F enzyme were shown in Table 3.54. The degree of hydrolysis value was found to be 7.0093. Four bioactive peptides had two properties.

Table 3.54 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with oligopeptidase F

DH _t [%]: 7.0093					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0047	0.0123	1.86046 E-6	0.00015
2	dipeptidyl peptidase IV inhibitor	0.0140	0.0239	1.77863 E-6	0.00377

Table 3.55 shows that 37.8505 DH_t value was found after proteinase P1 (lactocepin) cleavage. The results also shows that fragments are DPP-IV inhibitor and ACE inhibitors.

Table 3.55 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with proteinase P1 (lactocepin)

DH _t [%]: 37.8505					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0233	0.0611	0.00154	0.12733
2	dipeptidyl peptidase IV inhibitor	0.0186	0.0317	0	0

When the protein was hydrolysed by pepsin (pH>2), the value of the degree of hydrolysis was found to be higher than other results. As can be seen from Table 3.56 thirty-eight bioactive peptides with nine different properties were obtained.

Table 3.56 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with pepsin (pH>2)

DH _t [%]: 64.9533					
No	Activity	A _E	W	B _E	V
1	regulating	0.0047	0.3357	0	
2	dipeptidyl peptidase IV inhibitor	0.0837	0.1428	0.00013	0.28514
3	ACE inhibitor	0.0512	0.1342	0.00264	0.21873
4	antithrombotic	0.0047	0.3357	0	
5	antiamnestic	0.0047	0.3357	0	
6	antioxidative	0.0047	0.1263	0	
7	hypotensive	0.0047	0.2527	0.00051	0.39750
8	inhibitor	0.0047	0.2527	0	
9	stimulating	0.0140	0.3011	0	

As can be seen from Table 3.57 when the protein was hydrolysed by cocolysin enzyme, fifteen bioactive peptides with five different properties were obtained. These are ACE inhibitor, antioxidative, hypotensive, inhibitor, dipeptidyl peptidase IV inhibitor.

Table 3.57 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with cocolysin

DH _t [%]: 19.1589					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0186	0.0488	0.00010	0.00866
2	antioxidative	0.0093	0.2500	0	
3	hypotensive	0.0093	0.5000	0.00101	0.79501
4	inhibitor	0.0093	0.5000	0	
5	dipeptidyl peptidase IV inhibitor	0.0233	0.0398	0	0

Sixteen bioactive peptides with five different properties were observed after subtilisin *in silico* hydrolyses (Table 3.58.) These properties were categorized in five groups.

Table 3.58 DH_t, A_E, W, B_E and V values in kiwellin-like protein in *C. chinense* obtained through *in silico* hydrolysis with subtilisin

DH _t [%]: 23.8318					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0186	0.0488	0.000256	0.02117
2	hypotensive	0.0047	0.2527	0.00026	0.20499
3	inhibitor	0.0047	0.2527	0	
4	stimulating	0.0047	0.1011	0	
5	dipeptidyl peptidase IV inhibitor	0.0419	0.0715	7.02624 E-5	0.14886

In silico enzymatic cleavage was done by using other enzymes in BIOPEP such as V-protease. However, we observed no bioactive peptides from these enzymes.

3.5.3 Properties of bioactive peptides in *Vitis vinifera*

As can be seen from the Table 3.59 when kiwellin-like protein in *V. vinifera* (unnamed protein product, partial (BLASTp) - CBI16343.3) was hydrolysed by chymotrypsin, the value of the degree of hydrolysis was found high, thirty-five bioactive peptides were obtained, and these bioactive peptides have four different properties according to BIOPEP results (data not shown).

Table 3.59 DH_t, A_E, W, B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with chymotrypsin

DH _t [%]: 27.4314					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0249	0.0731	0.00171	0.12705
2	antioxidative	0.0075	0.1773	0	
3	stimulating	0.0025	0.0419	0	
4	dipeptidyl peptidase IV inhibitor	0.0522	0.0864	3.74948 E-5	0.04611

Kiwellin like protein in *V. vinifera* was hydrolysed by trypsin and it was revealed that this enzyme is not suitable for this application because the number of bioactive peptides and the value of DH_t was found to be 4.8246. The results were shown in Table 3.60.

Table 3.60 DH_t, A_E, W, B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with trypsin

DH _t [%]: 4.8246					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0044	0.0120	8.08669 E-7	6.84340 E-5

After enzymatic reaction, eight bioactive peptides were obtained and were categorized based of their mode of action. These bioactive fragments were obtained by an enzymatic reaction between kiwellin like protein in *V. vinifera* and pepsin. Dipeptidyl peptidase IV inhibitor has the highest observed A_E value. DH_t value of pepsin is 10.5263 and rest of the results can be seen in Table 3.61

Table 3.61 DH_t, A_E, W, B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with pepsin

DH _t [%]: 10.5263					
No	Activity	A _E	W	B _E	V
1	antioxidative	0.0044	0.1679	0	
2	ACE inhibitor	0.0131	0.0357	4.8580786026201E-6	0.0004111171371903
3	dipeptidyl peptidase IV inhibitor	0.0175	0.0348	3.383640827727E-5	0.090229018813326

Table 3.10 shows bioactive fragment parameters after hydrolytic reaction of proteinase K with kiwellin like protein in *V. vinifera*. Sixteen bioactive peptides were obtained but were revealed in Table 3.62. According to these results, four different activities shown in the Table 3.62

Table 3.62 DH_t, A_E, W, B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with proteinase K

DH _t [%]: 28.5088					
No	Activity	A _E	W	B _E	V
1	antioxidative	0.0087	0.3321	0	
2	ACE inhibitor	0.0131	0.0357	4.85808 E-6	0.00041
3	dipeptidyl peptidase IV inhibitor	0.0437	0.0870	9.07726 E-5	0.24206
4	hypotensive	0.0044	0.5057	0	0

Elastase is a suitable enzyme for this type of enzymatic reaction because of its DH_t value. When *in silico* hydrolysis was performed, 29 bioactive peptides were obtained and they were classified into four groups, which are shown in Table 3.63.

Table 3.63 DH_t , A_E , W , B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with elastase

DH_t [%]: 50.8772					
No	Activity	A_E	W	B_E	V
1	dipeptidyl peptidase IV inhibitor	0.0699	0.1392	6.78603 E-6	0.01810
2	activating ubiquitin-mediated proteolysis	0.0044	0.5057	0	
3	ACE inhibitor	0.0480	0.1309	0.00037	0.03134
4	antioxidative	0.0044	0.1679	0	

V-protease is not a suitable enzyme for hydrolysis of kiwellin-like protein. Because DH_t value was found to be 2.1930. Only two bioactive fragments are released it can be seen in Table 3.64.

Table 3.64 DH_t , A_E , W , B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with V-protease

DH_t [%]: 2.1930					
No	Activity	A_E	W	B_E	V
1	ACE inhibitor	0.0044	0.0120	0	0
2	dipeptidyl peptidase IV inhibitor	0.0044	0.0088	0	0

Table 3.13 reveals the results related to kiwellin-like protein in *V. vinifera* hydrolysis by thermolysin. DH_t value was found to be 25.8772 and the max A_E value was found as ACE inhibitor

Table 3.65 DH_t , A_E , W , B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with thermolysin.

DH_t [%]: 25.8772					
No	Activity	A_E	W	B_E	V
1	ACE inhibitor	0.0218	0.0594	2.73918 E-6	0.00023
2	antioxidative	0.0044	0.1679	0	
3	dipeptidyl peptidase IV inhibitor	0.0175	0.0348	0	0

When the protein was hydrolysed with the enzyme chymotrypsin, sixteen bioactive fragments were released. Although the DH_t value was 26.7544%, the resulting bioactive peptide diversity was high. As can be seen in Table 3.66, bioactive peptides have seven different properties. Enzyme-linked bioactive fragment frequency (A_E) value was highest in the DPP-IV inhibitor peptide.

Table 3.66 DH_t, A_E, W, B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with chymotrypsin

DH _t [%]: 26.7544					
No	Activity	A _E	W	B _E	V
1	regulating	0.0044	1.0000	0	
2	neuropeptide	0.0044	0.5057	0	
3	dipeptidyl peptidase IV inhibitor	0.0306	0.0609	8.74074 E-6	0.02331
4	antithrombotic	0.0044	1.0000	0	
5	antiemetic	0.0044	1.0000	0	
6	ACE inhibitor	0.0175	0.0477	2.14027 E-5	0.00181
7	stimulating	0.0044	0.1438	0	

According to the results obtained after hydrolysis of protein by plasmin enzyme, DH_t value was found to be quite low. Because the DH_t value was 4.8246, one bioactive fragment was released (Table 3.67). In addition, A_E was found to be very low.

Table 3.67 DH_t, A_E, W, B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with plasmin

DH _t [%]: 4.8246					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0044	0.0120	8.08669 E-7	6.84340 E-5

Table 3.68 shows the bioactive protein parameters after cathepsin hydrolysis. According to results, six bioactive peptides were obtained and these bioactive peptides were associated with two different activities. These were ACE inhibitor and DPP-IV inhibitor activity.

Table 3.68 DH_t, A_E, W, B_E and V values kiwellin-like protein in V. vinifera obtained through *in silico* hydrolysis with cathepsin

DH _t [%]: 15.7895					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0087	0.0237	3.49345 E-6	0.00030
2	dipeptidyl peptidase IV inhibitor	0.0175	0.0348	8.29008 E-6	0.02211

As can be seen from the Table 3.68, after by chymase *in silico* hydrolysis, eight bioactive peptides were obtained and these bioactive peptides have three different activities.

Table 3.69 DH_t, A_E, W, B_E and V values kiwellin-like protein in V. vinifera obtained through *in silico* hydrolysis with chymase.

DH _t [%]: 12.2807					
No	Activity	A _E	W	B _E	V
1	antioxidative	0.0044	0.1679	0	
2	ACE inhibitor	0.0131	0.0357	4.85808 E-6	0.00041
3	dipeptidyl peptidase IV inhibitor	0.0175	0.0348	3.38364 E-5	0.09023

As can be seen from Table 3.70 when the protein is hydrolysed by papain, the degree of hydrolysis value was found high, thirty-one bioactive peptides with five properties were obtained.

Table 3.70 DH_t, A_E, W, B_E and V values kiwellin-like protein in V. vinifera obtained through *in silico* hydrolysis with papain

DH _t [%]: 35.9649					
No	Activity	A _E	W	B _E	V
1	antioxidative	0.0131	0.5000	0	
2	ACE inhibitor	0.0480	0.1309	0.00294	0.24911
3	hypotensive	0.0087	1.0000	0.00024	1
4	inhibitor	0.0044	1.0000	0	
5	dipeptidyl peptidase IV inhibitor	0.0611	0.1217	4.53475 E-5	0.12092

As can be seen from the Table 3.71, the value of hydrolysis was found 43.4211. Because of enzymatic cleavage value was high twenty-four bioactive fragments were

obtained. These bioactive peptides are shown four different activities and max A_E value was found in the ACE inhibitor.

Table 3.71 DH_t , A_E , W , B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with ficin

DH_t [%]: 43.4211					
No	Activity	A_E	W	B_E	V
1	ACE inhibitor	0.0393	0.1071	0.00087	0.07338
2	antioxidative	0.0044	0.1679	0	
3	stimulating	0.0044	0.1438	0	
4	dipeptidyl peptidase IV inhibitor	0.0568	0.1131	0.00015	0.40821

After the kiwellin like protein hydrolysed by leukocyte elastase, nineteen bioactive peptides were obtained. As can be seen from Table 3.72 these bioactive peptides have three activities.

Table 3.72 DH_t , A_E , W , B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with leukocyte elastase

DH_t [%]: 35.5263					
No	Activity	A_E	W	B_E	V
1	dipeptidyl peptidase IV inhibitor	0.0437	0.0870	1.17957 E-5	0.03145
2	activating ubiquitin-mediated proteolysis	0.0044	0.5057	0	
3	ACE inhibitor	0.0349	0.0951	1.89468 E-5	0.00160

As it can be understood from the results of *in silico* hydrolysis by metridin enzyme DH_t value was found to be 12.2807. Eight bioactive fragments showed three different activities (Table 3.73).

Table 3.73 DH_t , A_E , W , B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with metridin

DH_t [%]: 12.2807					
No	Activity	A_E	W	B_E	V
1	antioxidative	0.0044	0.1679	0	
2	ACE inhibitor	0.0131	0.0357	4.85808 E-6	0.00041
3	dipeptidyl peptidase IV inhibitor	0.0175	0.0348	3.38364 E-5	0.09023

Table 3.74 shows the bioactive protein parameters after pancreatic elastase II hydrolysis. According to results, the degree of hydrolysis value was found as 11.8421 and eight bioactive peptides have three properties.

Table 3.74 DH_t, A_E, W, B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with pancreatic elastase II

DH _t [%]: 11.8421					
No	Activity	A _E	W	B _E	V
1	antioxidative	0.0044	0.1679	0	
2	ACE inhibitor	0.0131	0.0357	4.85808 E-6	0.00041
3	dipeptidyl peptidase IV inhibitor	0.0175	0.0348	3.38364 E-5	0.09023

According to Table 3.75, bromelain is a suitable enzyme for hydrolysis of potential allergen protein. We observed significant increased number of bioactive peptides. Twenty-seven bioactive peptides were found and these peptides showed six properties.

Table 3.75 DH_t, A_E, W, B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with stem bromelain

DH _t [%]: 49.1228					
No	Activity	A _E	W	B _E	V
1	dipeptidyl peptidase IV inhibitor	0.0524	0.1043	3.72827 E-5	0.09942
2	antioxidative	0.0087	0.3321	0	
3	ACE inhibitor	0.0437	0.1191	0.00060	0.05065
4	hypotensive	0.0044	0.5057	0.00024	1
5	inhibitor	0.0044	1.0000	0	
6	stimulating	0.0044	0.1438	0	

A low DH_t value indicates that the oligopeptidase B enzyme is not suitable for hydrolysis. From the Table 3.76, we can see that only one bioactive peptide is released and the activity of ACE inhibitor peptide was very low.

Table 3.76 DH_t, A_E, W, B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* o hydrolysis with oligopeptidase B

DH _t [%]: 4.8246					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0044	0.0120	8.08669 E-7	6.84340 E-5

The result of *in silico* hydrolysis using calpain 2 is shown in Table 3.77. Nineteen bioactive peptides have three activities were obtained. Although ACE inhibitor peptide and DPP-IV inhibitor peptide had similar A_E values, ACE inhibitor peptide activity of fragments (B_E) value was highest.

Table 3.77 DH_t, A_E, W, B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with calpain 2

DH _t [%]: 42.5439					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0437	0.1191	0.00387	0.32763
2	antioxidative	0.0044	0.1679	0	
3	dipeptidyl peptidase IV inhibitor	0.0349	0.0695	1.48509 E-5	0.03960

Table 3.78 reveals the results related to hydrolysis by glycyl endopeptidase. According to Table 3.78, the degree of hydrolysis value was found to be 14.9123 and three bioactive peptides revealed ACE inhibitor properties.

Table 3.78 DH_t, A_E, W, B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with glycyl endopeptidase

DH _t [%]: 14.9123					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0131	0.0357	1.72063 E-6	0.00014

The results of hydrolysis of potential allergen protein in *V. vinifera* by oligopeptidase F enzyme are shown in Table 3.79. The degree of hydrolysis value was found to be 10.5263 and eight bioactive peptides have three properties.

Table 3.79 DH_t, A_E, W, B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with oligopeptidase F

DH _t [%]: 10.5263					
No	Activity	A _E	W	B _E	V
1	antioxidative	0.0044	0.1679	0	
2	ACE inhibitor	0.0131	0.0357	4.85808 E-6	0.00041
3	dipeptidyl peptidase IV inhibitor	0.0175	0.0348	3.38364 E-5	0.09023

Table 3.80 showed that 38.1579 DH_t value after proteinase P1 (lactocepin) cleavage. The results also showed that fragments are DPP-IV inhibitor, stimulating, antioxidative, dipeptidyl peptidase IV inhibitor.

Table 3.80 DH_t, A_E, W, B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with proteinase P1 (lactocepin)

DH _t [%]: 38.1579					
No	Activity	A _E	W	B _E	V
1	ACE inhibitor	0.0087	0.0237	0.00036	0.03009
2	stimulating	0.0044	0.1438	0	
3	antioxidative	0.0044	0.1679	0	
4	dipeptidyl peptidase IV inhibitor	0.0175	0.0348	5.19859 E-5	0.13863

When the protein is hydrolysed by pepsin (pH>2) the value of the degree of hydrolysis was found to be higher than other results. As can be seen from Table 3.81 thirty-six bioactive peptides with five different properties were obtained.

Table 3.81 DH_t, A_E, W, B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with pepsin (pH>2)

DH _t [%]: 63.5965					
No	Activity	A _E	W	B _E	V
1	dipeptidyl peptidase IV inhibitor	0.0873	0.1738	0.00017	0.46526
2	antioxidative	0.0044	0.1679	0	
3	activating ubiquitin-mediated proteolysis	0.0044	0.5057	0	
4	ACE inhibitor	0.0480	0.1309	0.00227	0.19183
5	stimulating	0.0131	0.4281	0	

As can be seen from Table 3.82 when the protein is hydrolysed by cocolysin enzyme, thirteen bioactive peptides with four different properties were obtained. These are DPP-IV inhibitor, stimulating, antioxidative, dipeptidyl peptidase IV inhibitor.

Table 3.82. DH_t , A_E , W , B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with cocolysin

DH_t [%]: 21.0526					
No	Activity	A_E	W	B_E	V
1	ACE inhibitor	0.0262	0.0714	7.30564 E-6	0.00062
2	antioxidative	0.0044	0.1679	0	
3	stimulating	0.0044	0.1438	0	
4	dipeptidyl peptidase IV inhibitor	0.0218	0.0434	0	0

Twenty-one bioactive peptides with five different properties were observed after subtilisin *in silico* hydrolysis.

Table 3.83 DH_t , A_E , W , B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with subtilisin

DH_t [%]: 25.4386					
No	Activity	A_E	W	B_E	V
1	dipeptidyl peptidase IV inhibitor	0.0349	0.0695	0.00015	0.40694
2	antioxidative	0.0087	0.3321	0	
3	ACE inhibitor	0.0349	0.0951	0.00313	0.26466
4	stimulating	0.0087	0.2843	0	
5	alpha-glucosidase inhibitor	0.0044	1.0000	0.00019	1

Table 3.84 reveals the results related to kiwellin-like protein in *V. vinifera* hydrolysis by V-8 protease (Glutamyl endopeptidase). The value of the degree of hydrolysis was found to be 9.2105 and eight bioactive with two different activities were observed. These are ACE inhibitor and DPP-IV inhibitor.

Table 3.84 DH_t , A_E , W , B_E and V values kiwellin-like protein in *V. vinifera* obtained through *in silico* hydrolysis with V-8 protease (Glutamyl endopeptidase)

DH_t [%]: 9.2105					
No	Activity	A_E	W	B_E	V
1	ACE inhibitor	0.0044	0.0120	0	0
2	dipeptidyl peptidase IV inhibitor	0.0044	0.0088	0	0

The BIOPEP tool also form

s a table containing the properties of the bioactive peptides, such as peptide ID, sequence, location, name, function, monoisotopic mass, and chemical mass. These tables are not shown in the thesis. When BIOPEP results are examined, it is seen that pepsin is the best-hydrolysed enzyme in three potential allergen proteins. The results of the three proteins show slight differences: for example, although *C. chinense* has 85% similarity, the prolyl oligo peptidase enzyme hydrolyses only the potential allergen protein found in the *C. chinense* species (in the table 3.38) and the V-protease enzyme does not hydrolysed the same protein.

A_E values (the frequency of release of fragments with given activity by selected enzymes) of three proteins were found to be relatively higher in *C.chinense*. Three proteins hydrolysed by the enzyme pH> 2 pepsin (EC 3.4.23.1) were found to have the highest value DH_t . In addition, the A_E values of the DPP-IV inhibitor peptides were highest in *V.vinifera*.

CHAPTER FOUR

CONCLUSION

Based on the tools used in the study, kiwellin-like proteins might show similar allergenicity in people who have kiwellin allergenicity. The identification of an unknown protein in a fruit sample by using *in silico* tools is so easier to estimate compared to wet-lab methodology. This strategy can reduce the cost of medicine and/or therapy costs spent for allergenicity. Therefore, there is a great need for development of novel allergenicity tools with better accuracy. Kiwi is an important and highly consumed fruit because of its rich ingredients such as vitamins and antioxidant molecules. Natural production place of kiwi is China. On the other hand, kiwi is also produced in Italy, New Zealand, Iran and Chile. According to FAOSTAT, total production of kiwi is 4,274,840 tons (FAOSTAT, 2018).

In the study conducted in 2012, it was mentioned that kiwi fruit has anticancer, antiviral, anti-inflammatory, antimicrobial, prebiotic, keratinocyte proliferation, and stimulant activity and has immune system regulating properties (Lim, 2012). Although kiwi is known as a healthy fruit, it has 13 allergen proteins in it, according to allergen.org. Kiwi has also been started to produce in different country. For instance, kiwi trees have been planted in the northern part of Turkey and now the production has significantly been increasing. Since some people have not consumed this fruit previously, people should be informed about the possible allergenicity of the allergen proteins of kiwi. In the present study, 10 kiwellin-like proteins have been studied in 10 different plants based on the similarities. It is very important to note that these proteins have still not been taken place in allergen.org.

Many scientific studies have been published on kiwi. Tamburrini et al (2005) purified kiwellin by ion exchange chromatography and defined it as an allergen protein. They also mentioned that kiwellin is one third of the total protein of kiwi fruit. In their study, it is selected because of high abundance in kiwi fruit compared to other allergen proteins. They proved its allergenicity by using Skin Prick test, western blot, specific IgE and total IgE tests.

Tuppo et al. (2008) also showed that Kissper part of kiwellin is a proteolysis-resistant protein and it constitutes pore forming in lipid membrane of cell. pH dependent and thionine containing Kissper also show its function in ion-channels. These functions cause allergenicity in human. The researchers suggest that after the Kissper removal, KiTH may be a secret IgE binding epitope, and the Kissper may be an IgE-binding epitope on its own. Considering the Clustal w results of 10 kiwellin-like proteins, we show the presence of protected regions and we think that IgE binding sites may be present in similar proteins.

Ciardiello et al (2008) identified two domains in kiwellin. First, it is known as Kissper consisted of first 39 amino acids and six of them is Cys. Kissper is located in the N-terminal. Second domain is known as KiTH and located in C-terminal. The residue number of this location is 40-189. Eight Cys are also reported for this residue. They mentioned that kiwellin based ion channel disruption can be associated with cystic fibrosis. Although the researchers mention that the high rate of Cys amino acids in the kiwellin, it was found that the percentage of Gly and Ser was higher in the seven kiwellin-like proteins that are identified as allergens.

Bernardi et al (2010) investigated the effect of protein conformational change in allergen-IgE interaction. Circular dichroism, immunoblotting, IgE reactivity based experiments were used to obtain the information about the protein conformation. They found out that Act d 5 is highly sensitive to the chemical environments.

Offerman et al (2015) reported X-crystallography of kiwellin protein in *A.deliciosa* (Act d 5). Offerman et al (2015) also explained pore-forming structure of kiwellin. The researchers mentioned that the properties of kiwellin are highly dependent on environmental conditions.

Hamiaux et al (2014) investigated crystal structure of kiwellin and they found that there is a binding region on the surface of kiwellin for endogenous ligands. They found Kissper part of kiwellin section has pore-forming activity in synthetic lipid layers and high flexibility.

Uberti et al (2015) studied 13 allergen proteins in kiwi. They defined three of them as major allergen proteins (Act d 1, Act d 2, Act d 6). Uberti et al (2015) defined act d

5 and Act d 8/11 as minor allergen proteins. Three major allergens were identified because of experiments conducted in the 10-person group of subjects. According to the study Act d 5 protein showed 40% allergic reaction.

Jenkins et al (2015) analysed the plant genomes to study allergen proteins. It is reported that 65% of the food allergens are originated from four different protein family.

According to an interesting paper by Ciacci et al (2013), they explained antioxidant and anti-inflammatory effects of Kissper peptide.

As it is seen in the study conducted in 2016, bioinformatics tools were used not only to investigate food allergens but also to investigate insect allergy (Yang et al., 2016). Yang and colleagues used three bioinformatics tools to obtained eight peptides that cause cockroach allergy in their study and they predicted T cell epitope sequences of Per a 10 allergen.

Agyei and Danquah (2012) reported on the importance of bioactive peptides and their effects on the human body in their study in 2012. They mentioned that peptides with antihypertensive, antimicrobial and immunomodulatory properties could be used as a preventive or controlling agent against some disease such as hypertension, cancer, cardiovascular diseases, diabetes, osteoporosis, stress and obesity. Also they reported that immunomodulatory peptides may be included in food and drug formulations that may be consumed additionally. In our study, we were obtained many bioactive peptides by hydrolysing potential allergen proteins. The use of these bioactive peptides from allergen proteins is thought to be more useful than any non-allergic source.

Consequently, kiwellin like proteins can be existed in not only kiwi but also in different fruits. By using *in silico* tools, it is easier to define possible allergen proteins. Since *in silico* tools have recently been developed, more input will be released in near future. The outputs from *in silico*-based investigations will most likely decrease number of allergen-based disorders. More scientific researches will be needed for development of new *in silico* tools and applications to find allergen proteins in foods. Developments in the field of artificial intelligence will most likely to increase the quality of *in silico* tools and applications in near future.

REFERENCES

- Agyei, D., & Danquah, M. K. (2012). Rethinking food-derived bioactive peptides for antimicrobial and immunomodulatory activities. *Trends in Food Science & Technology*, 23(2), 62-69.
- Ahmed, J., Preissner, S., Dunkel, M., Worth, C. L., Eckert, A. & Preissner, R. (2011). Super Sweet a resource on natural and artificial sweetening agents. *Nucleic Acids Research*, 39, D377-D382.
- Algpred (2019). Prediction of allergenic proteins and mapping of IgE epitopes. Retrieved July 27, 2019 from <http://crdd.osdd.net/raghava/algpred/submission.html>.
- Anthony, S., C., Masuyer, G., D Sturrock, E., & R Acharya, K. (2012). Structure based drug design of angiotensin-I converting enzyme inhibitors. *Current medicinal chemistry*, 19(6), 845-855.
- Attwood, T. K., Gisel, A., Eriksson, N. E., & Bongcam-Rudloff, E. (2011). Concepts, historical milestones and the central place of bioinformatics in modern biology: A European perspective. In *Bioinformatics-Trends and Methodologies*. IntechOpen.
- Ballereau, S., Glaab, E., Kolodkin, A., Chaiboonchoe, A., Biryukov, M., Vlassis, N., & Schneider, R. (2013). Functional genomics, proteomics, metabolomics and bioinformatics for systems biology. In *Systems Biology* 3, 41. Dordrecht, Springer.
- Baumgartner, C., Osl, M., Netzer, M., & Baumgartner, D. (2011). Bioinformatics-driven search for metabolic biomarkers in disease. *Journal of Clinical Bioinformatics*, 1(1), 2.
- Benkert, P., Biasini, M. & Schwede, T. (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 27, 343–350.

- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic Acids Research*, 41, 36-42.
- Bergholz, T. M., Switt, A. I. M. & Wiedmann, M. (2014). Omics approaches in food safety: fulfilling the promise? *Trends in Microbiology*, 22, 275-281.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., & Weissig, H. (2000). The Protein Data Bank, *Nucleic Acids Research*, 28, 235-242.
- Bernstein, K.E., Ong, F.S., Blackwell, W.L., Shah, K.H., Giani, J.F., Gonzalez-Villalobos, R.A., Shen, X.Z., Fuchs, S., & Touyz, R.M. (2012). A modern understanding of the traditional and non-traditional biological functions of angiotensin-converting enzyme. *Pharmacological Reviews*, 65, 1–46.
- Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L. & Schwede, T. (2017). Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. *Scientific Reports*. 7, 10480.
- Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., & Bordoli, L. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Resources*, 42, 252–258.
- Biasini, M., Schmidt, T., Bienert, S., Mariani, V., Studer, G., Haas, J., Johner, N., Schenk, A. D., Philippsen, A. & Schwede, T. (2013). OpenStructure: an integrated software framework for computational structural biology. *Acta Crystallographica Section D: Biological Crystallography*, 69, 701–709.
- Bienert, S., Waterhouse, A., de Beer, T. A., Tauriello, G., Studer, G., Bordoli, L., & Schwede, T. (2016). The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Research*, 45, 1, 313-319.

- Bjorklund, A.K., Soeria-Atmadja, D., Zorzet, A., Hammerling, U. & Gustafsson, M.G. (2005). Supervised identification of allergenrepresentative peptides for in silico detection of potentially allergenic proteins. *Bioinformatics*, 21, 39–50.
- BLAST (2019). *Basic Local Alignment Tool*. Retrieved July 27, 2019 from <http://blast.ncbi.nlm.nih.gov/>
- Boyce, J. A., Assa'ad, A., Burks, A. W., Jones, S. M., Beck, L. A., Byrd-Bredbenner, C., & Eichenfield, L. (2010). “Guidelines for the diagnosis and management of food allergy in the United States: summary of the NIAID-Sponsored expert panel report”, *Journal of Allergy and Clinical Immunology*, 126(6), 1105-1118.
- Charoentong, P., Angelova, M., Efremova, M., Gallasch, R., Hackl, H., & Galon, J. (2012). Bioinformatics for cancer immunology and immunotherapy. *Cancer Immunology, Immunotherapy*, 61, 1885-1903
- Chen, L., Lucas, J. S., Hourihane, J. O., Lindemann, J., Taylor, S. L., & Goodman, R. E. (2006). Evaluation of IgE binding to proteins of hardy (*Actinidia arguta*), gold (*Actinidia chinensis*) and green (*Actinidia deliciosa*) kiwifruits and processed hardy kiwifruit concentrate, using sera of individuals with food allergies to green kiwifruit. *Food and Chemical Toxicology*, 44(7), 1100-1107.
- Chen, L., Morrow, J. K., Tran, H. T., Phatak, S. S., Du-Cuny, L. & Zhang, S. (2012). From laptop to benchtop to bedside: structure-based drug design on protein targets. *Current Pharmaceutical Design*, 18, 1217.
- Chial, H. (2008). DNA sequencing technologies key to the Human Genome Project. *Nature Education*, 1, 219.
- Ciechanover, A. & Schwartz, A. L. (1989). How are substrates recognized by the Ubiquitin-mediated proteolytic system? *Trends in Biochemical Science*. 14, 483–488.

- Clustal w (2019). *Multiple Sequence Alignment*. Retrieved July 27, 2019 from <https://www.ebi.ac.uk/Tools/msa/clustalo/>
- Coates D (Jun 2003). The angiotensin converting enzyme (ACE). *The International Journal of Biochemistry & Cell Biology*. 35 (6): 769–73.
- Cooper, K. K., Mandrell, R. E., Louie, J. W., Korlach, J., Clark, T. A., Parker, C. T. (2014a). Comparative genomics of enterohemorrhagic *Escherichia coli* 77 O145: H28 demonstrates a common evolutionary lineage with *Escherichia coli* O157: H7. *BMC Genomics*, 15 (1), 17.
- Cooper, K. K., Mandrell, R. E., Louie, J. W., Korlach, J., Clark, T. A., Parker, C. T. (2014b). Complete genome sequences of two *Escherichia coli* O145: H28 outbreak strains of food origin. *Genome Announcements*, 2 (3), e00482-14.
- Ddjb (2019). *DNA Data Bank of Japan*. Retrieved July 27, 2019, from www.ddbj.nig.ac.jp/.
- Derks, K. W., Hoeijmakers, J. H. & Pothof, J. (2014). The DNA damage response: The omics era and its impact. *DNA Repair*, 19, 214-220.
- Dimitrov, I., Flower, D. R. & Doytchinova, I. (2013). AllerTOP-a server for in silico prediction of allergens. *BMC Bioinformatics*, 14, S4.
- Dziuba, M., Dziuba, B. & Iwaniak, A. (2009). Milk proteins as precursors of bioactive peptides. *Acta Scientiarum Polonorum Technologia Alimentaria*, 8, 71-90.
- Edwards, A. W. (1964). Reconstruction. *Phenetic and Phylogenetic Classification*, 6, 67-76.

Ehlers, M.R. & Riordan, J.F. (1989). Angiotensin-converting enzyme: new concepts concerning its biological role. *Biochemistry* 28, 5311–5318.

EMBL (2019). European Molecular Biology Laboratory. Retrieved July 27, 2019 from <http://www.ebi.ac.uk/>

Entrez (2019). Retrieved July 27, 2019, from www.ncbi.nlm.nih.gov/gquery/.

Erdmann, K., Cheung, B. W. & Schröder, H. (2008). The possible roles of food derived bioactive peptides in reducing the risk of cardiovascular disease. *The Journal of Nutritional Biochemistry*, 19, 643-654.

Eriksson, U., Danilczyk, U., & Penninger, J.M. (2002). Just the beginning: novel functions for angiotensin-converting enzymes. *Current. Biology*. 12, R745–R752.

ExPASy (2019). Retrieved July 27, 2019, from <http://www.expasy.org/>.

Fang, Y. Z., Yang, S. & Wu, G. (2002). Free radicals, antioxidants, and nutrition. *Nutrition*, 18, 872-879.

FASTA (2019). Retrieved July 27, 2019 from <http://blast.ncbi.nlm.nih.gov/>

Fiers, M. W., Kleter, G. A., Nijland, H., Peijnenburg, A. A., Nap, J. P. & Van Ham, R. C. (2004). Allermatch™, a web tool for the prediction of potential allergenicity according to current FAO/WHO Codex alimentarius guidelines. *BMC Bioinformatics*, 5 (1), 133.

FitzGerald, R. J., Murray, B. A., & Walsh, D. J. (2004). Hypotensive peptides from milk proteins. *Journal of Nutrition*, 134, 980S–988S.

Garcia-Horsman, J. A., Männistö, P. T. & Venäläinen, J. I. (2007). On the role of prolyl oligopeptidase in health and disease. *Neuropeptides*, 41(1), 1-24.

- Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook* (571-607). Humana Press.
- Genbank (2019). Retrieved July 27, 2019, from <http://www.ncbi.nlm.nih.gov/Genbank/>
- Gill, S. C., & Von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Analytical Biochemistry*, 182(2), 319-326.
- Gobbetti, M., Minervini, F., & Rizzello, C. G. (2004). Angiotensin I converting-enzyme-inhibitory and antimicrobial bioactive peptides. *International Journal of Dairy Technology*, 57, 172–188.
- Gu, Y., Majumder, K. & Wu, J. (2011). QSAR-aided in silico approach in evaluation of food proteins as precursors of ACE inhibitory peptides. *Food Research International*, 44, 2465-2474.
- Guruprasad, K., Reddy, B. V. B., & Pandit, M. W. (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Engineering, Design and Selection*. 4, 155–161.
- Hagen, J. B. (2000). The origins of bioinformatics. *Nature Reviews Genetics*, 1, 231-236.
- Hamiaux, C., Maddumage, R., Middleditch, M. J., Prakash, R., Brummell, D. A., Baker, E. N., & Atkinson, R. G. (2014). Crystal structure of kiwellin, a major cell-wall protein from kiwifruit. *Journal of structural biology*, 187(3), 276-281.
- Hogeweg, P. (2011). The Roots of Bioinformatics in Theoretical Biology. *PLoS Computational Biology*, 7 (3), e1002021.

- Holton, T. A., Vijayakumar, V. & Khaldi, N. (2013). Bioinformatics: Current perspectives and future directions for food and nutritional research facilitated by a Food-Wiki database. *Trends in Food Science & Technology*, 34 (1), 5-17.
- Horgan, R. P. & Kenny, L. C. (2011). Omic technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetrician & Gynaecologist*, 13 (3), 189-195. <http://www.intechopen.com/books/bioinformatics-trends-and-methodologies/>.
- Ikai, A.J. (1980). Thermostability and aliphatic index of globular proteins. *The Journal of Biochemistry*. 88, 1895-1898.
- INSDC (2019). International Nucleotide Sequence Database. Retrieved July 27, 2019, from <http://www.insdc.org/>.
- Ivanciuc, O., Schein, C. H. & Braun, W. (2003). SDAP: Database and computational tools for allergenic proteins. *Nucleic Acids Research*, 31 (1), 359-362.
- Iwaniak, A., Minkiewicz, P., & Darewicz, M. (2014). Food-originating ACE inhibitors, including antihypertensive peptides, as preventive food components in blood pressure reduction. *Comprehensive Reviews in Food Science and Food Safety*, 13(2), 114-134.
- Kandemir-Cavas, C., Pérez-Sanchez, H., Mert-Ozupek, N., & Cavas, L. (2019). In Silico Analysis of Bioactive Peptides in Invasive Sea Grass *Halophila stipulacea*. *Cells*, 8(6), 557.
- Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R. G., Wyckoff, H. & Phillips, D. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181, 662-666.

- Kim, J. H., Kim, S. I. & Song, K. S. (2001). Prolyl endopeptidase inhibitors from green tea. *Archives of Pharmacal Research*, 24, 292-296.
- Kitts, D. D., & Weiler, K. (2003). Bioactive proteins and peptides from food sources. Applications of bioprocesses used in isolation and recovery. *Current Pharmaceutical Design*, 9, 1309–1323.
- Korhonen, H., & Pihlanto, A. (2003). Food-derived bioactive peptides opportunities for designing future foods. *Current Pharmaceutical Design*, 9, 1297–1308.
- Korhonen, H., & Pihlanto, A. (2006). Bioactive peptides: production and functionality. *International Dairy Journal*, 16(9), 945-960.
- Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*. 157, 105–132.
- Lacroix, I. M. & Li-Chan, E. C. (2012). Evaluation of the potential of dietary proteins as precursors of dipeptidyl peptidase (DPP)-IV inhibitors by an in silico approach. *Journal of Functional Foods*, 4, 403-422.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, (2001). Initial sequencing and analysis of the human genome. *Nature*, 409, 860-921.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21).
- Lesk, A. M. (2005). Preface. *Introduction to Bioinformatics* New York: Oxford University Press Inc.

- Li, G., Le, G., Shi, Y., & Shrestha, S. (2004). Angiotensin I-converting enzyme inhibitory peptides derived from food proteins and their physiological and pharmacological effects. *Nutrition Research*, 24, 469–486.
- Lim, T. K. (2012). *Actinidia chinensis*. In *Edible Medicinal and Non-Medicinal Plants* (12-19). Dordrecht, Springer.
- Löw, P. (2011). The role of ubiquitin–proteasome system in ageing. *General and Comparative Endocrinology*, 172 (1), 39-43.
- Lucas, J. S., Lewis, S. A., Trewin, J. B., Grimshaw, K. E., Warner, J. O., & Hourihane, J. O. B. (2005). Comparison of the allergenicity of *Actinidia deliciosa* (kiwi fruit) and *Actinidia chinensis* (gold kiwi). *Paediatric Allergy and Immunology*, 16(8), 647-654.
- Luscombe, N. M., Greenbaum, D. & Gerstein, M. (2001). What is bioinformatics? A proposed definition and overview of the field. *Methods of Information in Medicine*, 40(04), 346-358.
- Madden, T. (2013). The BLAST sequence analysis tool. In *The NCBI Handbook [Internet]. 2nd edition*. National Center for Biotechnology Information (US); 2 January 2015.
- Masuyer, G., Yates, C. J., Sturrock, E. D., & Acharya, K. R. (2014). Angiotensin-I converting enzyme (ACE): structure, biological roles, and molecular basis for chloride ion dependence. *Biological Chemistry*, 395(10), 1135-1149.
- McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., & Buso, N. (2013). Analysis tool web services from the EMBL-EBI. *Nucleic Acids Research*, 41, 597-600.

- Meisel, H., & FitzGerald, R. J. (2003). Bio functional peptides from milk proteins: Mineral binding and cytomodulatory effects. *Current Pharmaceutical Design*, 9, 1289–1295.
- Minkiewicz, P., Dziuba, J. ve Michalska, J. (2011). Bovine meat proteins as potential precursors of biologically active peptides a computational study based on the BIOPEP database. *Food Science and Technology International*, 17 (1), 37–45.
- Minkiewicz, P., Dziuba, J., Iwaniak, A., Dziuba, M. ve Darewicz, M. (2008). BIOPEP database and other programs for processing bioactive peptide sequences. *Journal of AOAC International*, 91, 965-980.
- Mokharraq, W., Al Khalaf, N., & Altman, T. (2012). Application of Bioinformatics and Data Mining in Cancer Prediction. *In Proceedings of the International Conference on Information and Knowledge Engineering (IKE) (1)*. The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).
- NCBI (2019). National Centre for Biotechnology Information. Retrieved July 27, 2019, from www.ncbi.nlm.nih.gov/protein
- Nelson, D. L. & Cox, M. M. (2004). Amino acids, peptides and proteins. *Lehninger Principles of biochemistry* (4th ed.) (75-115). USA: W. H. Freeman.
- Nongonierma, A. B. & FitzGerald, R. J. (2013). Dipeptidyl peptidase IV inhibitory and antioxidative properties of milk protein-derived dipeptides and hydrolysates. *Peptides*, 39, 157-163.
- Offermann, L. R., Giangrieco, I., Perdue, M. L., Zuzzi, S., Santoro, M., Tamburrini, M., & Chruszcz, M. (2015). Elusive structural, functional, and immunological features of Act d 5, the green kiwifruit kiwellin. *Journal of agricultural and food chemistry*, 63(29), 6567-6576.

- Osborne, N. J., Koplin, J. J., Martin, P. E., Gurrin, L. C., Lowe, A. J., Matheson, M. C., & Allen, K. J. (2011). Prevalence of challenge-proven IgE-mediated food allergy using population-based sampling and predetermined challenge criteria in infants. *Journal of Allergy and Clinical Immunology*, 127(3), 668-676.
- Pastorello, E. A., Conti, A., Pravettoni, V., Farioli, L., Rivolta, F., Ansaloni, R., & Ortolani, C. (1998). Identification of actinidin as the major allergen of kiwi fruit. *Journal of Allergy and Clinical Immunology*, 101(4), 531-537.
- PDB (2019). Retrieved July 27, 2019, from <http://www.rcsb.org/pdb/home/home.do>
- Pearson, W. R. (2013). An introduction to sequence similarity (homology) searching. *Current Protocols in Bioinformatics*, 42(1), 3-1.
- Phelan, M. & Kerins, D. (2011). The potential role of milk-derived peptides in cardiovascular disease. *Food & Function*, 2, 153-167.
- Pihlanto, A., & Korhonen, H. (2003). Bioactive peptides and proteins. In S. L. Taylor (Ed.), *Advances in Food and Nutrition Research*, 47, (175–276). Elsevier Inc, San Diego, USA.
- Power, O., Jakeman, P. & FitzGerald, R. J. (2013). Antioxidative peptides: enzymatic production, in vitro and in vivo antioxidant activity and potential applications of milk-derived antioxidative peptides. *Amino Acids*, 44, 797-820.
- Protparam (2019). Retrieved July 27, 2019 from <http://www.expasy.org/tools/protparam.html>
- Reuter, S., Gupta, S. C., Chaturvedi, M. M. & Aggarwal, B. B. (2010). Oxidative stress, inflammation, and cancer: how are they linked? *Free Radical Biology and Medicine*, 49, 1603-1616.

- Rojas-Ronquillo, R., Cruz-Guerrero, A., Flores-Nájera, A., Rodríguez-Serrano, G., Gómez-Ruiz, L., & Reyes-Grajeda, J. P., (2012). Antithrombotic and angiotensin-converting enzyme inhibitory properties of peptides released from bovine casein by *Lactobacillus casei* Shirota. *International Dairy Journal*, 26 (2), 147-154.
- Ryan, J. T., Ross, R. P., Bolton, D., Fitzgerald, G. F. & Stanton, C. (2011). Bioactive peptides from muscle sources: meat and fish. *Nutrients*, 3, 765-791.
- Saha, S., & Raghava, G. P. S. (2006). AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic acids research*, 34, 202-209.
- Sampson, H. A., O'mahony, L., Burks, A. W., Plaut, M., Lack, G., & Akdis, C. A. (2018). Mechanisms of food allergy. *Journal of Allergy and Clinical Immunology*, 141(1), 11-19.
- Sanger, F. & Tuppy, H. (1951). The amino-acid sequence in the phenylalanyl chain of insulin. 1. The identification of lower peptides from partial hydrolysates. *Biochemistry Journal*, 49, 463–481.
- Sawey, E. T., Chanrion, M., Cai, C., Wu, G., Zhang, J., Zender, L., & French, D. M. (2011). Identification of a therapeutic strategy targeting amplified FGF19 in liver cancer by Oncogenomic screening. *Cancer Cell*, 19(3), 347-358.
- Shulaev, V. (2006). Metabolomics technology and bioinformatics. *Briefings in Bioinformatics*, 7, 128-139.
- Sicherer, S. H., & Sampson, H. A. (2018). Food allergy: a review and update on epidemiology, pathogenesis, diagnosis, prevention, and management. *Journal of Allergy and Clinical Immunology*, 141(1), 41-58.

- Sørensen, R., Kildal, E., Stepaniak, L., Pripp, A. H. & Sørhaug, T. (2004). Screening for peptides from fish and cheese inhibitory to prolyl endopeptidase. *Food/Nahrung*, 48 (1), 53-56.
- Tamburrini, M., Cerasuolo, I., Carratore, V., Stanziola, A. A., Zofra, S., Romano, L., & Ciardiello, M. A. (2005). Kiwellin, a novel protein from kiwi fruit. Purification, biochemical characterization and identification as an allergen. *The protein journal*, 24(7-8), 423-429.
- Tateno, Y., Imanishi, T., Miyazaki, S., Fukami-Kobayashi, K., Saitou, N., & Sugawara, H. (2002). DNA Data Bank of Japan (DDBJ) for genome scale research in life science. *Nucleic Acids Research*, 30, 27-30.
- Trifonov, E. N. (2000). Earliest pages of bioinformatics. *Bioinformatics*, 16, 5-9.
- Tuppo, L., Giangrieco, I., Palazzo, P., Bernardi, M. L., Scala, E., Carratore, V., & Ciardiello, M. A. (2008). Kiwellin, a modular protein from green and gold kiwi fruits: evidence of in vivo and in vitro processing and IgE binding. *Journal of Agricultural and Food Chemistry*, 56(10), 3812-3817.
- Turner, A.J. & Hooper, N.M. (2002). The angiotensin-converting enzyme gene family: genomics and pharmacology. *Trends in Pharmacological Sciences*. 23, 177–183.
- Uberti, F., Peñas, E., Manzoni, Y., di Lorenzo, C., Ballabio, C., Fiocchi, A., & Restani, P. (2015). Molecular characterization of allergens in raw and processed kiwifruit. *Pediatric Allergy and Immunology*, 26(2), 139-144.
- Udenigwe, C. C. & Aluko, R. E. (2012). Food protein-derived bioactive peptides: Production, processing, and potential health benefits. *Journal of Food Science*, 77 (1), 11-24.

- Udenigwe, C. C. & Mohan, A. (2014). Mechanisms of food protein-derived antihypertensive peptides other than ACE inhibition. *Journal of Functional Foods*, 8, 45-52.
- Udenigwe, C. C., Gong, M. & Wu, S. (2013). In silico analysis of the large and small subunits of cereal RuBisCO as precursors of cryptic bioactive peptides. *Process Biochemistry*, 48, 1794-1799.
- Uenishi, H., Kabuki, T., Seto, Y., Serizawa, A. & Nakajima, H. (2012). Isolation and identification of casein-derived dipeptidyl-peptidase 4 (DPP-4)-inhibitory peptide LPQNIPPL from gouda-type cheese and its effect on plasma glucose in rats, *International Dairy Journal*, 22 (1), 24-30.
- UniProt (2019). Retrieved July 27, 2019, from <http://www.uniprot.org/>
- UWM (2019). Retrieved July 27, 2019 from www.uwm.edu.pl/biochemia
- Valdes, A. M., Glass, D., & Spector, T. D. (2013). Omics technologies and the study of human ageing. *Nature Reviews Genetics*, 14(9), 601.
- Valdés, A., Simó, C., Ibáñez, C. & García-Cañas, V. (2013). Foodomics strategies for the analysis of transgenic foods. *TrAC Trends in Analytical Chemistry*, 52, 2-15.
- Vermeirssen, V., van Camp, J., & Verstraete, W. (2004). Bioavailability of angiotensin I-converting enzyme inhibitory peptides. *British Journal of Nutrition*, 92, 357–366.
- Wang, X. & Liotta, L. A. (2011). Clinical bioinformatics: A new emerging science. *Journal of Clinical Bioinformatics*, 1, 1.
- Watson, J.D. & Crick, F.H.C. (1953). Genetical implications of the structure of deoxyribonucleic acid. *Nature*, 171, 964–967.

- Wiener, A., Shudler, M., Levit, A. & Niv, M. Y. (2012). BitterDB: a database of bitter compounds. *Nucleic Acids Research*, 40 (1), 413-419.
- Wilson, J., Hayes, M. & Carney, B. (2011). Angiotensin-I-converting enzyme and prolyl endopeptidase inhibitory peptides from natural sources with a focus on marine processing by-products. *Food Chemistry*, 129, 235-244.
- Wu, D., & Wang, X. (2015). Application of clinical bioinformatics in lung cancer-specific biomarkers. *Cancer and Metastasis Reviews*, 34(2), 209-216.
- Yamamoto, N., Ejiri, M., & Mizuno, S. (2003). Biogenic peptides and their potential use. *Current Pharmaceutical Design*, 9, 1345–1355.
- Yanai, T., Suzuki, Y. & Sato, M. (2003). Prolyl endopeptidase inhibitory peptides in wine. *Bioscience, Biotechnology, and Biochemistry*, 67, 380-382.
- Yang, H., Chen, H., Jin, M., Xie, H., He, S., & Wei, J. F. (2016). Molecular cloning, expression, IgE binding activities and in silico epitope prediction of Per a 9 allergens of the American cockroach. *International journal of molecular medicine*, 38(6), 1795-1805.
- Yang, W. G., Wang, Z. & Xu, S. Y. (2007). A new method for determination of antithrombotic activity of egg white protein hydrolysate by microplate reader. *Chinese Chemical Letters*, 18, 449-451.
- Zhang R, Xu X, Chen T, Li L, & Rao P (May 2000). An assay for angiotensin-converting enzyme using capillary zone electrophoresis. *Analytical Biochemistry*. 280 (2): 286–90.
- Zou, W. (2014). Bioinformatics analysis of *Salmonella* flic gene diversity from next generation sequencing data. *International Association for Food Protection (Iafp)*. 2014, Indianapolis.