T.C.
MANISA CELAL BAYAR UNIVERSITY
THE GRADUATE SCHOOL



M.Sc. THESIS
DEPARTMENT OF BIOENGINEERING
SCIENCE OF BIOENGINEERING

BREAST CANCER MODELLING WITH GRAPE SEED FLAVONOIDS: A MOLECULAR DOCKING, MOLECULAR DYNAMICS SIMULATION (MDS) AND DENSITY FUNCTIONAL THEORY (DFT) STUDY

Ezgi SOYLU

MANISA–2024

LETTER OF COMMITMENT

This thesis was written in Manisa Celal Bayar University The Graduate School, Department of Bioengineering, in accordance with the academic and ethical rules, and all the literature information used is included in the thesis by reference. It is entirely my own work. I declare that I do not use artificial intelligence or programs.

Ezgi SOYLU

ÖZET

Yüksek Lisans Tezi

Üzüm Çekirdeği Flavonoidleri ile Meme Kanseri Modellemesi: Bir Moleküler Yerleştirme, Moleküler Dinamik Simülasyonu ve Yoğunluk Fonksiyonel Teorisi Çalışması

Ezgi SOYLU

Manisa Celal Bayar Üniversitesi
Lisansüstü Eğitim Enstitüsü
Biyomühendislik Anabilim Dalı

Bu tezde, üzüm çekirdeği flavonoidlerinin meme kanseri üzerindeki potansiyel etkilerini araştırmak amacıyla moleküler kenetlenme, moleküler dinamik simülasyonları (MDS) ve yoğunluk fonksiyonel teorisi (DFT) kullanılmıştır. Bu tez çalışmasında, meme kanseriyle ilişkili 7UJM (ER), 1R5K (ER), 4IFI (BRCA1) ve 1N0W (BRCA2) olmak üzere dört önemli reseptör hedeflenmiştir. Bu reseptörler, meme kanseri tedavisinde önemli biyomoleküler hedefler olarak seçilmiştir. Moleküler kenetlenme çalışmaları, üzüm çekirdeği flavonoidlerinin ve bilinen anti-kanser ilaçlarının bu reseptörlerle etkileşimlerini incelemek için gerçekleştirilmiştir. Çalışmanın ilk aşamasında, reseptör-ligand kompleksleri arasında en düşük bağlanma enerjisine sahip kompleksler belirlenmiştir. Elde edilen sonuçlar, üzüm çekirdeği flavonoidlerinin, bazı mevcut ilaçlarla karşılaştırıldığında, hedef reseptörlerle güçlü etkileşimler gösterdiğini ortaya koymuştur. Moleküler dinamik simülasyonları, seçilen ligand-reseptör komplekslerinin kararlılığını değerlendirmek amacıyla gerçekleştirilmiştir. Simülasyon sonuçları, flavonoidlerin reseptörlerle stabil kompleksler oluşturduğunu ve bu komplekslerin kararlı olduğunu göstermiştir. Ayrıca, yoğunluk fonksiyonel teorisi (DFT) kullanılarak, bu komplekslerin elektronik özellikleri detaylı bir şekilde incelenmiştir. Sonuç olarak, bu tez çalışması, üzüm çekirdeği flavonoidlerinin meme kanseri tedavisinde potansiyel birer terapötik ajan olarak kullanılabileceğini öne sürmektedir.

Anahtar Kelimeler: moleküler kenetlenme, moleküler dinamik simülasyon, meme kanseri, yoğunluk fonksiyonel teorisi, üzüm çekirdeği flavonoidleri, ilaç etkileşimi.

2024, 156 sayfa

ABSTRACT

M.Sc. Thesis

Breast Cancer Modelling with Grape Seed Flavonoids: A Molecular Docking, Molecular Dynamics Simulation (MDS), and Density Functional Theory (DFT) Study

Ezgi SOYLU

Manisa Celal Bayar University
The Graduate School
Department of Bioengineering

In this thesis, molecular docking, molecular dynamics simulations (MDS) and density functional theory (DFT) were used to investigate the potential effects of grape seed flavonoids on breast cancer. In this thesis study, four important receptors related to breast cancer were targeted: 7UJM (ER), 1R5K (ER), 4IFI (BRCA1) and 1N0W (BRCA2). These receptors have been selected as important biomolecular targets in the treatment of breast cancer. Molecular docking studies have been conducted to study the interactions of grape seed flavonoids and known anti-cancer drugs with these receptors. In the first stage of the study, the complexes with the lowest binding energy among the receptor-ligand complexes were identified. The results obtained have revealed that grape seed flavonoids show strong interactions with target receptors, compared to some existing drugs. Molecular dynamics simulations were performed to evaluate the stability of the selected ligand-receptor complexes. The simulation results showed that flavonoids form stable complexes with receptors and that these complexes are stable. In addition, the electronic properties of these complexes have been studied in detail using density functional theory (DFT). As a result, this thesis study suggests that grape seed flavonoids can be used as a potential therapeutic agent in the treatment of breast cancer.

Keywords: molecular docking, molecular dynamics simulation, breast cancer, density functional theory, grape seed flavonoids, drug interaction.

2024, 156 pages

FOREWORD AND ACKNOWLEDGMENTS

I wish to extend my sincere gratitude to the distinguished professors and administrative staff of the Manisa Celal Bayar University Department of Bioengineering, and The Graduate School for granting me the opportunity to undertake and complete this thesis.

Ezgi SOYLU
Manisa, 2024

SYMBOLS AND ABBREVATIONS INDEX

AC	Adriamycin Cyclophosphamide
AC-T	Adriamycin Cyclophosphamide- Taxol
AIs	Aromatase Inhibitors
AJCC	American Joint Committee on Cancer
ALA	Alanine
ARG	Arginine
ASN	Asparagine
ASP	Aspartic Acid
BC	Before Christ
BCS	Breast Conserving Surgery
BRCA1	Breast Cancer 1 Gene
BRCA2	Breast Cancer 2 Gene
BSE	Breast-Self Examinations
CAF	Cyclophosphamide Adriamycin Fluorouracil
ChEMBL	Chemical Database of Bioactive Molecules
CMF	Cyclophosphamide Methotrexate Fluorouracil
Coul-SR	Coulombic potential-Short Range
CPU	Cental Processing Unit
cryo EM	cryoelectron microscopy
CT	Computed Tomography
DCIS	Ductal Carcinoma *In Situ*
DFT	Density Functional Theory
DNA	Deoxyribonucleic Acid
EPR	Electron Paramagnetic Resonance
ER	Estrogen Receptor
ER+	Estrogen Receptor Positive
ER-	Estrogen Receptor Negative
FDA	Food and Drug Administration
FEC	Fluorouracil Hydrochloride Cyclophosphamide
FNAB	Fine Needle Aspiration Biopsy
GPU	Graphics Processing Unit
GLN	Glutamine

GLU	Glutamic Acid
HER2	Human Epidermal Growth Factor 2
HER2+	Human Epidermal Growth Factor 2 Positive
HF-SCF	Hartree-Fock Self Consistent Field
HIS	Histidine
IBC	Inflammatory Breast Cancer
IC50	Half Maximal Inhibitory Concentration
IDC	Invasive Ductal Carcinoma
IE	Interaction Energy
ILC	Invasive Lobular Carcinoma
ILE	Isoleucine
IR	Infrared Radiation
LCIS	Lobular Carcinoma *In Situ*
LDL	Low-density Lipoproteins
LEU	Leucine
LH	Luteinizing Hormone
LHRH	Luteinizing Hormone-Releasing Hormone
LJ-SR	Lennard-Jones poteintial-Short Range
LYS	Lysine
MD	Molecular Docking
MDS	Molecular Dynamic Simulation
MET	Methionine
MM	Mechanics Method
MRI	Magnetic Resonance Imaging
MSD	Mean Square Displacement
NMR	Nuclear Magnetic Resonance
OH	Hydroxyl
ORCA	Open-source Reaction Calculator
PBC	Periodic Boundary Conditions
PET-CT	Positron Emission Tomography
PDB ID	Protein Data Bank Identification Number
PDBQT	Protein Data Bank, Partial Charge (Q), Atom Type (T)
pH	Power of Hydrogen
PHE	Phenylalaline

PME	Particle Mesh Ewald
PR	Progesterone Receptor
PR+	Progesterone Receptor Positive
PR-	Progesterone Receptor Negative
PRO	Proline
PTEN	Phosphatase and Tensin Homolog
PubChem	Public Chemical Database
ρ	Density
RDF	Radial Distribution Functions
Rg	Gyration Radius
RMSD	Root Mean Square Deviation
RNA	Ribonucleic Acid
SASA	Solvent Accessible Surface Area
SER	Serine
SERDs	Selective Estrogen Receptor Degraders
SERMs	Selective Estrogen Modulators
TAC	Taxotere Adriamycin Cyclophosphamide
THR	Threonine
TNM	Tumor-Node-Metastasis
TRP	Tryptophan
TYR	Tyrosine
UCSF Chimera	University of California, San Francisco Chimera
UHeM	The National Center for High Performance Computing
UICC	Union International Center Cancer
UniProtKB	Universal Protein Resource Knowledgebase
USG	Ultrasonography
UV	Ultraviolet
VAL	Valine
WHO	World Health Organization

FIGURES INDEX

	Page
Figure 1. The interaction between flavonoids and breast cancer: Apoptosis.....	4
Figure 1.1. Classification of phonolic compounds and flavonoids.....	7
Figure 1.2. Types of flavonoids and basic flavonoid structure.....	8
Figure 2.1. Interaction between target and ligand.....	12
Figure 2.2. Schematic image of protein-ligand classification.....	13
Figure 2.3. The basic of MD simulation process	15
Figure 2.4. Interactional Density Functional Theory.....	19
Figure 2.5. Breast anatomy of a woman with a normal breast.....	28
Figure 2.6. Mammograohic view of breast anatomy	29
Figure 2.7. Demonstration of breast self examination positions.....	35
Figure 2.8. Needle core biopsy image of a patient with DCIS	39
Figure 2.9. MRI image of rib extension and skin ulceration in the left breast....	45
Figure 2.10. MRI image with lung metastases in the right breast.....	46
Figure 2.11. Schematic image of breast cancer treatment methods.....	50
Figure 2.12. PET-CT image of the patient with locally advanced breast cancer	53
Figure 3.1. Image showing the identification of hydrogen bonding with acceptor and donor atoms in a molecular dynamics simulation.....	69
Figure 3.2. Graph showing curves over different time periods for MSD calculation.....	72
Figure 4.1. Discovery Studio 2D ligand-protein interaction diagram showing ligand82 amino acid interactions with the 7UJM receptor	79
Figure 4.2. Discovery Studio 2D ligand-protein interaction diagram showing flav100 amino acid interactions with the 7UJM receptor.....	80
Figure 4.3. SwissADME results for potential drug candidacy of flav100 amino acid with 7UJM receptor.....	81
Figure 4.4. Discovery Studio 2D ligand-protein interaction diagram showing drug5 amino acid interactions with the 7UJM receptor.....	82
Figure 4.5. Discovery Studio 2D ligand-protein interaction diagram showing ligand29 amino acid interactions with the 1R5K receptor.....	86
Figure 4.6. Discovery Studio 2D ligand-protein interaction diagram showing flav93 amino acid interactions with the 1R5K receptor.....	87
Figure 4.7. SwissADME results for potential drug candidacy of flav93 amino acid with 1R5K receptor	88
Figure 4.8. Discovery Studio 2D ligand-protein interaction diagram showing drug16 amino acid interactions with the 1R5K receptor	89
Figure 4.9. Discovery Studio 2D ligand-protein interaction diagram showing flav100 amino acid interactions with the 4IFI receptor	93
Figure 4.10. SwissADME results for potential drug candidacy of flav100 amino acid with 4IFI receptor.....	94
Figure 4.11. Discovery Studio 2D ligand-protein interaction diagram showing drug17 amino acid interactions with the 4IFI receptor.....	95
Figure 4.12. Discovery Studio 2D ligand-protein interaction diagram showing ligand70 amino acid interactions with the 1N0W receptor.....	99
Figure 4.13. Discovery Studio 2D ligand-protein interaction diagram showing flav83 amino acid interactions with the 1N0W receptor	100
Figure 4.14. SwissADME results for potential drug candidacy of flav83 amino acid with 1N0W receptor.....	101

Figure 4.15. Discovery Studio 2D ligand-protein interaction diagram showing drug17 amino acid interactions with the 1N0W receptor 102
Figure 4.16. RMSD analysis results graph of 7UJM receptor-bound ligands in MD simulation with apo state 105
Figure 4.17. RMSF analysis results graph of 7UJM receptor-bound ligands in MD simulation with apo state 106
Figure 4.18. Rg analysis results graph of 7UJM receptor-bound ligands in MD simulation with apo state 107
Figure 4.19. Hydrogen bonding analysis results graph of 7UJM receptor-bound CHAINA_ligand82 in MD simulation 108
Figure 4.20. Hydrogen bonding analysis results graph of 7UJM receptor-bound Epicathecin_ligand100 in MD simulation 109
Figure 4.21. Hydrogen bonding analysis results graph of 7UJM receptor-bound Drug_ligand5 in MD simulation 109
Figure 4.22. RDF analysis results graph of 7UJM receptor-bound ligands in MD simulation with apo state 111
Figure 4.23. RMSD analysis results graph of 1R5K receptor-bound ligands in MD simulation with apo state 112
Figure 4.24. RMSF analysis results graph of 1R5K receptor-bound ligands in MD simulation with apo state 113
Figure 4.25. Rg analysis results graph of 1R5K receptor-bound ligands in MD simulation with apo state 113
Figure 4.26. Hydrogen bonding analysis results graph of 1R5K receptor-bound CHAINA_ligand29 in MD simulation 114
Figure 4.27. Hydrogen bonding analysis results graph of 1R5K receptor-bound Drug_ligand16 in MD simulation 115
Figure 4.28. RDF analysis results graph of 1R5K receptor-bound ligands in MD simulation with apo state 116
Figure 4.29. RMSD analysis results graph of 4IFI receptor-bound ligands in MD simulation with apo state 117
Figure 4.30. RMSF analysis results graph of 4IFI receptor-bound ligands in MD simulation with apo state 118
Figure 4.31. Rg analysis results graph of 4IFI receptor-bound ligands in MD simulation with apo state 118
Figure 4.32. Hydrogen bonding analysis results graph of 4IFI receptor-bound Epicathecin_ligand100 in MD simulation 119
Figure 4.33. Hydrogen bonding analysis results graph of 4IFI receptor-bound Drug_ligand17 in MD simulation 119
Figure 4.34. RDF analysis results graph of 4IFI receptor-bound ligands in MD simulation with apo state 121
Figure 4.35. RMSD analysis results graph of 1N0W receptor-bound ligands in MD simulation with apo state 122
Figure 4.36. RMSF analysis results graph of 1N0W receptor-bound ligands in MD simulation with apo state 123
Figure 4.37. Rg analysis results graph of 1N0W receptor-bound ligands in MD simulation with apo state 124
Figure 4.38. Hydrogen bonding analysis results graph of 1N0W receptor-bound CHAINB_ligand70 in MD simulation 125
Figure 4.39. Hydrogen bonding analysis results graph of 1N0W receptor-bound Cathecin_ligand83 in MD simulation 126

Figure 4.40. Hydrogen bonding analysis results graph of 1N0W receptor-bound Drug_ligand17 in MD simulation 126
Figure 4.41. RDF analysis results graph of 1N0W receptor-bound ligands in MD simulation with apo state..... 128
Figure 4.42. Final geometries after DFT calculation for ligand/7UJM residues..... 129
Figure 4.43. Final geometries after DFT calculation for ligand/1R5K residues..... 130
Figure 4.44. Final geometries after DFT calculation for ligand/4IFI residues..... 131
Figure 4.45. Final geometries after DFT calculation for ligand/1N0W residues..... 132
Figure B.1. Protein-Ligand Complex EM.mdp file parameters..... 152
Figure B.2. Protein-Ligand Complex NVT.mdp file parameters..... 154
Figure B.3. Protein-Ligand Complex NPT.mdp file parameters 155
Figure B.4. Protein-Ligand Complex MD.mdp file parameters 156

TABLE INDEX

Page

Table 2.1. Interaction between target-ligand and image of molecular docking.. 10
Table 2.2. Issues about molecular modeling..... 12
Table 2.3. Breast cancer risk factors 27
Table 2.4. Symptoms of breast cancer 31
Table 2.5. TNM staging in breast cancer 50
Table 2.6. Cancer drugs approved by the FDA for breast cancer 57
Table 2.7. Chemotherapy drug combinations used in breast cancer treatment... 58
Table 2.8. Classification of drugs used in breast cancer hormone therapy..... 60
Table 3.1. 3D image of 7UJM and 1R5K receptors obtained from UCSF Chimera 61
Table 3.2. 3D image of 4IFI and 1N0W receptors obtained from UCSF Chimera 62
Table 3.3. Grape seed flavonoids, receptor-ligand and drug CHEMBL codes... 63
Table 3.4. Dimensions and centers of the Grid box of the receptors with AutoDock Vina 65
Table 3.5. Ligand-receptor complexes and CHEMBL ID 66
Table 3.6. Ligand files and GitHub GROMACS Protein-Ligand files and descriptions 67
Table 3.7. Flavonoid conformations with the lowest binding energy selected for ORCA analysis after Autodock Tools..... 73
Table 4.1. Ligand-receptor complexes with lowest binding energy with AutoDock Vina 75
Table 4.2. Grid parameter file of 7UJM receptor and ligands with AutoDock Tools..... 77
Table 4.3. Docking parameter file of 7UJM receptor and ligands with AutoDock Tools..... 77
Table 4.4. Rigid and blind docking results of 7UJM receptor with the best score 78
Table 4.5. Flexible docking results of 7UJM with the best score after AutoDock Vina 83
Table 4.6. Grid parameter file of 1R5K receptor and ligands with AutoDock Tools..... 84
Table 4.7. Docking parameter file of 1R5K receptor and ligands with AutoDock Tools..... 84
Table 4.8. Rigid and blind docking results of 1R5K receptor with the best score 85
Table 4.9. Flexible docking results of 1R5K with the best score after AutoDock Vina 90
Table 4.10. Grid parameter file of 4IFI receptor and ligands with AutoDock Tools..... 91
Table 4.11. Docking parameter file of 4IFI receptor and ligands with AutoDock Tools..... 92
Table 4.12. Rigid and blind docking results of 4IFI receptor with the best score 92
Table 4.13. Flexible docking results of 4IFI with the best score after AutoDock Vina 96

Table 4.14. Grid parameter file of 1N0W receptor and ligands with AutoDock Tools..... 97
Table 4.15. Docking parameter file of 1N0W receptor and ligands with AutoDock Tools..... 98
Table 4.16. Rigid and blind docking results of 1N0W receptor with the best score 98
Table 4.17. Flexible docking results of 1N0W with the best score after AutoDock Vina 103
Table 4.18. MD simulations parameters 103
Table 4.19. Protein-ligand interaction energy results of 7UJM receptor-bound ligands in MD simulation..... 110
Table 4.20. Protein-ligand interaction energy results of 1R5K receptor-bound ligands in MD simulation..... 115
Table 4.21. Protein-ligand interaction energy results of 4IFI receptor-bound ligands in MD simulation..... 120
Table 4.22. Protein-ligand interaction energy results of 1N0W receptor-bound ligands in MD simulation..... 127
Table 4.23. Orca interaction energy results after DFT geometry optimization .. 133
Table A.1. 7UJM and 1R5K receptors (ER) molecular docking ChEMBL codes and Autodock Vina docking results 138
Table A.2. 4IFI receptor (BRCA1 gene) molecular docking ChEMBL codes and Autodock Vina docking results 140
Table A.3. 1N0W receptor (BRCA2 gene) molecular docking ChEMBL codes and Autodock Vina docking results 141
Table A.4. 7UJM (ER), 1R5K (ER), 4IFI (BRCA1) and 1N0W (BRCA2) receptors molecular docking flavonoids and Autodock Vina docking results 142
Table A.5. 7UJM (ER), 1R5K (ER), 4IFI (BRCA1) and 1N0W (BRCA2) receptors molecular docking drugs and Autodock Vina docking results 145

CONTENTS

	Page
ÖZET	I
ABSTRACT	II
FOREWORD AND ACKNOWLEDGMENTS	III
SYMBOLS AND ABBREVATIONS INDEX	IV
FIGURES INDEX	VII
TABLES INDEX	X
CONTENTS	XII
INTRODUCTION	1
CHAPTER I	
ORGANIZATION OF THESIS	
1.1. Overview of Grape Seed Flavonids: Functions and Potential Effects	5
1.1.1. Types of Flavonoids in Grape Seeds	6
CHAPTER II	
GENERAL INFORMATIONS	
2.1. Molecular Docking and Basic Concepts	12
2.2. Interaction of Protein-Ligand Docking	13
2.3. Molecular Dynamics Simulation Techniques	14
2.3.1. Ewald Summation Techniques	17
2.3.2. Energy-Minimization Methods	18
2.3.3. Solvent Models	18
2.3.4. Thermostats	18
2.3.5. Periodic Boundary Conditions (PBC)	19
2.3.6. Particle Mesh Ewald Method	19
2.4. Density Functional Theory (DFT) and Its Applications	19
2.4.1. Fundementals of DFT	21
2.5. Interactions between Molecular Docking, Molecular Dynmacis Simulation, Density Functional Theory and Drug Development	22
2.6. Breast Cancer: Epidemiology and Risk Factors	24
2.6.1. Breast Anatomy and Breast Tissue	27
2.6.2. Clinical Symptoms of Breast Cancer	30
2.6.3. Early Detection Methods in Breast Cancer	32
2.6.4. Staging of Breast Cancer	38
2.6.4.1. Histological Classification of Breast Cancer	38
2.6.4.2. Pathological Classification of Breast Cancer	42
2.6.4.3. Molecular Claassification of Breast Cancer	43
2.6.4.4. TNM Staging Method in Breast Cancer	44
2.6.4.5. Stage Groups for Breast Cancer	47
2.6.5. Breast Cancer Treatment Methods	50
2.6.5.1. Surgical Methods	51
2.6.5.2. Radiotherapy	53
2.6.5.3. Chemotherapy	54
2.6.5.4. Hormone Therapy	58

CHAPTER III
MATERIALS AND METHODS

3.1. Grape Seed Flavonoids 62
3.2. Protein Preparation 63
3.3. Ligand Preparation 63
3.4. Protein-Ligand Docking 63
3.5. Molecular Docking Procedure with Autodock Tools & Autodock Vina 64
3.6. Molecular Dynamics Simulatios Procedure with GROMACS Tutorial 65
3.7. Molecular Dynamics Simulation Tutorial-5 Analysis..... 68
3.7.1. Clustering Analysis (Frame Selection) 68
3.7.2. RMSD (Root Mean Square Deviation)..... 68
3.7.3. RMSF (Root Mean Square Fluctuations) 69
3.7.4. Hydrogen Bond Analysis 69
3.7.5. Radius of Gyration (Rg) 70
3.7.6. Interaction Energy of Protetin Ligand (IE)..... 71
3.7.7. SASA (Solvent Accessible Surface Area) 71
3.7.8. Radial Distribution Funtions (RDF) 72
3.7.9. MSD (Mean Square Displacement) 72
3.8. Density Functional Theory (DFT) analysis with ORCA 5.0 73

CHAPTER IV
RESEARCH FINDINGS AND DISCUSSION

4.1. Molecular Docking Findings 75
4.1.1. Molecular Docking Studies with 7UJM (ER), Ligands and Drugs..... 76
4.1.2. Molecular Docking Studies with 1R5K (ER), Ligands and Drugs..... 83
4.1.3. Molecular Docking Studies with 4IFI (BRCA1), Ligands and Drugs..... 90
4.1.4. Molecular Docking Studies with 1N0W (BRCA2), Ligands and Drugs..... 96
4.2. Molecular Dynamics Simulations Results (Stability of Ligand Receptor Complexes) 103
4.2.1. Molecular Dynamics Simulations with 7UJM (ER), Ligands and Drugs 104
4.2.2. Molecular Dynamics Simulations with 1R5K (ER), Ligands and Drugs 111
4.2.3. Molecular Dynamics Simulations with 4IFI (BRCA1), Ligands and Drugs 116
4.2.4. Molecular Dynamics Simulations with 1N0W (BRCA2), Ligands and Drugs 121
4.3. Geometry Optimization Results for Selected Conformations..... 128
4.4. Comparison with Experimental Data 128

CHAPTER V
CONCLUSION AND SUGGESTIONS

5.1. Overall Conclusions 134
5.2. Suggestions and Final Thoughts 129
REFERENCES..... 136

APPENDICES 143
APPENDIX A. Molecular Docking Results with Autodock Vina..... 143
APPENDIX B. GROMACS Tutorial/ Protein-Ligand Complex..... 151

1. INTRODUCTION

Cancer occurs when body cells grow uncontrollably and continuously, and if left untreated, it has a fatal effect. Breast cancer occurs when cells in the breast grow and divide uncontrollably, forming a tissue mass called a tumor. B.C. The first written information about breast cancer was recorded in the Edwin Smith Papyrus from Ancient Egypt between 3000 and 2500 BC. This document is the oldest known surgical document in the world and it is stated that there is no successful treatment for the disease (Oldenburg, Meijers-Heijboer, Cornelisse, & Devilee, 2007). Breast cancer is among the most common and deadly diseases worldwide. A genetic disease, which is one of the significant health issues of our era, is not attributed to a single cause. Environmental factors include malnutrition, however, lifestyle factors such as obesity and lack of physical activity can affect cancer metabolism in different ways. It is known that these conditions are directly related to the risk of breast cancer (Harbeck et al., 2019; Veronesi, Boyle, Goldhirsch, Orecchia, & Viale, 2005).

Breast tissue; While it consists of breast lobes, fatty tissue, ligaments, cavities (sinuses), glands and milk ducts, breast cancer develops due to the rapid proliferation of cells in the milk ducts or milk glands in this tissue. Breast cancer, which can also be seen in men, is a type of cancer that is most common in women over the age of 50 around the world. The incidence of this type of cancer in men is 0.5-1%. While 80% of breast cancer cases, seen in one in every 8 women around the world, occur due to the uncontrolled proliferation of cells in the milk ducts, the remaining 20% develop in the milk glands, defined as invasive lobular carcinoma. Apart from all these, there are also different types of breast cancer such as medullary, mucinous and tubular. In the presence of breast cancer, cancer cells multiply over time and form a mass. After this situation, which occurs more slowly compared to other types of cancer, cancer cells can spread to the lymph nodes and then to different parts of the body through the bloodstream. As with many types of cancer, there is a possibility of metastasis in breast cancer. Metastasis refers to the spread of cancerous cells to other parts of the body. Tumors in breast cancer can be benign or malignant. These tumors, which can also be called cancerous or non-cancerous, cause different effects on the body. The proliferation of cells that lead to malignant tumors is usually quite slow in the initial period and does not cause symptoms. Therefore, the person often does not notice this

situation in the initial period. This causes the cancer to spread, that is, metastasize, first to the lymph nodes and then to different parts of the body through the circulatory system (Harbeck et al., 2019).

Additionally, according to World Health Organization (WHO) data; it is stated that after the age of 5, one of the top three causes of death in developed and developing countries is cancer, 10% of all deaths occur due to cancer, 6.4 million new cancer cases occur every year in the world and 4.8 million people die due to cancer also, it caused 670,000 deaths worldwide in 2022, and breast cancer is recorded as the most common type of cancer in women in 157 of 185 countries. In line with these data, 2.3 million women worldwide were diagnosed with breast cancer in 2022, and approximately half of breast cancer was seen in women who did not have any specific risk factors other than gender and age. Although many risk factors for breast cancer have been identified, the most common and important known factor is genetic mutations (Oldenburg et al., 2007; Organization, 2024).

Additionally, breast cancer 1 gene (BRCA1), which is known to be associated with genetic factors, and breast cancer 2 gene (BRCA2), p53, PTEN or other familial/genetic factors including gene mutations; It includes environmental factors such as radiotherapy applied to the thorax area before the age of 30, hormone replacement therapy, alcohol use, number of breast biopsies, atypical hyperplasia, dense breast tissue, and body mass index. In addition, estrogen and progesterone hormone receptors (ER & PR) on the surface of tumor cells play an important role in the treatment selection of breast cancer patients (Oldenburg et al., 2007; Yersal & Barutca, 2014).

Depending on the stage of the breast cancer cell, there are a wide variety of treatment methods such as Surgical Methods, Radiotherapy, Chemotherapy, Targeted Drug Therapy, Immunotherapy, and Hormonotherapy. Many of the modern drugs used in the treatment of many types of cancer to suppress the proliferation of relevant cancer cells are derived from plants. Studies have shown that phytochemical compounds, which are naturally occurring components in plant foods, have anticarcinogenic activity, and the development of new generation anticancer agents has become an important research area. Phytochemicals reduce the side effects and complications

caused by chemotherapeutic agents in cancer treatment. There are 264.949 clinical trials registered (Barrios, Reinert, & Werutsky, 2018). According to a clinical trial conducted on women with early-stage breast cancer, it was understood that 59% of the early-stage breast cancer population was not well represented in clinical studies (Rocque et al., 2021). In addition, the effect of the drug on the tumor at the molecular level has almost never been explained in clinical interventions. Only the risks and benefits of treatment options, such as the effectiveness of the drug and the healing process of the disease, are explained through analysis such as tissue and blood samples.

In order to eliminate the deficiencies in the studies in the literature and also to determine the optimum drug amount and conformation, breast cancer was modeled at the atomistic level in this master's thesis study. The first reason that makes the study unique is that Molecular Docking (MD), Molecular Dynamic Simulation (MDS) and Density Functional Theory (DFT) methods were used together. The second important reason is that grape seed receptors, which have been on the agenda for use in breast cancer treatment in recent years, were addressed for the first time in a computational study. Many of the modern drugs used in the treatment of many types of cancer to suppress the proliferation of relevant cancer cells are derived from plants. Studies have shown that phytochemical compounds, which are naturally occurring components in plant foods, have anticarcinogenic activity, and the development of new generation anticancer agents has become an important research area (Ferraz da Costa et al., 2020). Grape seeds are a rich source of flavonoids. The flavonoids here are gallic acid, monomeric flavon-3-ols (such as (+)-catechin, (-)- epicatechin) and dimeric, trimeric, tetrameric procyanidins. Proanthocyanidins are a group of polyphenolic bioflavonoids known to have high pharmacological activity and are abundant in grape seeds (Rydberg, Jørgensen, & Olsen, 2014). Although the strong effect of polyphenols in grape seed extract on cancer is known, there are limited studies on the beneficial roles of these compounds in clinical studies (Ferraz da Costa et al., 2020; Rydberg et al., 2014).

In a nutshell, one of the goals of this thesis is to generate data for clinical studies that will reduce the side effects and complications caused by chemotherapeutic agents by using synthetic receptors from phenolic compounds of phytochemicals obtained from plants used in modern medicines today. It has been understood that

phytochemical compounds, which are naturally occurring components in plant foods, have anticarcinogenic activity. It has been revealed whether grape seed receptors can be an alternative to clinical ones. While aiming to pioneer the development of new generation anticancer agents, breast cancer modeling and drug research were carried out with computerized calculation methods such as Molecular Docking, MDS and DFT, taking as a guide the clinical studies conducted on drugs without using physical resources throughout the thesis study. Thus, preliminary resources were created for clinical and experimental drug research for breast cancer. It is expected that the results obtained in this study will contribute significantly to the literature if they are supported by experimental and clinical studies.

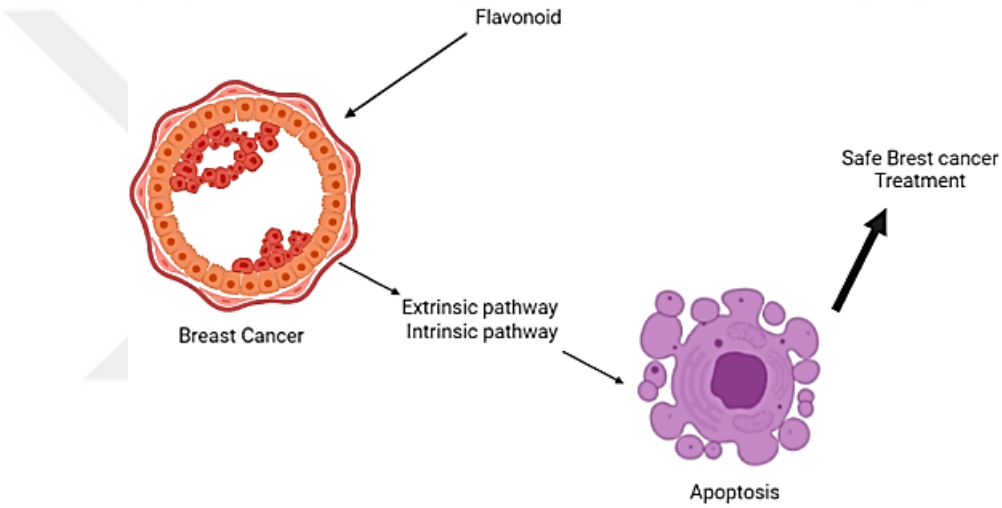


Figure 1. The interaction between flavonoids and breast cancer: Apoptosis (Park et al., 2022).

CHAPTER I. ORGANIZATION OF THESIS

1.1. Overview of Grape Seed Flavonoids: Functions and Potential Effects

Throughout history, people have processed plants from nature in different ways and used them as treatments for many diseases. Although developing technology and new treatment techniques come with new drugs, the dangerous side effects of these drugs cannot be ignored, so interest in scientific studies on plants has increased and research on their bioactive properties has deepened (Kursat M, 2009). The population that benefits from plants containing phytochemical components with various bioactive properties to treat/prevent health problems is 64% of the world's population, and this rate is higher in developing countries. The treatment approach called "Integrative Medicine" has gained more importance from past to present. The World Health Organization (WHO), in its report prepared based on scientific studies on medicinal plants conducted by researchers from 91 countries, reported that there are approximately 20,000 species of plants used in the treatment of diseases (Mukherjee, Kumar, & Houghton, 2007; Sheeja & Kuttan, 2007).

Grape (*Vitis vinifera*), which belongs to the Vitaceae family, is one of the most grown fruit species in the World. Grape seeds obtained from grape processes are an important by-product. This unfermented raw material subsequently became a valuable material for polyphenol extraction. Seeds constitute approximately 20% of the fruit weight (Jayaprakasha, Selvi, & Sakariah, 2003; Shi, Yu, Pohorly, & Kakuda, 2003).

The pharmacological benefits and nutritional benefits of grape seeds come from the free radical scavenging capacity of the polyphenols it contains. It has been reported in many studies that grape seed polyphenols reduce the risk of cancer and heart diseases by inhibiting the oxidation of low-density lipoproteins (LDL) (Jayaprakasha et al., 2003; Shi et al., 2003). Polyphenols (procyanidins) have antioxidative, free radical scavenging, cardioprotective, anticarcinogenic, antihypertension, antimutagenic, antiviral, antiallergic, immune system supportive properties and inhibit some undesirable enzyme activities. In addition, grape seed polyphenols show antibacterial and antiulcer activity (Fuleki & Ricardo da Silva, 1997; Jayaprakasha et al., 2003; Shi et al., 2003).

1.1.1. Types of Flavonoids in Grape Seeds

It has been reported that consumption of plant-based foods such as fruits and vegetables with high polyphenol content reduces the incidence of cancer. Studies conducted with cell culture and experimental animals have provided a better understanding of the anticarcinogenic properties of polyphenols. Polyphenols; It has been shown to prevent oxidation, inflammation and metastasis, induce apoptosis, improve immune system functions and affect signal transduction pathways related to the progression of carcinogenesis. In addition, polyphenols are found in parts of plants such as flowers, leaves, fruits, stems and roots. Flavonoids are divided into four classes: phenolic acids, stilbenes and lignans (Ilaiyaraja & Khanum, 2011).

The use of plants for medicinal purposes has increased with the discovery that the chemical compounds used in medicine are similar to the plant active ingredients. While the use of herbal medicines was more common in the past; As a result of the development of chemical applications, this rate has gradually decreased; However, in recent years, with the research and development of new areas of use of plants for therapeutic purposes, the demand for natural herbal products has increased due to the rich chemical structures of plants, their use to develop new and highly effective drug formulations has become one of the research areas of pharmacology (Ilaiyaraja & Khanum, 2011).

Phenolic compounds or polyphenols are organic compounds containing one or more hydroxyl groups on the benzene ring. Phenolic compounds commonly found in plants have many biological effects, including antioxidant activity. Phenolic compounds, or more commonly known as polyphenols, contain a benzene ring. Hydroxy benzene, often referred to as phenol, is a compound containing one OH group attached to the benzene ring. All other phenolic compounds are derived from here. Phenolic compounds are basically divided into two main groups: phenolic acids and flavonoids. Phenolic acids and flavonoids consist of subgroups, as seen in Figure 1.1 (Lopes, Dias de Queiros, Ávila, Monteiro, & Macedo, 2017).

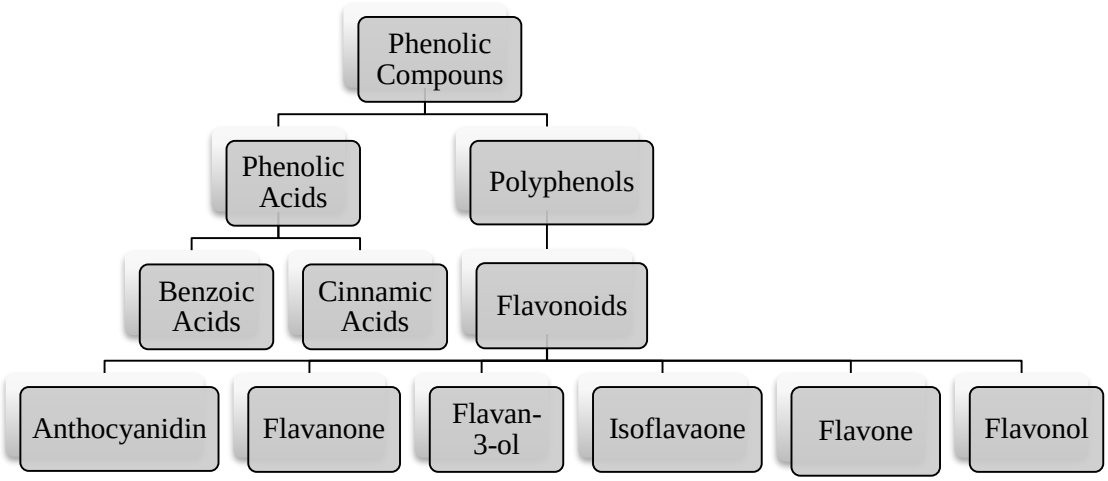


Figure 1.1. Classification of phenolic compounds and flavonoids (Lopes et al., 2017).

Phenolic acids are divided into two groups: hydroxybenzoic acid and hydroxycinnamic acid, both of which have a single ring structure. Hydroxybenzoic acid, gallic acid and ellagic acid can be classified as members of this group. This type of acid is found in onions, black radishes and especially berries. Hydroxycinnamic acid, caffeic acid, ferulic acid, p-coumaric acid and sinapic acid are examples of this group. Hydroxycinnamic acids are sensitive to temperature. These types of acids can be obtained from kiwi, apples, blueberries and cherries (Lopes et al., 2017; Nishiumi et al., 2011).

Flavonoids, which represent an important group of phenolics, have a relatively wide distribution in the plant flora. The difference between flavonoids arises from the number of hydroxyl groups in their structures, the degree of unsaturation and the oxidation level of the triple carbon segment. As shown in Figure 1.2, flavonoids generally consist of two benzene rings bonded heterocyclically with oxygen (flavone ring). They have a structure based on the diphenylpropane skeleton (C6-C3 C6). Additionally, its skeleton containing 15 carbon atoms is arranged in the C6-C3-C6 configuration (Nishiumi et al., 2011).

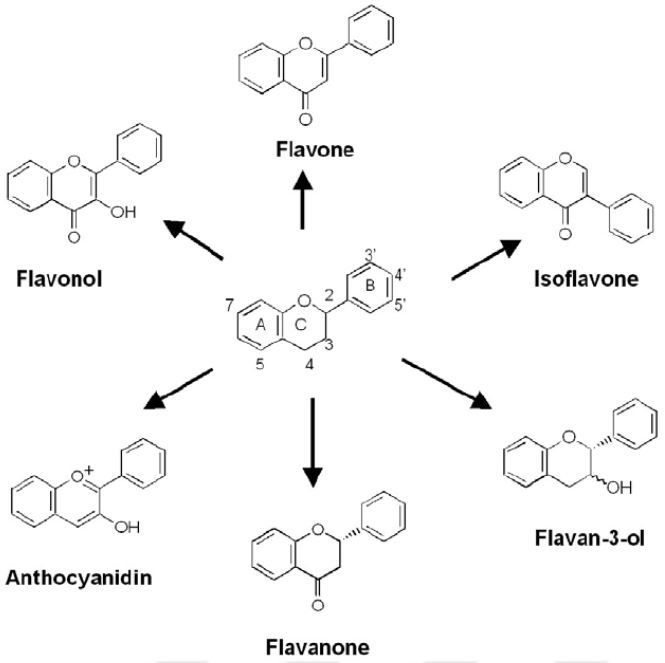


Figure 1.2. Types of flavonoids and basic flavonoid structure (Nishiumi et al., 2011).

Anthocyanidins are found as glycosides in sugars, and these flavonoids are the substances that create the red and blue color found in the skins of blue, red and black grapes. Flavones, which are phenolic groups containing one carboxyl group, consist mainly of apigenin and luteolin. Quercetin, kaempferol and myricetin are the most common flavones, and quercetin is found in glycoside form in grape skins. Flavanols can be found both in monomer form, such as catechins, and in polymer form, such as proanthocyanidins. The monomer form is found in apricots, while the polymer form can be obtained from red wine, grape seeds, black tea and chocolate (Nishiumi et al., 2011). Catechins, a group of flavanols, are widely found in the plant kingdom and are colorless. In terms of their chemical structure, they are flavan-3-ols. They contain an OH group on the third carbon atom. In addition, grape seeds are a rich source of flavonoids. The flavonoids herein are gallic acid, monomeric flavon-3-ols (such as (+)-catechin, (-)-epicatechin) and dimeric, trimeric, tetrameric procyanidins. Proanthocyanidins are a group of polyphenolic bioflavonoids known to have high pharmacological activity and are abundant in grape seeds. Although the strong effect of polyphenols in grape seed extract on cancer is known, there are limited studies on the beneficial roles of these compounds in clinical studies (Nishiumi et al., 2011).

CHAPTER II. GENERAL INFORMATIONS

Bioinformatics, which is an interdisciplinary science, is the process of examining biological information with the help of computers, which develops, organizes and analyzes biological data storage techniques and techniques for retrieving from storage. Also, the name bioinformatics was first used by Paulien Hogeweg in 1970 to study the knowledge of living systems, and bioinformatics uses computer engineering and statistics as well as biology science. In this way, the type of data processed by bioinformatics science, which gained momentum with the announcement of the human genome project in 2001, is mostly genetic data and related gene expression (Maloy & Hughes, 2013). The usage areas of bioinformatics are shown in Table 1 below;

Table 2.1. Interaction between target and ligand and image of molecular docking (Proteomics, 2024).

• The use of computers in molecular biology
• Three-dimensional graphic representation of molecules
• Molecular arrays
• Three-dimensional structure molecular databases
• Generating large amounts of data in a short dime
• Investigation of biologically active molecules
• Protein-protein relationship

The developments in the last 10 years, the most rapidly and effectively progressing, revolutionary projects, multidisciplinary projects carried out by large working groups in institutes or private companies, synthetic biology, bioengineering or cybernetics research, geneticists, cell biologists, chemists, engineers, bioinformatics and even physicians. showed that they work together (Antoraz, Santamaría, Díaz, Sanz, & Rodríguez, 2015; Maloy & Hughes, 2013).

Additionally, bioinformatics are people who process genetic data. They are expected to know at least one programming language, along with a very good genetic

background. In addition, in our country, there are some major bioinformatics companies, mostly together with biotechnology research. Also, bioinformatics workers are employed within pharmaceutical companies, research laboratories and universities (Antoraz et al., 2015; Maloy & Hughes, 2013).

Additionally, one of the first medical applications of bioinformatics is rational drug design. Possible amino acid sequences can be determined from nucleotide sequences using translation software. Later, homologous regions are determined with the help of sequence search techniques in model organisms, and human proteins can be modeled over experimentally characterized structures using sequence similarities. Thus, drugs can be designed with a molecule that can be connected to the model structure with the algorithm and biological activity can be tested on the real protein (Antoraz et al., 2015; Maloy & Hughes, 2013).

Molecular modelling covers all theoretical or computerized methods used to model or simulate the behavior of molecules. In this manner, simulation applications have begun to be used due to the methods and algorithms developed to obtain real-time and physically correct deformations of objects. Objects to be deformed such as human organs can be modeled superficially or volumetrically by molecular modeling using methods such as Mass-Spring Systems. On the other hand, molecular modeling method and technology is a method developed for surface modeling that works fast and produces realistic internal forces, and generally triangles are used in this modeling method (Antoraz et al., 2015; Gu & Li, 2011; Maloy & Hughes, 2013).

Additionally, mechanism-based drug design is the most ideal drug design method in which all studies are carried out at molecular level with biological pathways. Natural endogenous substances in the organism are effectors such as ligands. On the other hand, the biopolymers and macromolecules in which the ligands react in the organism are target molecules. The rational definition of structure-effect relationships began with the study of effector-target or ligand receptor interactions. Molecular modeling issues are shown in table 2.2 below (Antoraz et al., 2015; Gu & Li, 2011; Maloy & Hughes, 2013).

Table 2.2. Issues about molecular modelling.

Determination of physicochemical properties of molecules	Determination of hydrophobic, electronic, steric properties
The interaction of an enzyme with a substrate	Examining the compatibility of a molecule with the enzyme's active site and designing more effective molecules
IR, UV, NMR Spectra	Knowing the chemical structure, finding the spectrum
Chemical Reactivity	Nucleophilic and electrophilic regions
Molecular geometry	The shapes of molecules (bond lengths, angles, dihedral angles)

There are some methods used in molecular modelling. The first of these methods is the Ab-initio method. Two different mathematical approaches are used in this method. These methods are Hartree-Fock Self Consistent Field (HF-SCF) and Density Functional Theory (DFT). In the HF model, it is applied to calculate molecular frequencies and determine molecular geometry based on an average potential for electron-electron interactions. On the other hand, Density Functional Theory (DFT) is based on the Schrödinger equation, instead of molecular wave functions, the electron probability density (ρ) is calculated, it is a method that gives much more accurate results in the determination of molecular properties. Also, the Schrödinger equation is a function called the instrumental wave function that provides all information about a quantum system, and there are multiple electrons involved (Antoraz et al., 2015; Cygan, 2003; Gu & Li, 2011). The second method of molecular modeling is Molecular Mechanics Method (MM). It is a quick form of calculation that describes the interactions between atoms in a chemical system using the classical laws of mechanics. In this method, even large molecules like steroids and enzymes are quickly optimized within seconds. While docking studies are carried out with this method, features that depend on the electronic structure cannot be obtained by this method. In addition, even for large systems such as enzymes, quantities such as reaction temperature and conformational stability can be calculated (Antoraz et al., 2015; Cygan, 2003; Gu & Li, 2011). The third method of molecular modeling is Semi-empirical method. Quantum mechanics is used in this method, which is based on the Schrödinger

equation. Also, experimental values are also included in the calculations as parameters. On the other hand, this method is not as slow as Ab initio calculations, but not as fast as the molecular mechanics method. Semi-empirical method is between both methods (Antoraz et al., 2015; Cygan, 2003; Gu & Li, 2011).

2.1. Molecular Docking and Basic Concepts

Molecular docking method is the method of placing the molecule in receptor pockets. On the other hand, this method, which means interlock, involves the interaction of a molecule with another site, usually found in the protein structure. Also, it is applied by placing the molecule in certain parts of the protein structure, taking into account many quantities such as the electro negativity of the atoms, their relative positions, and the conformation of the molecule to be placed. However, it simulates the placement of the protein structure with certain score algorithms. Molecular docking studies play a major role in determining whether millions of compounds synthesized are effective drug active substances. The image of the molecular docking method resulting from the interaction between the target molecule and the ligand is shown in Figure 2.1 (Cygan, 2003; Morris & Lim-Wilby, 2008).

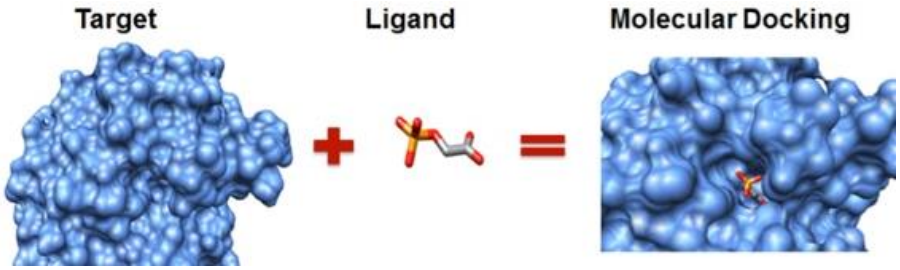


Figure 2.1. Interaction between target and ligand and image of molecular docking (Proteomics, 2024).

Molecular docking studies have a very important role in selecting the most effective substances, as it is impossible to study each of the millions of chemical substances individually in vitro. It is a method that saves a lot of money by transforming the effect of a chemical substance on a protein structure into points, preventing the *in vitro* and *in vivo* studies of molecules that are impossible to be effective. In addition, it establishes a strategic infrastructure for the synthesis of

molecules that are likely to be more effective by preparing the ground for the modification of the molecule with the correct estimation of the binding modes of the relevant molecule. The docking process plays an important role in explaining the receptor-ligand, enzyme-ligand relationship. Suitable antagonists have an important place in enzyme inhibition studies in finding its compounds (Cygan, 2003; Morris & Lim-Wilby, 2008).

2.2. Interaction of Protein-Ligand Docking

Molecular docking, also known as protein-ligand docking, is a computational chemistry method used in many areas. In this method, which stands out as one of the first steps in drug discovery, computer-aided drug design, which is the first step of drug discovery studies, includes molecular approximation calculations. The best examples are processes such as docking and drug interactions used in cancer treatment. Also, protein-ligand docking, that is, molecular docking, is one of the first methods used for drug discovery in computational chemistry. The classification of protein ligand docking methods is shown in table 3.1 (Huang & Zou, 2010; Proteomics, 2024).

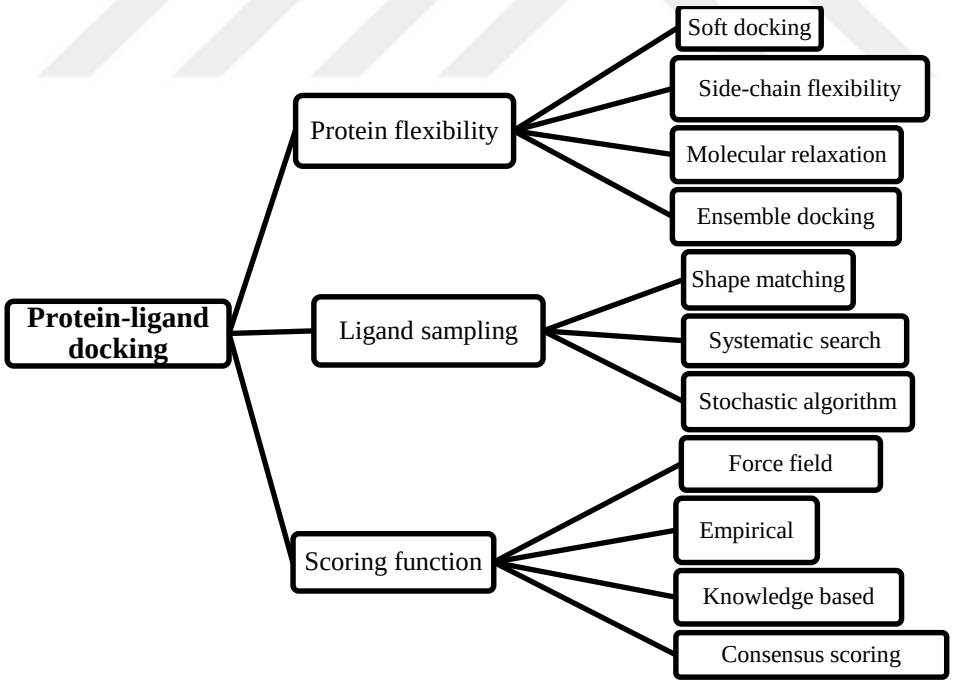


Figure 2.2. Schematic image of protein-ligand classification (Huang & Zou, 2010).

Intercalarily, in order for the protein structure to be suitable for protein-ligand coupling, the protein side must be prepared for this process. After the protein data is obtained, it should be checked whether there are missing amino acids in the chain and it should be made sure that there is no problem (Huang & Zou, 2010; Proteomics, 2024).

2.3. Molecular Dynamics Simulation Techniques

There are various tools that enable the accurate and rapid description of the physical properties of atomic-sized systems. Depending on the positions of all atoms in a biomolecular system, the force that these atoms exert on each other can be calculated. Molecular dynamics (MD) algorithms were developed towards the end of the 19th century and gained widespread use due to the rapid advancements in computer technology. The first realistic MD simulation was performed on liquid argon in 1964. With this technique, which is difficult to achieve through experimental methods, the movement and position of each atom can be tracked at every moment in time. Molecular dynamics (MD) simulation techniques were used for these calculations, enabling further development of the algorithms. The MD simulation method is based on Newton's second law or equation of motion, $F = m$a. Here, F is the external force acting on the particle, m is the mass, and a is the acceleration (Hospital, Goñi, Orozco, & Gelpí, 2015). It is possible to determine the acceleration of each atom in the system by using the force acting on each atom. MD simulation is an effective method used to study the physical motions of molecules Brreand atoms. This method is widely used to examine the structural, dynamic and thermodynamic properties of many high-tech materials such as semiconductors, metals, polymers and nanomaterials, metallic alloys. Atoms and molecules are allowed to interact for a certain period of time, giving a view of the dynamic evolution of the system. The resulting potential energies are calculated using molecular mechanics methods with different force fields (Hospital, Goñi, Orozco, & Gelpí, 2015).

MD simulations have become a powerful tool for studying biophysical systems, with the computational power and software availability chosen to investigate the properties of the system and its potential to accurately model the behavior of the atoms in the system. In MD simulation, a protein whose ligands are bound to it;

Simulation conditions such as initial conformation, mutations or post-translational modifications of other molecules in the environment, temperature, voltage value across a membrane must be known exactly, and one of the most important tasks is the selection of appropriate potential functions from which the forces are derived. Using the molecular mechanics force field model, the forces in the MD simulation are calculated via quantum mechanics. The main elements required to evaluate the mobility or flexibility of various regions of a biomolecule are flexibility, accuracy, providing accurate results at high temperatures and calculation speed. MD simulation not only demonstrates structural variation upon environmental changes such as pH, temperature, and mutations, but can also demonstrate the dynamic process of misfolding of proteins or peptides. Time steps in each simulation should be short, such as a few femtoseconds (10^{-15}s). Figure 2.3 shows a simple figure of the MD simulation process (Hospital et al., 2015; Lindahl, 2008).

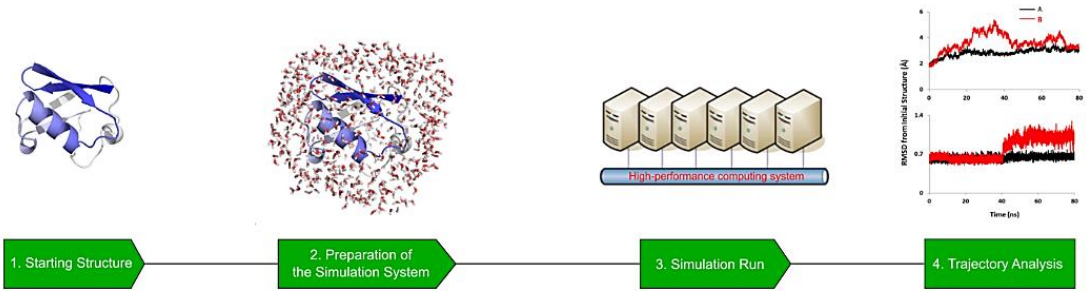


Figure 2.3. The basic of MD simulation process (Lindahl, 2008).

Most biochemical events (e.g., functionally important structural changes in proteins) occur over long periods of time, such as nanoseconds or microseconds, involving millions or billions of time steps. MD simulations reveal the positions of all atoms at femtosecond scale temporal resolution; It can capture a wide range of important biomolecular processes, such as conformational change, ligand binding, and protein folding, and predict how each atom in a protein or other molecular system will behave over time based on a general physics model that governs interatomic interactions. MD simulation, which provides an understanding of processes such as ligand separation and protein folding with high spatial resolution, is used together with structural techniques such as Forster resonance energy transfer, nuclear magnetic

resonance (NMR), X-ray crystallography, electron paramagnetic resonance (EPR) and cryoelectron microscopy (cryo EM). When MD simulation is used together with experimental structure determination methods such as X-ray crystallography and cryo-EM, the level of structural fluctuation of the equilibrium state in various regions of the molecule can be measured. In addition, simulation is used to test and refine the accuracy of a modeled structure by also revealing the dynamic behavior of water molecules and salt ions that are critical for protein function and ligand binding (Chan-Yao-Chong, Durand, & Ha-Duong, 2019).

In an MD-based protocol, simulation is used to generate high-resolution structures of individual components of the complex separately, to generate atomic-level molecular models from low-resolution cryo-EM density maps, and to recover ensembles of conformers in response to a single structure from NMR data. In order to see how a biomolecular system will react to some disruptions, the ligand in the protein is removed from the structure and it is checked how this deficiency affects the conformation of the protein (Chan-Yao-Chong et al., 2019).

MD simulation studies aim to observe functional processes. Biomolecular processes such as protein folding, membrane transport, ligand or voltage-dependent conformational change, and ligand binding are facilitated by MD simulations. Simulations generate hypotheses that lead to new experimental studies. Biophysical techniques that provide information about the structural dynamics of a biomolecule, such as EPR spectroscopy or NMR spectroscopy, changes in the chemical environment are reported (Chan-Yao-Chong et al., 2019).

MD simulations provide information about the functions of molecules within the cell by examining the movements of biological molecules at certain temperatures, pressures and salt concentrations. There are many simulation packages available to use MD simulations. The most commonly used simulation software packages include GROMACS, NAMD, AMBER, CHARMM, Desmond and OpenMM. The computing hardware used are GPU (Graphics Processing Unit) and CPU (Central Processing Unit). GPU is preferred because it performs faster simulations at lower cost. On the

other hand, it runs faster on supercomputers and traditional central processing units (CPUs). Most modern simulation software packages support multiple force fields. All of these software packages perform similar calculations but differ in how efficient the various hardware and supported features are (Chan-Yao-Chong et al., 2019).

In certain applications in ligand and protein design, complex structures can be tested experimentally through simulations. Simulations are used to create a measurable understanding of how the biomolecule or drug works. Simulations generate experiments that directly investigate structural properties by generating hypotheses that lead to new experimental studies and enable the structure to be solved in reality. In cases where simulations show that a particular protein-ligand interaction is important for binding, the relevant part of the protein or ligand can be modified. By checking its effect on the functional properties of the protein, the simulation of the structure is examined, and it can be measured how much the equilibrium state in various regions of the molecule moves and what kind of structural fluctuations it undergoes. How a biomolecular system will respond to some deteriorations is determined by MD simulation. After the ligand in an experimentally determined protein is removed from the structure, the system is simulated to observe how the absence of the ligand affects the protein conformation. Functional processes such as protein folding, ligand-dependent conformational change, and ligand binding are also simulated (Chan-Yao-Chong et al., 2019).

2.3.1. Ewald Summation Techniques

The Ewald Summation Technique, an MD simulation technique based on the Dynamo program written between 1982 and 1984, calculates long distances in real space by examining the interaction with electrostatic energies (Toukmaji & Board, 1996).

2.3.2. Energy-Minimization Techniques

Energy minimization, one of the important MD simulation techniques, first selects low-energy regions and follows a trend towards minimum energy geometries

in structures. While the lowest energy geometric structures of the clusters and the potential energy surface were scanned, the minimum energy geometries and isomers corresponding to the phase transition region were obtained using the microcanonical Molecular Dynamics (MD) thermal minimization method. Thus, it is based on obtaining the lowest energy geometry by slowly reducing the total energy until all the total energy of the system becomes potential energy (Bent, 2008) .

2.3.3. Solvent Models

Solvent models, which are MD simulation methods, require aqueous environments to operate the system. Water molecules are very much needed in MD simulations. Various water models such as SPC, SPC/E, TIP3P, TIP5P are used according to the physical properties of the water molecule such as its function, diffusion, density anomaly and radial distribution. The SPC model is a solvent model in which water molecules are simply represented, while the SPC/E model is an improved version of the SPC model. In addition, the TIP3P model is a solvent model used as another simple representation of water molecules, and while it works with the assumption that the water molecule contains charges at three different points, the TIP5P model offers a more complex structure than the TIP3P model. (Tanner, 2011).

2.3.4. Thermostats

In the MD simulation system, keeping the system temperature at a constant value **(**canonical ensemble**)** while examining the properties and dynamics of biomolecules is ensured by advanced thermostat algorithms such as Berendsen, Andersen, Nose-Hoover and Langevin. The purpose of the Berendsen thermostat is to smoothly harmonize the system temperature with the target temperature, while the purpose of the Andersen thermostat is to harmonize the speeds of the molecules with the target temperature. Also, while the purpose of the Nose Hoover thermostat is to keep the system temperature at the target temperature and to provide energy conservation principles, the purpose of the Langevin thermostat is to harmonize the speeds of the molecules with the target temperature and at the same time regulate the energy losses of the system (Shvab, 2014).

2.3.5. Periodic Boundary Conditions (PBC)

In Periodic Boundary Conditions used in MD Simulation, an infinite number of images are obtained in space after the system is placed in a unit cell. An infinite lattice is created with this 3D extension. Thus, surface effects are eliminated by simulation. At this stage, the box size and box types will be decided according to the system (Shvab, 2014).

2.3.6. Particle Mesh Ewald Method

Particle Mesh Ewald (PME) Method, one of the MD Simulation methods, works with an efficient and fast algorithm. This method, which is one of the most preferred methods, divides the interactions into short (fast decreasing) and long (slow decreasing) range interactions. Direct binary interactions between particles, short ranges are Real Space (Particle). Long range slowly decaying electrostatic interactions are Fourier Space (Mesh). In order to use the mesh method, the load must be neutral (de Souza & Ornstein, 1997).

2.4. Density Functional Theory (DFT) and Its Applications

Density Functional Theory (DFT); It is one of the widely used computational tools for investigating and predicting the properties of systems such as isolated molecules, crystalline solids, interfaces, and surfaces. In order to use naturally occurring or artificially produced materials in technology, it is important to know the properties of the materials well, and the properties of the molecules or atoms that make up the material can be calculated both experimentally and theoretically. By contributing to old research with new research on the material, the costs of new areas of use are reduced. All information about a quantum system can be obtained with the Schrödinger wave equation. For some systems wave functions are obtained by solving the Schrödinger equation and then the energy states of the system are determined. However, analytical and numerical solution of the Schrödinger equation of multi-electron systems is not yet possible (Ana M. Sanchez, 2023).

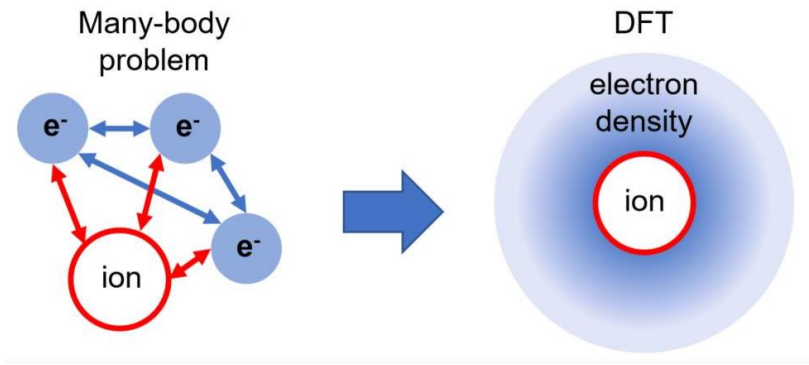


Figure 2.4. Interactional Density Functional Theory, which is created by converting the multi-electron system into electron density and relating the energy to this density (Posysaev, 2018).

In the 1964s, Hohenberg and Kohn developed a method of calculating using the electron density, which is a function of position and time, instead of solving the solution using the multi-electron wave function of the system. Calculation of the electronic structure of crystal structures by determining the ground state density by solving equations called Kohn-Sham equations is based on ab inito methods, which are based on Kohn-Sham equations and enable exact solutions of the quantum mechanical ground states of electron systems. Density functional theory, one of the most successful theories in determining the properties of matter, makes it possible to determine many properties such as the band structure of solids and the calculation of bond energy in molecular chemistry. In addition, superconductivity, interaction of atoms with laser, relativistic effect in heavy nucleated atoms, magnetic properties of classical liquids and alloys can be determined by DFT (Ana M. Sanchez, 2023).

The active role of density functional theory (DFT) in the search for new two- and one-dimensional materials has enabled the rapid development of materials science and the discovery of new one- or two-dimensional materials. It has emerged as a standard in solving the many-body problem for atomic-scale systems, and in recent years it has been used to examine the bulk and surface properties of solids as well as the physical properties of nanotubes. Density functional theory, one of the most successful theories in determining the properties of matter, makes it possible to determine many properties such as the band structure of solids and the calculation of bond energy in molecular chemistry. In addition, superconductivity, interaction of

atoms with laser, relativistic effect in heavy nucleated atoms, magnetic properties of classical liquids and alloys can be determined by this method (Ana M. Sanchez, 2023).

Calculations using density functional theory can be used to support experimental studies, as well as to lead studies that have not been conducted experimentally. While a compound has not yet been synthesized, many of its properties can be calculated theoretically, whether it is structurally stable or not, leading to experimental and engineering studies and making it easier to carry out high-cost studies. In addition, DFT allows research to be carried out in extreme conditions such as high pressure. Electronic, optical, elastic, structural and thermodynamic properties of many structures have been studied using Density functional theory and the results obtained from such studies have been found to be quite consistent with experimental results. Using density functional theory to examine multi-particle systems further accelerates the process (Ana M. Sanchez, 2023).

2.4.1. Fundementals of DFT

While the physical properties of materials are obtained depending on their electronic structures, information about the electronic structures of materials is obtained from the solution of the multi-particle Schrödinger equation. The first term in the Schrödinger wave equation; kinetic energy operator for electrons, the second term is the potential energy created by electron-electron interaction, and the third term is the external potential term created by the nucleus on the electrons. Solving the Schrödinger equation of a multi-particle system is almost impossible. For this reason, the equation is made solvable by making some approximations. Density Functional Theory (DFT) is one of the methods that enable the approximate solution of the multi-particle problem and is a method developed to explain the ground state properties and electronic structures of systems with electrons. In DFT, the fundamental quantity is the electron density rather than the multi-particle wave function (Ana M. Sanchez, 2023).

The purpose of DFT is; is to find the total energy of the system in terms of a functional of density without any reference wave function. The minimum value of this total energy corresponds to the ground state energy of the system. DFT started with the first studies, the Thomas and Fermi model. It was developed with the Hohenberg-

Kohn and Kohn-Sham theorems. It states that the exact ground state density of a system in external potential can be found by minimizing the energy functional. If the ground state of the system is not degenerate, there is a single charge density corresponding to the given external potential and minimizing the energy, giving the ground state energy. In addition, there is a single charge density that corresponds to the given external potential and minimizes the energy. Considering the existence of an auxiliary non-interacting system with the same electron density as the real system, Kohn and Sham propose that the electron density of the interacting system can be written in terms of single electron wave functions of the non-interacting system. The electronic and magnetic properties of materials and how these properties change with the effect of an external field can be effectively obtained by using DFT methods. The data obtained as a result of these calculations; magnetic stagnation, heat capacity, hysteresis curves etc. of the structure. It enables the determination of magnetic properties such as magnetic properties under different external influences (Ana M. Sanchez, 2023).

2.5. Interactions between Molecular Docking, Molecular Dynamics Simulation and Density Functional Theory and Drug Development

In every civilization, people have used medicines of plant or animal origin to prevent or cure disease. Since ancient times, fighting against disease has been a basic need, almost like searching for food and shelter. Organic synthesis methods have been developed to produce natural compounds that are biologically active but difficult to obtain in quantity. Since the cost of this is too high, developments in computational techniques and computer hardware have enabled *in silico* (computerized) methods. Computer-aided drug design, which is used in drug research and development by many of the world's leading pharmaceutical companies, now forms a part of modern technology. The binding interactions of drugs with the target macromolecule are mostly through intermolecular bonds, and electrostatic bonds are the most common bonds. Also, hydrogen bonds and Van der Waals interactions. Hydrogen bonds and Van der Waals interactions are used in the optimization of compounds that have potential as drug candidates. Substituents are identified and added to these interactions, and their effects on activity and selectivity are investigated, and their possible effects are investigated through studies such as structure-activity relationships, molecular docking and pharmacophore in the computer environment, before proceeding with the

synthesis of the molecules. Basically, computer-aided drug design methods and computer data banks are used. Structure/activity relationships, genetic algorithms, statistical tools, data analysis tools, and visualization methods are commonly used computer-aided drug design methods (Guimin & Zhu, 2016).

Additionally, one of the most important goals of computational biology is to determine the binding affinities of enzymes, proteins and small molecules. Molecular docking is used to determine the positions of small molecules in the active site of the target protein. Although docking methods have been providing more reliable results in recent years, thanks to the development of the parameters of the scoring methods used, the results of the scoring functions are not always in agreement with the experimental data. In this context, one of the aims of computational medicinal chemistry is to calculate the binding energy between ligand and enzyme (protein) with sensitive methods before drug synthesis (Guimin & Zhu, 2016).

The prominent areas in the field of computer-aided drug development in recent years are multi-functional drug development studies and the discovery of new molecules by quickly and accurately screening large ligand banks containing synthesized ready-made small molecules (such as ZINC, Otava) in the studied target structure, taking into account their interaction scores. Binding positions determined by docking are verified by MD simulations, and this combination provides more reliable and detailed binding models. In addition, MD simulation method is used to examine the movement of particles. Numerical solving of Newton's equation of motion forms the basis of this calculation. The interaction of each particle with each other is calculated, and as a result of the calculation, the change that the system undergoes over a certain period of time depending on external conditions is monitored. Atoms can interact with each other using potential energy functions based on experiments and observations, or various force fields chosen depending on the area of study, and each forces acting on the atoms are calculated for a particular configuration based on these functions or force fields. The effect of the molecular docked ligand in analyzing the relationship between protein and ligand structure interactions in drug design is supported by this method and presented with more precise results. The feature of these studies is that they produce results by taking into account the interaction of each atom in the entire structure with another atom. Mathematical functions are used in

computational chemistry to calculate chemical bonds and interactions on an energy basis. The potential energy of the system to be analyzed by computational chemistry is defined by force fields as a function of atomic positions. If the ligand docked to the protein structure prevents the movement of the protein structure, this protein cannot perform its function or performs it less than normal. After revealing the effectiveness of the active and catalytic regions of the proteins both by docking and simulations, the amino acids in the active and catalytic region, as well as the ligand molecules found there, and finally Density Functional Theory (DFT), one of the quantum chemical methods, are used for high-accuracy modeling. The electronic properties and energy calculations of the instantaneous structures obtained during MD simulations are calculated with DFT to obtain more detailed and precise energy results. Thus, binding affinity and binding energies are predicted more accurately (Guimin & Zhu, 2016).

2.6. Breast Cancer: Epidemiology and Risk Factors

Cancers occur when a single cell loses control of its normal growth and replication processes. Carcinogenesis takes place in 3 phases: initial, progress and development phases. Initial phase; it occurs with the first genetic mutation that results in the cell being exposed to an agent. However, this is not enough for the development of cancer. The initial cell must be activated with the advanced agent that causes cellular proliferation, and this phase is called the progression phase. In the development phase, the cells formed in the initial and progression phases constitute the tumor mass. At the end of carcinogenesis, the cell may have some or all of the characteristic features of cancer cells, such as independent growth signaling, insensitivity to anti-growth signals, avoidance of apoptosis, maintenance of angiogenesis, tissue invasion and metastasis. Factors related to food, nutrition and physical activity can affect many cellular processes, including carcinogenesis (Harbeck & Gnant, 2017).

The first written information about breast cancer dates back to B.C., known as the world's oldest surgical document. It is included in the Edwin Smith Papyrus dating back to 3000-2500 BC. In this ancient Egyptian text, it is reported that successful treatment of the disease is not possible. Breast cancer is among the most common and deadly diseases worldwide. According to research conducted by the World Health Organization (WHO) in recent years, it has been recorded that 2.3 million women were diagnosed with breast cancer and 685.000 women died of breast cancer in 2020 (Islam

et al., 2023). Environmental factors include malnutrition, however, lifestyle factors such as obesity and lack of physical activity can affect cancer metabolism in different ways. It, whose incidence is increasing in developed and developing countries, is a growing public health problem. Although breast cancer is rare in women under the age of 30, it is known that it increases more rapidly in the years following this age and this increase continues to rise slowly after menopause. Breast cancer, which ranks first with more than 4.4 million women living with breast cancer in the world and accounting for approximately 31% of all cancers seen in women, continues to be a universal health problem. According to the data of the American Cancer Society, it is stated that the risk of developing breast cancer in a lifetime is 1/8 in Canada and the USA, 1/5 in Australia, and 1/50 in Japan. The probability of an American woman developing breast cancer during her lifetime is calculated as 12.5%, and the probability of dying from breast cancer is 3.4% (Budny et al., 2019).

The risk of developing the disease is directly related to age, and the incidence of the disease gradually increases as age increases. Breast cancer is rare before the age of 30, with a rapid increase in the reproductive years following this age. Following a slight decrease during menopause, a steady increase occurs with a slow slope in the postmenopausal years. The frequency of breast cancer varies from country to country around the world. While Hawaii, California and Canada rank first with an incidence of 80-90 per hundred thousand per year, the same value is only between 12-15 per hundred thousand in Japan. This difference in disease frequency between countries in the world is especially seen in post-menopausal women, and the differences between countries are very small in the pre-menopausal period. Similarly, large differences in incidence are observed among different ethnic groups living in the same country and between white and black races, and this difference is attributed to environmental factors, lifestyles and socioeconomic status (Budny et al., 2019).

Breast cancer accounts for less than 1% of all organ cancers in men. Approximately 1500 new cases are reported annually in the USA, and the prevalence in Europe is stated to be 1/100,000. Although it is reported in the literature that cases are most commonly diagnosed between the ages of 60-70, cases are also encountered under the age of 30 and even in childhood. Male breast cancers are similar to female breast cancers in terms of clinical course and pathological features (Budny et al., 2019).

Additionally, changes have occurred in the structure of diseases due to the increase in the average age of societies, the extension of life expectancy and the emergence of developments in the field of preventive medicine. Cancer, which is considered one of these diseases, affects organs and systems in the body and is used for many different diseases. It is stated that it usually takes many years for the cancer to grow and increase the number of its cells to reveal signs and symptoms sufficient to prompt a person to consult a physician. The most important feature of cancers and malignant tumors is that they can invade nearby organs and tissues, spread, and break away from the original tumor and be carried to distant places in the body via blood. Cancer cells acting in this way can cause metastasis or new tumors called secondaries. In terms of cancer types, it is stated that breast cancer is the most common type of cancer in women. Raising women's awareness about breast cancer and its risk factors plays an important role in the fight against breast cancer. It is very important to understand these risk factors affecting the person and to determine, implement and develop early diagnosis methods. When the results obtained from breast cancer incidence and epidemiological studies are examined, it is seen that no single factor is responsible for the formation of breast cancer, but there are many different risk factors. The incidence of breast cancer, which is associated with many risk factors, varies depending on the decrease or increase in risk factors. These factors that increase the risk of breast cancer in women are listed as follows:

Table 2.3. Breast cancer risk factors (Sun et al., 2017).

• Aging
• Fertile age duration
• Ethnicity/race
• Reproduvtive factors
• Menopause age
• Breast cancer 1 gene (BRCA1)
• Breast cancer 2 gene (BRCA2)
• Familial genetic factors
• Excessive use of alcohol
• Familial genetic factors

• Number of breast biopsies
• Body mass index
• Excessive be on the birth control pills
• Smoking
• Fatty diet and resulting obesity

The most obvious risk factors for breast cancer are female gender and age. While it is possible to change some of the risk factors, some of them cannot be changed. Risk assessment models used today are based on combinations of risk factors and calculate the risk of breast cancer over a certain period and/or lifetime. Also, some of the risk assessment methods are Gail Model, Claus Model, Ford Model, BRCAPro, Bodian, Myriad, Tyrer-Cuzick, Manual Model, Rosner-Colditz, Couch Model and Hultson Murday Model (Kim & Bahl, 2021).

2.6.1. Breast Anatomy and Breast Tissue

In women, breast tissue is located on the chest wall above the muscle called the pectoral muscle. Its size varies from person to person and its shape and structure are specific to each person. Both breasts may be slightly different in size. Although breast tissue is present in both men and women, the mammary glands are functional only in the postpartum period and milk is secreted from the mammary glands to nourish the newborn baby. Breast tissue in humans consists of fat and connective tissue, as well as mammary glands. It is located between the superficial fascia of the lateral and anterior chest wall of the breast. Each breast, which is cone-shaped with a base diameter of 10-12 cm, weighs approximately 150-225 g. Characteristically, the breast is hemispherical in shape with an elliptical base. The architectural structure of the breast consists of lactiferous canals and lobes located radially around the nipple-areola structure (nipple-areola complex). Depending on embryological development, the upper outer quadrant of the breast contains more glandular structure than the other quadrants. Breast tissue consists of mammary glands, fatty tissue and connective tissue. It is the amount of fatty tissue that determines the actual size of the breast. Fibrous thickenings in the connective tissue extend into the parenchymal tissue of the breast, these fibrous structures are called Cooper ligaments (Astley Cooper Ligaments). Cooper ligaments provide support to the breast by extending within the

mammary gland from the deep fascia of the superficial fascia covering the pectoralis major muscle to the superficial layer under the skin. These suspensory ligaments are especially evident in the upper part of the breast. In patients with breast cancer, these ligaments are affected, causing an orange peel appearance on the skin. Cooper ligaments are of great importance in both the spread of cancer and the first signs of it. The breast consists of two parts: lobules (milk glands) and ducts (milk ducts). Support and fatty tissue fill the space between lobules and ducts. The milk-secreting part of the breast, lobules and ducts, open to the nipple in the dark colored area called the areola, located in the middle of the breast. Lobes are formed in the breast by the union of lobules. The area around the nipple consists of a circular pigment area called areola. In the aerola epithelium are small hairs, sebaceous and sweat glands, and accessory mammary glands. There are also motgomeri glands (small bumps) around the areola. Breast tissue is mostly found in the upper outer quadrant. The presence of large amounts of breast tissue in the upper outer quadrant, including the axillary region, causes more tumors to form in this area (Pandya & Moore, 2011).

In an adult woman, the breasts are cone-shaped. The shapes and sizes of breasts vary depending on the amount of fat tissue they contain. It is a common situation that there is a difference in size between both breasts. Generally, there is no underlying endocrine anomaly, but it should be investigated whether this size difference is due to a tumor. In most women, an increase in the size, density and nodularity of the breasts is detected before the onset of menstruation. The most important factor in this increase is fluid accumulation in the breasts. Total breast volume is at its lowest between the 6th and 15th days of the menstrual cycle. The average upper and lower diameter of the breast is 10-12 cm, and its maximum thickness in the central region is approximately 5-7 cm. The weight of a non-lactating breast is 150-200 g and in lactation it is 400-500 g (Pandya & Moore, 2011).

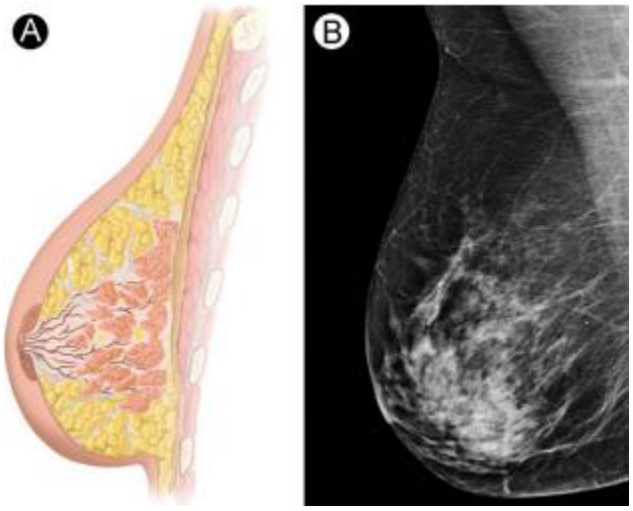


Figure 2.5. Breast anatomy of a woman with a normal breast: (A) vertical visualization (B) oblique mammographic view of a real and normal female breast (Jesinger, 2014).

Breast tissue consists of the following structures:

- Lobules: Mammary glands
- Ducts: Milk channels that carry the produced milk to the nipple. Fat and connective tissue: Fat and connective tissue surrounding the milk glands and ducts give the breast its shape and serve as support.
- Areola: It is the pink-brown circular structure around the nipple. It contains small sweat glands that make breastfeeding easier.
- Nipple: Milk ducts open here.

Breast tissue contains blood vessels and lymphatic vessels. Breast lymphatic vessels carry lymph fluid to the armpit lymph nodes. This is especially important in the treatment of breast cancer. Breast cancer is often carried to the armpit lymph nodes through these vessels. Breast lymph fluid may also drain into groups of internal mammary lymph nodes above the collarbone (supraclavicular), below the collarbone (subclavicular), and just below the breastbone. In the treatment of breast cancer, it is important to evaluate the spread to these lymph nodes. Throughout a woman's life, changes occur in adult breast tissue related to the menstrual period, pregnancy, lactation and menopause. Changes occurring in the breast, which is a part of the reproductive system, are under a strict hormonal control mechanism. Knowing these

normal changes that occur in the breast is important in distinguishing between pathological and physiological conditions. There are many hormones involved in breast development. Estrogen, progesterone, adrenal corticoids, prolactin, insulin, thyroxine and growth hormone coordinate breast development. The pituitary gland affects breast development by controlling the endocrine system. Progesterone has the greatest effect on the alveolar system. Prolactin affects physiological changes such as the differentiation of alveolar epithelial cells into mature milk-producing cells, as well as other structural changes. The breast has three developmental phases. The first stage of breast development occurs between birth and puberty. During this period, there is a gradual increase in glandular tissue and surrounding stroma. Other phases are the menstrual cycle and menopause (Jesinger, 2014).

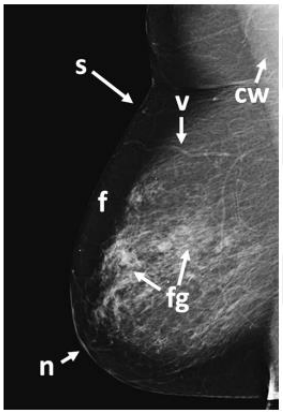


Figure 2.6. Mammographic view of breast anatomy: (cw) chest Wall. The breast is covered by skin (s) with a centrally positioned nipple (n). Within the breast is fat (f), a variable amount of fibroglandular breast tissue (fg), and neurovascular structures (v) (Jesinger, 2014).

2.6.2. Clinical Symptoms of Breast Cancer

The symptoms of breast cancer vary depending on the extent of spread of the disease in the body and from person to person. Although breast cancer occurs painlessly in most women at first, attention should be paid to the symptoms. For this reason, if these changes are noticed during a breast examination, a physician should be consulted without delay. The important point about the symptoms of breast cancer is that there are many reasons for the changes that occur in the breast. Although the majority of these changes are harmless, it has been observed that there is a small possibility that they may be the first signs of breast cancer. For this reason, it is

recommended that women know what is normal for them, examine the natural structure of their breasts, identify the changes and report them without delay, and the importance of participating in breast screening programs commensurate with their age is emphasized. Early stage breast cancer is usually asymptomatic. For this reason, it is of great importance for women to have breast self-examinations (BSE) and screening (=early diagnosis) programs regularly in order to detect breast cancer at an early stage before symptoms develop (Barlow et al., 2002).

Table 2.4. Symptoms of breast cancer (Barlow et al., 2002).

• Mass in the breast: The mass, which is often painless, is found incidentally by women; It is generally hard and immobile. With this feature, the mass that does not move with the surrounding breast tissue can be easily distinguished from the movement of a fibroadenoma.
• Nipple discharge: A small number of women with breast cancer have nipple discharge, which is usually unilateral and spontaneous.
• Lump in the underarm area: The tumor disrupts the lymphatic drainage in the breast, causing a slowdown in lymph flow and subsequent development of lymphedema.
• Nipple turning inward: In tumors located in the upper outer and inner quadrants of the breast, the nipple can be pulled towards the quadrant where the tumor is located. This situation causes an asymmetrical appearance in the breasts and nipples.
• Peau D'orange Symptom (Orange Peel Appearance): Tumor cells in the breast, on the one hand, travel to the regional lymph nodes, and on the other hand, they are carried to the breast skin lymphatics. The increasing amount of tumor cells in the lymph vessels causes the lymphatic flow to slow down and gradually leads to lymphatic blockage, due to narrowing of the lymph vessels.
• Skin retraction: When the tumor growing within the breast infiltrates the Cooper ligaments, it causes these ligaments to shorten. The shortening of the Cooper ligaments pulls the skin towards the tumor, causing the skin to become hollow (bellying, retraction).

Most palpable lumps in the breast may not be cancer, but when women notice a different lump or hardness in the breast, adequate precautions must be taken. However, it is stated that many women delay going to the doctor despite experiencing some symptoms. Breast cancer first spreads to regional lymph nodes and often involves the armpit lymph nodes. Cancer cells that exceed the regional lymph nodes enter the bloodstream and can spread to the lungs, pleura (lung membrane), bones (ribs, lumbar vertebrae, skull, arm bones), liver, peritoneum (abdominal membrane), adrenal glands, brain and ovaries, respectively. In advanced stages of breast cancer, the spread of cancer to other organs other than the breast is called metastasis. This is one of the situations frequently seen in breast cancer patients. From the perspective of developed countries, the development of breast cancer screening programs and the advancement of treatment methods have not changed the fact that metastasis is still a problem (Barlow et al., 2002).

2.6.3. Early Detection Methods in Breast Cancer

Breast cancer is identified based on the structure of the breast. There are units called lobules formed by secreting cells in the breast, and lobules are formed by the union of lobules. The lobules are connected to each other by milk ducts, and these milk ducts unite towards the nipple. Breast cancer develops with the uncontrolled proliferation of cells forming the lobules or milk ducts. Cancer originating from the milk ducts is called ductal carcinoma, and cancer originating from the lobules is called lobular carcinoma. Breast cancer is a malignant development that occurs in the structure of the breast. Tumor diameter and axillary lymph node involvement affect prognosis in breast cancer; It is one of the most important factors that increase mortality and recurrence. Therefore, early diagnosis of breast cancer affects the prognosis positively and reduces mortality. In order to ensure early diagnosis, it is very important to educate and inform women on this issue and their participation in screening programs (Bhushan, Gonsalves, & Menon, 2021).

In the diagnosis of breast cancer; Radiological Methods, Biopsy and Breast Self-Examination are performed.

- **Radiological Methods**

It is the most effective radiological method that is frequently used in the diagnosis of breast cancer and reduces mortality by 20-70% with early diagnosis.

Mammography is used for diagnostic purposes after breast examination in women over 30 years of age with symptoms, for screening purposes in women who have no symptoms or complaints about breast cancer, to investigate the possibility of treating clinically hidden breast cancer at a time when there is a possibility of complete or high recovery. It is a method used for treatment planning and post-treatment follow-up in diagnosed patients.

Ultrasonography (USG); It is a complementary imaging method to mammography in the diagnosis of breast cancer and is the first imaging method used in all symptomatic women under the age of 35. In addition, since it does not contain radiation, it is preferred during pregnancy and lactation periods, since mammographic compression is painful.

Magnetic Resonance Imaging (MRI); It can be performed in patients who cannot be definitively evaluated by ultrasonography and mammography; It can also be used to examine palpable masses after surgery or radiation treatment, and to investigate hidden breast cancer in patients who have a mass in the armpit and whose breast lesion cannot be detected by mammography and ultrasound. Apart from these frequently used radiological applications, radiological methods such as xerography and thermography are also used to detect breast cancer (Bhushan et al., 2021).

- **Biopsy**

It should be performed in cases where there is a palpable mass and continuous discharge from the nipple (except during pregnancy and lactation). As an invasive procedure in diagnosis; Fine Needle Aspiration Biopsy (FNAB), Marking and Excisional Biopsy and Stereotactic Excisional Biopsy can be performed (Bhushan et al., 2021).

Fine Needle Aspiration Biopsy (FNAB) is the process of removing cells from a non-palpable mass with a needle, with the help of mammography and ultrasonography. It is advantageous compared to surgical biopsy because it is a less

invasive and more economical procedure, has less physical and psychological trauma, and does not cause complications; It is a procedure that has disadvantages due to the possibility that the material taken is insufficient and it is not very sufficient for the diagnosis of benign tumors of the breast (Bhushan et al., 2021).

Marking and Excisional Biopsy is a widely used method for the diagnosis of non-palpable tumors detected only by mammography. It is a method that involves removing the lesion, which is usually located with the help of mammography or ultrasonography in the radiology unit, by following the guide wire in the operating room, taking a direct radiograph in radiology and taking it to pathology for a definitive diagnosis (Bhushan et al., 2021).

Stereotactic Excisional Biopsy is the excisional biopsy of an impalpable lesion thanks to newly developed special systems. When the pathological result of the removed lesions is taken as invasive cancer, this system can be used for diagnosis, not for treatment purposes, since it is not known whether the surgical margins are positive or not (Bhushan et al., 2021).

Diagnosing breast cancer at an early stage increases the chance of treatment success and survival. While the 5-year survival probability is 10% in cases with advanced breast cancer, it increases to 80% in cases diagnosed at an early stage. There are basically three complementary methods recommended for early diagnosis;

- Personal (Breast Self Examination) breast checks,
- Clinical breast checks (Physical Breast Examination)
- Mammography

- **Breast Self-Examination (BSE)**

Breast self examination is a recommended method among breast cancer screening methods. The most important reason behind this recommendation is that there are opinions that regular breast examination will help diagnose breast cancer at an early stage. For this reason, it is recommended that every conscious woman learn to perform breast self examination and make it a habit to perform this examination regularly, and from this perspective, BSE is considered an important health habit and behavior. In order for this method, which has a very important place in detecting breast

cancer at an early stage, to be successful, it must be performed continuously, regularly and at periodic intervals, people must be willing to undergo breast examination, know the examination technique well, perform it regularly every month, and they feel responsible for practicing it. It is stated in the literature that 80% of breast masses, both benign and malignant, are discovered by women either by chance, while showering, dressing, or through deliberate breast self examination. However, it is stated that the masses discovered by chance are generally large masses, and as a result of properly performed BSE, it will be possible to discover smaller masses and some findings that may indicate breast cancer. It is mentioned that if the BSE method is understood and applied monthly, the chance of dealing with this type of cancer, which has a high chance of cure if diagnosed early, will be high (Baines, 1992).

Regular breast self examination allows a woman to have an idea of how her breasts should normally be, find out what the normal situation is, and determine any changes as soon as they develop. The best time for a breast exam is indicated as the second or third day after the end of menstruation for menstruating women, and this exam is expected to be performed regularly every month from the age of 20, while women who do not menstruate are advised to choose a specific day of the month and do BSE every month without skipping days. There are three distinct stages of breast self-examination. These;

- Visual evaluation of the breasts,
- Manual evaluation of the breasts in the supine position,
- Standing and manual evaluation of the breasts.

Additionally, when visually evaluating the breasts, it is checked whether there are any significant swelling in the breasts, the nipple pulling inward, changes in shape and color of the nipple, redness, a state of prominence in the superficial veins that did not exist before, or findings resembling an "orange peel" appearance on the skin. After the visual evaluation of the breasts, a manual evaluation examination is performed and the presence of a tissue and an abnormal mass that should not be in the breast tissue is tried to be determined through examination procedures. For this, the entire breast tissue is scanned, from the armpit of the breast to the breastbone, from the collarbone. All areas of the breast up to the lower border are evaluated by carefully feeling (Baines, 1992).

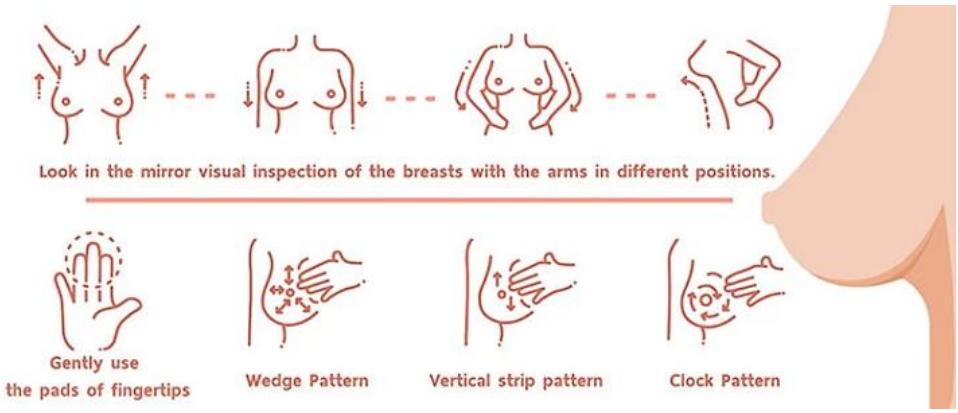


Figure 2.7. Demonstration of breast self examination positions (Massivebio, 2021).

In order to manually evaluate the breasts in the supine position, it is necessary to lie on your back and place a pillow or a folded towel under the right shoulder. After this process, the right hand is placed under the head and the manual scanning methods described above are used and the entire right breast is scanned (Baines, 1992).

Additionally, after the evaluation of the right breast is completed, the same procedures are applied to the left breast. It is recommended that the examination be performed while standing, ideally under a shower and with soapy hands, so that possible masses in the breast tissue can be more easily identified (Baines, 1992).

- **Clinical Breast Examination**

Breast self-examination is not the only method used in the early detection of breast cancer. There is information that both breast examination by a physician or nurse and mammography may be effective in reducing the death rate from breast cancer. It seems that wide and different recommendations have been developed regarding the age of clinical breast examination and screening intervals. From this perspective, according to the generally accepted view; It is recommended that women between the ages of 20 and 39 have this examination every 3 years, and women who are 40 and over have this examination annually (Baines, 1992).

- **Mammography**

Mammography is defined as radiological x-ray imaging of the breast and reveals the structure of the breast. Mammography is stated to be a good and accurate way to diagnose breast cancer at an early stage, and this method is accepted as a routine

screening examination. It is underlined that mammography is an effective method in reducing deaths due to breast cancer, and it is stated that this rate varies between 30% and 40% Also, it is possible for women who have a family history of breast cancer and are at risk to be controlled and monitored by the physician before 40 age with mammography. Mammography, which is defined as taking films of the breast with special devices made for the breast, is divided into two: for screening purposes and for diagnostic purposes. In screening mammography, films of both breasts are taken from the top and side, and if necessary, additional films are requested in different positions. This method is briefly expressed as “cancer screening or screening mammography”. The most important benefit of screening mammography is that it can detect and treat breast cancer at an early stage, thus reducing deaths from this disease. Mammography can be performed on people who have symptoms such as a lump in the breast, pain, and nipple discharge, and who are diagnosed with early stage breast cancer, and who have abnormal findings during screening mammography, upon the recommendation of the physician. This method is also referred to as "diagnostic mammography". If there is a suspicious finding as a result of this diagnostic mammography, breast ultrasonography is performed and it is investigated whether the detected mass is a cyst (fluid-filled mass). When it is understood that the mass in the breast is not a cyst, a piece of this mass is taken with a needle and examined by biopsy method. The definitive conclusion as to whether a person has breast cancer is reached through the biopsy method. The most important advantage of mammography, which is described as a special x-ray technique, is that it can detect changes in the breast that can be expressed as cancer long before physical symptoms appear. In this case, in breast cancer diagnosed at an early stage, the size of the tumor will be smaller and the risk of metastasis will decrease. However, the important point is that mammography cannot always be described as perfect because it misses some cancers. For this reason, physicians occasionally recommend performing some tests in addition to regular mammography to determine whether the changes found on mammography are cancer or not. In order to get the maximum benefit from mammograms, it is recommended to perform a clinical breast examination shortly before the mammogram. This is also important in terms of increasing the reliability of the scans to be performed. Another issue that should be noted at this point is that although there is insufficient evidence that mammography directly causes cancer; Exposure to intense radiation and having

mammograms done more frequently than necessary could hypothetically be dangerous for women (Baines, 1992).

2.6.4. Staging of Breast Cancer

Breast cancer is a complex and genetic disease that does not have a single cause. It is examined in 3 groups: hereditary, familial and sporadic. If one or more people in the family have breast cancer, it is called familial breast cancer. Breast cancer is a heterogeneous disease with a spectrum of many subtypes with distinct biological properties that lead to differences in response patterns to various treatment approaches and clinical outcomes. The heterogeneity of breast cancer affects diagnosis, treatment and therefore prognosis. Mutation in the BRCA1 and BRCA2 genes, which have been found to be directly related to breast cancer, the occurrence of cancer in the family at an early age, and the occurrence of breast cancer in men in the family indicate that the disease is hereditary. People with hereditary breast cancer have a very high risk of developing breast cancer during their lifetime, at 85%. Those outside these two groups are sporadic (incidental) breast cancers that do not have a familial or hereditary connection. Although many genes (such as BRCA1 and BRCA2) that cause breast cancer and play a role in its development have been identified, the molecular mechanisms involved in its regulation are still not fully explained. The development of high-throughput microarray-based expression profiling technologies has facilitated large-scale studies in breast cancer, and a lot of data has been obtained to elucidate the causes and formation of breast cancer, and new biomarkers have been developed that are specific and suitable for treatment. Breast tumors are heterogeneous tumors with different characteristics according to morphological, clinical, hormone receptor level, response to treatment. The reason for this difference is the difference in the number of underlying cancer cells, different combinations of oncogenes activation and/or tumor suppressor gene function losses. For this reason, in addition to histological and pathological classification, breast cancer has also been classified as molecular due to expression microarray studies (Weigelt, Geyer, & Reis-Filho, 2010).

2.6.4.1. Histological Classification of Breast Cancer

Traditional classification of breast cancer is done by microscopic method. This classification, called histological or morphological classification, is based on the size, shape and arrangement of breast cancer cells. In addition, the region where breast

cancer develops is an important criterion. It is important to answer two questions in this classification. First, whether the tumor is limited to the breast epithelial component where it developed or whether it is spreading to surrounding tissues; Secondly, whether the source of the tumor is the milk duct or the mammary gland. If the cancer developed in the milk ducts, ductal; If it is developed in the mammary glands, it is called lobular. It is observed that, although rare, it can develop in the connective tissue outside the mammary glands and ducts (Weigelt et al., 2010).

Additionally, the spread of breast cancer also affects its type. In this context, breast cancer is divided into two: *in situ* and *invasive. In situ* breast cancer is a type that is limited to the area where it started and does not spread to the rest of the breast tissue. *Invasive* breast cancer refers to the type that spreads into the surrounding breast tissue. Among the breast cancer types, the most dominant type with a rate of 70% is invasive ductal carcinoma. *Invasive* lobular carcinoma, one of the special histological subtypes, ranks second with approximately 10%. Subtypes such as tubular, mucinous, cribriform and papillary carcinoma are generally associated with favorable prognoses. The incidence of medullary and mucinous carcinoma is approximately 5%, while the incidence of tubular carcinoma varies between 1-5%. Other microscopic types are rarely seen (Weigelt et al., 2010).

- **Ductal Carcinoma *In Situ* (DCIS)**

DCIS is a breast disease that develops when milk duct cells become abnormal and show typical features of transformation into cancer cells. It does not cause membrane invasion but has the potential to be a leading lesion in the development of invasive carcinoma. It can also be defined as a malignant proliferation of ductal epithelial cells. DCIS is considered a precursor to invasive breast carcinoma, but this does not mean that it will occur before every carcinoma. Carcinomas independent of DCIS can also occur. Due to this pre-invasive lesion feature, its early diagnosis becomes extremely important due to the increased awareness of breast cancer and the widespread use of breast cancer screening, DCIS has now become defined as a disease that can be diagnosed early. Death due to DCIS is a rare condition, and the presence of an undetected invasive component can lead to death from the disease (Badve & Gökmen-Polar, 2019).

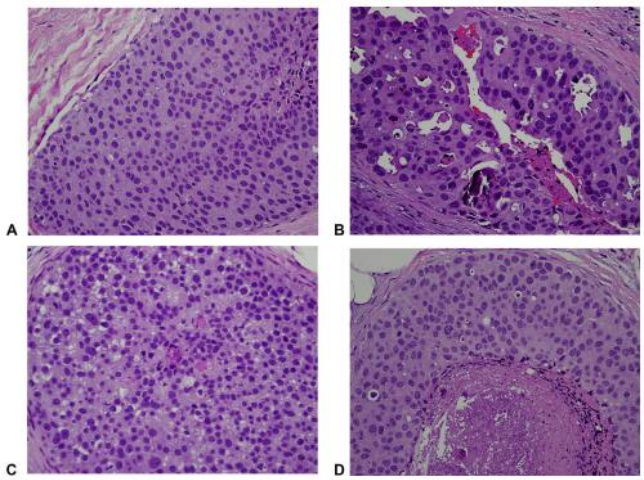


Figure 2.8. Needle core biopsy image of a patient with ductal carcinoma *in situ* (DCIS) (Badve & Gökmen-Polar, 2019).

DCIS can be diagnosed as high-grade, intermediate-grade and low-grade. Its high grade indicates that it is more prone to turn into an invasive type, with features such as rapid proliferation and recurrence after treatment. In cases diagnosed as high and medium grade, surgery is used as the main treatment and the tumor and some normal tissue in the immediate vicinity of the tumor are removed. It is considered that in low-grade cases, surgery has no effect on patient survival and therefore does not require treatment (Badve & Gökmen-Polar, 2019).

- **Invasive Ductal Carcinoma (IDC)**

An invasive type of cancer occurs when the cancer spreads beyond the upper layer of the cell, where it started. Most breast cancers are invasive carcinomas and spread. Among cancers that can spread, ductal carcinoma, which arises from the cells that form the breast ducts, is the most common type of breast cancer. Cancer first begins inside the canal, then crosses the canal wall and spreads to breast tissue, sometimes outside the breast. It accounts for about 80% of invasive breast cancers (Badve & Gökmen-Polar, 2019).

- **Lobular Breast Cancer**

It is a type of cancer that develops from the mammary glands and is less common than ductal breast cancers (Provenzano, Ulaner, & Chin, 2018).

- **Lobular Carcinoma *In Situ* (LCIS)**

It is the presence of abnormal cells in the lobules. LCIS is defined as the cells in the lobules where milk production begins to become abnormal, and it is not considered cancer. Cells that resemble cancer cells grow inside the lobules but do not invade to spread outside the lobules. LCIS is diagnosed through a biopsy, in which small pieces of breast tissue are removed and checked in the laboratory. Most of the time, LCIS does not cause a lump that can be felt or changes that can be seen on a mammogram. Usually, LCIS is found when a biopsy is performed for another nearby breast problem. DCIS can often be diagnosed randomly as a result of a biopsy performed with a different purpose or surgical interventions targeting another lesion. Variants such as pleomorphic LCIS and LCIS with central necrosis can be detected by mammography. Since it is not defined as a cancer, it does not require treatment if diagnosed. However, it is recommended to be monitored with routine check-ups as it may increase the risk of developing cancer. Unlike DCIS, which is defined as a precursor lesion and is seen only on one side, LCIS can be seen in both breasts and is defined only as a risk factor. However, since this risk factor is the source of both ductal and lobular breast carcinomas, it is recommended that the diagnosed patient be followed up throughout life (Provenzano et al., 2018).

- **Invasive Lobular Carcinoma (ILC)**

It is the form in which cancer cells pass outside the lobules. Cancer has spread inside or outside the breast. It accounts for 10-15% of all breast cancers. Invasive lobular carcinoma begins in the epithelial cells of the mammary glands and has the ability to spread to the breast tissue and other parts of the body. This spread usually occurs through internal organs. It is generally more common in older women and can be more difficult to diagnose with mammography or physical examination than IDC. Research on lobular cancers is increasing day by day, and it is expected that the progress of studies in this field will contribute to the development of personalized treatment methods. The treatment to be applied in ILC depends on the aggressiveness of the tumor, its stage and the general health condition of the patient. Treatment is planned locally or systemically. Local treatments target the breast tissue and its immediate surroundings. It involves removing part or all of the breast tissue followed by radiation therapy. Systemic treatment methods are preferred in cases of cancer that

has spread beyond the breast tissue. Chemotherapy and hormone therapy are the most basic system of treatment (Provenzano et al., 2018).

- **Inflammatory Breast Cancer (IBC)**

Cancer cells block the lymphatic flow of the breast skin and the breast becomes red, swollen and its skin thickened. It is a rare but aggressive breast cancer. There is an increase in temperature in the breast. Inflammatory breast cancer, one of the most aggressive types of cancer, is a type of cancer that blocks the lymph vessels in the breast skin and therefore causes inflammation in the breast and effects such as swelling, warmth and skin erythema in the breast tissue due to this inflammation. In addition, early diagnosis of the disease is not possible and the disease can be diagnosed at an advanced stage. Although the method used in treatment may vary from patient to patient, systemic treatment is commonly applied primarily to shrink the tumor. Then, the tumor is removed by surgical operation and radiation therapy is followed to prevent regional recurrence of the disease following the surgery (Barazi, 2024).

- **Paget Disease of the Breast**

It is a rare type of cancer that affects the skin on the nipple or areola. It may be associated with other invasive cancer in the breast. Sometimes it may be mistaken for eczema and there may be a delay in diagnosis and treatment.

2.6.4.2. Pathological Classification of Breast Cancer

Pathological classification of breast cancer is basically made according to the amount of certain receptors. Pathological classifications based on the presence of the receptor are of great importance in deciding on the treatment process of breast cancer patients and in general, in terms of the correct use of drugs developed to target receptors (such as trastuzumab and tamoxifen). If there is an excess of estrogen, progesterone and HER2 receptors, it is called positive, otherwise it is called negative. If it does not have all three receptor types, it is known as triple negative. Breast cancer, pathologically; There are 7 types: ER+, ER-, PR+, PR-, HER2+, HER2 and triple negative. Trastuzumab (Herceptin) is a drug designed to block the function of HER2 receptors by targeting the extracellular part and is used in HER2+ patients. It has been shown that the lifespan of patients receiving this treatment increases and 30% of the patients respond to the drug. In addition, approximately 70% of breast cancers are

estrogen positive (ER+) and develop in response to the estrogen hormone (Rakha, Tse, & Quinn, 2023).

2.6.4.3. Molecular Classification of Breast Cancer

The fact that breast cancer is a complex disease and breast tumors are heterogeneous tumors with different characteristics makes it difficult to explain the mechanisms in the regulation of the disease and, as a result, the treatment. Traditional classification systems for biological features such as tumor size, lymph node involvement, histological grade, patient age, estrogen receptors (ER), progesterone receptors (PR), and human epidermal growth factor receptor 2 (HER2 or c-erbB2) status patient-specific treatment strategies It remains limited for . In addition, histological imaging of tumors is not sufficient to elucidate the complex genetic changes involved in the mechanism and biological events involved in cancer development and progression. It has become necessary to search for new biomarkers, and in addition to histological and pathological classification, breast cancer is also classified molecularly with this deficiency. Tumors with similar clinical and pathological presentations may behave differently. Therefore, studies in recent years have focused on defining more detailed biological features to improve patient risk stratification and ensure the highest chance of benefit and least toxicity from a given treatment modality. In the modern medical world, morphological classification and clinical pathological parameters in breast cancer are insufficient to predict the pathophysiology of breast tumor. For this reason, studies focus on the analysis of molecular models of breast cancer. Comprehensive and groundbreaking gene expression studies conducted in 2000 with the analysis study, it was first divided into 4 groups as luminal cell-like, basal cell-like, normal epithelial-like and HER2+ groups. Indicator gene groups such as estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor (HER2) played a role in determining these subtypes. As a result of subsequent gene expression profile studies, 4 molecular subtypes were determined. These; Luminal A, Luminal B are HER2 Positive and Triple Negative. In general, luminal A and B types have a benign course, while HER2+ and triple negative groups have a more aggressive course. However, after the use of the anti-HER2 antibody (smart drug) called Herceptin, the prognosis began to improve dramatically in the HER2+ group (Rakha et al., 2023).

Luminal type A breast cancer, which is estrogen and/or progesterone receptor positive and HER2 negative, has a good prognosis and the tumor recurrence rate is lower than other subtypes. In addition, it is the most common subgroup, accounting for 50-60% of breast cancer cases (Yersal & Barutca, 2014).

Luminal type B (triple positive) breast cancer is a type of cancer that is estrogen and/or progesterone receptor positive and HER2 positive. Unlike Luminal A, the expression of genes related to proliferation is higher. The presence of luminal B tumors also indicates that growth receptor signaling genes are increased. This subtype, which accounts for 15-20% of breast cancer cases, is Luminal B and has a higher histological grade, a more aggressive progression and, accordingly, a lower prognosis and a higher recurrence rate than Luminal A (Yersal & Barutca, 2014).

HER2 Positive breast cancer, which is estrogen and/or progesterone receptor positive exhibits a more aggressive biological and clinical behavior than other types (Yersal & Barutca, 2014).

Triple negative breast cancer, which is estrogen and/or progesterone receptor negative and Her2 negative, is not sensitive to the three hormones, unlike other breast cancers. None of the estrogen hormone receptor, progesterone hormone receptor and HER2 human epidermal growth factor receptors are present (Yersal & Barutca, 2014).

2.6.4.4. TNM Staging Method in Breast Cancer

Correct staging of breast cancer before and after surgery is important. After breast cancer is diagnosed, it is investigated whether cancer cells have spread within the breast or to other parts of the body. This process is called "staging". Knowing the stage of the cancer is important to choose the best treatment. While the patient's surgery plan is made during preoperative staging, postoperative pathological staging is absolutely necessary for the decision on systemic and local treatment for the patient. In addition, staging also provides information about the patient's prognosis. Staging can be divided into clinical and pathological. In clinical staging, the status of the tumor and peripheral lymph nodes are taken into consideration using palpation, inspection and radiological findings. In pathological staging, the size of the tumor and the presence of metastasis in the lymph nodes removed by axillary dissection are evaluated

macroscopically and microscopically. The most commonly used system in staging today is the TNM classification adopted by UICC (Union International Center Cancer) and AJCC (American Joint Committee on Cancer) (Koh & Kim, 2019).

Tumor (T): It indicates the tumor (It gives information about the size of the tumor and how much it has spread to the breast and surrounding organs.)

Node (N): It provides information about the condition of the lymph nodes.

Metastasis (M): It provides information about whether the cancer has metastasized.

Additionally, some numbers and letters after T, N and M are used for details about the tumor, lymph node and metastasis.

Primary tumor (T):

TX: Primary tumor cannot be evaluated

T0: No primary tumor

Tis: Cancer in situ (ducal carcinoma in situ (DCIS), lobular carcinoma in situ (LCIS), or Paget's Disease of the nipple without tumor)

T1: Tumor ≤ 20 mm in greatest dimension

T1mi: Tumor ≤ 1 mm in greatest dimension

T1a: Tumor > 1 mm but ≤ 5 mm in greatest dimension

T1b: Tumor > 5 mm but ≤ 10 mm in greatest dimension

T1c: Tumor > 10 mm but ≤ 20 mm in greatest dimension

T2: Tumor > 20 mm but ≤ 50 mm in greatest dimension

T3: Tumor > 50 mm in greatest dimension

T4: Tumor can be any size; tumor that spreads directly to the chest wall or skin, breast edema, ulceration, inflammatory breast cancer

T4a: Tumor invades the thoracic wall

T4b: Breast skin edema, ulceration, skin nodules confined to the same breast

T4c: T4a + T4b

T4d: Inflammatory carcinoma



Figure 2.9. MRI image of rib extension and skin ulceration in the left breast (Koh & Kim, 2019).

Clinical regional lymph nodes (N):

cNX: Primary tumor cannot be evaluated (e.g. previously removed)

cN0: No lymph node metastasis

cN1: Mobile lymph node metastasis in ipsilateral axilla

cN2: Presence of metastasis detected by imaging methods in the mammary interna in the ipsilateral axilla, adherent to each other or fixed to the surrounding tissues.

cN2a: Fixed lymph node metastasis in the ipsilateral axilla

cN2b: Only clinically detected internal mammarian lymph node metastasis on the same side and no clinically detected axillary lymph node metastasis

cN3: Ipsilateral infraclavicular lymph node metastasis, or clinically detected internal mammary lymph node metastasis and clinically detected axillary lymph node metastasis, or supraclavicular lymph node metastasis on the same side

cN3a: Ipsilateral infraclavicular lymph node metastasis

cN3b: Ipsilateral internal mammarian lymph node metastasis and axillary lymph node metastasis

cN3c: Ipsilateral supraclavicular lymph node metastasis

Pathological regional lymph nodes (N):

pNX: Primary tumor cannot be evaluated (e.g. previously removed)

pN0: No metastasis to regional lymph node

pN0 (i+): Malignant cell clusters no larger than 0.2 mm and there are isolated tumor cells

pN0 (mol+): Positive molecular findings by reverse transcriptase-polymerase chain reaction (RT-PCR)

pN1: Mobile metastasis to ipsilateral axillary lymph node(s)

pN1 (mi): Micrometastasis (0.2-2 mm)

pN1a: Metastasis to 1-3 axillary lymph nodes

pN1b: Microscopic metastasis in internal mammarian lymph node

pN1c: 1-3 axillary lymph node metastasis + microscopic metastasis in internal mammarian lymph node (pN1a + pN1b)

pN2: Axillary lymph node metastases on the same side, fixed to each other or to surrounding structures and metastasis in 4-9 axillary lymph nodes

pN2a: Metastasis in 4-9 axillary lymph nodes

pN2b: Clinical metastasis in internal mammarian lymph node

pN3: Metastasis to ipsilateral mammarian internal lymph nodes

Distant metastasis (M):

MX: Distant metastasis cannot be evaluated

M0: No distant metastasis

M1: There is distant metastasis

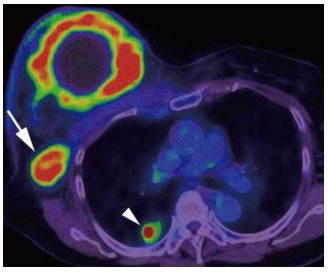


Figure 2.10. MRI image with lung metastases in the right breast (Koh & Kim, 2019).

2.6.4.5. Stage Groups for Breast Cancer

In breast cancer, determining the stage of the disease plays an important role in planning the treatment to be applied, and staging uses imaging methods and laboratory to find out whether the cancer has spread and, if so, to which parts of the body it has spread (Koh & Kim, 2019).

Stage 0 (Carcinoma in situ)

In situ cancers are cancers that have not spread and have not spread to surrounding tissues. There are two types of in situ cancer.

- Ductal carcinoma *in situ* (DCIS): A pre-cancerous condition in the milk ducts. The abnormal cells have not spread out of the duct and into the surrounding breast tissue. However, sometimes if DCIS is not treated, cancer that can spread occurs.
- Lobular carcinoma *in situ* (LCIS): It is not cancer. It is the presence of abnormal cells in the lobules. These abnormal cells are a sign of high risk. This means that a woman with LCIS has a higher risk of developing cancer that may spread to both breasts.

Stage I

- The tumor size is 20 mm or less and the cancer cells have not spread outside the breast (to the lymph nodes).

Stage IIA

- The tumor is 20 mm or smaller and has spread to the lymph nodes in the armpit,
- The tumor is larger than 20 mm but less than 50 mm and has not spread to the armpit lymph nodes.

Stage IIB

- The tumor is larger than 20 mm and smaller than 50 mm and has spread to the armpit lymph nodes,
- The tumor is larger than 50 mm but has not spread to the armpit lymph nodes.

Stage IIIA

- The tumor is 50 mm or smaller and has spread to the armpit lymph nodes, and the lymph nodes are adherent to surrounding tissues or to each other,
- The tumor is larger than 50 mm and has spread to the armpit lymph nodes, and the lymph nodes are adherent to the surrounding tissues or to each other.

Stage IIIB

- The tumor can be of any size and has spread to tissues adjacent to the breast (skin or chest wall, ribs or muscles in the chest wall) and either has an orange peel appearance on the breast,
- Tumor cells have spread to the lymph nodes in the rib cage (under the ribs),
- Inflammatory breast cancer is considered stage IIIB.

Stage IIIC

- The tumor can be of any size, may have spread to tissues adjacent to the breast (skin or chest wall, ribs or muscles in the chest wall),
- The tumor cells have spread to both the armpit lymph nodes and the lymph nodes within the rib cage (under the ribs),
- The tumor has spread to the lymph nodes below the collarbone also, there may or may not be spread to the armpit lymph nodes.
- Tumor cells have spread to the lymph nodes above the collarbone; There may or may not be spread to the lymph nodes in the armpit and rib cage.
- **Stage IV** (**Metastatic)**
- Cancer cells have spread outside the breast to other parts of the body (such as bones, lung, liver or brain).

According to doctors, cancer between Stage I and Stage IIA is considered "early stage", while cancer between Stage IIB and Stage III is considered "locally advanced stage" (Koh & Kim, 2019).

Table 2.5. TNM staging in breast cancer

Stage 0	Tis	N0	M0
Stage I	T1	N0	M0
Stage IIA	T0	N1	M0
	T1	N1	M0
	T2	N0	M0
Stage IIB	T2	N1	M0

	T3	N0	M0
Stage IIIA	T0	N2	M0
	T1	N2	M0
	T2	N2	M0
	T3	N1	M0
	T3	N2	M0
Stage IIIB	T4	No	M0
	T4	N1	M0
	T4	N2	M0
Stage IIIC	T	N3	M0
Stage IV	T	N	M1

2.6.5. Breast Cancer Treatment Methods

The chance of success in the treatment of breast cancer is quite high. However, treatment methods vary depending on the stage of cancer. The earlier breast cancer is detected, the easier and more effective the treatment becomes. After breast cancer is diagnosed by biopsy, most patients require surgery to remove the cancer. In more advanced stages, surgery may be possible by preserving the nipple and breast skin and applying an implant. Depending on the stage of the cancer and the type of operation, radiotherapy and chemotherapy may be added to the surgical treatment. In the early stages, breast-conserving surgery, that is, removal of only cancerous tissue, may be sufficient. The armpit glands on the side where the cancer is located are also removed with this surgery. The surgically removed tumor and lymph nodes are examined under a microscope and a report is written. It is necessary to determine estrogen and progesterone receptors in the removed cancerous tissue because this test shows whether the patient will benefit from hormone therapy. Many important features of the tumor written in the pathology report, such as the size of the tumor, the appearance of cancer cells, whether the lymph nodes are involved by cancer cells, and the presence of estrogen and progesterone receptors, provide guidance in determining the treatment plan. Thus, the stage of the disease is determined and a committee formed by the general surgeon, medical oncologists, and radiation oncologists decides whether additional treatment is needed after the surgery and, if so, what treatment should be given, taking into account the patient's characteristics in the pathology report, age,

whether she has entered menopause, and her general condition (Maughan, Lutterbie, & Ham, 2010).

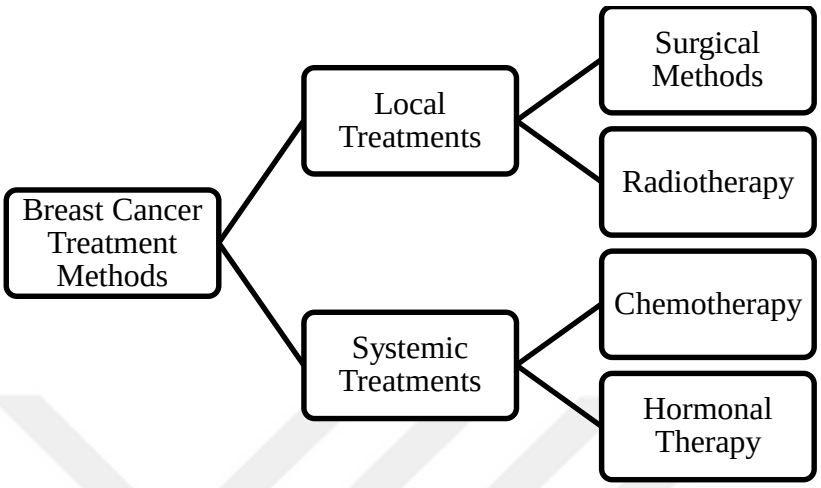


Figure 2.11. Schematic image of breast cancer treatment methods (Maughan et al., 2010).

2.6.5.1. Surgical Methods

Breast cancer is a treatable disease and when detected early, it is possible to get rid of this disease completely. After a diagnosis of breast cancer is made, the first stage of treatment is usually surgery. The most effective treatment of breast cancer known today is the surgical removal of the tumor and the lymph nodes where the tumor cells have spread. The method to be followed in the treatment of early stage breast cancer is to first surgically remove the tumor and clean the armpit lymph nodes (Maughan et al., 2010).

Mastectomy

Mastectomy is one of the best treatment methods in case of breast recurrence after breast-conserving surgery. Mastectomy is the name given to surgery to remove cancer tissue in the breast. Mastectomy is performed especially in multifocal disease, in the presence of a large tumor proportional to the size of the breast, when there is tumor in more than one quadrant of the breast, when there is a history of collagen vascular disease and if the person chooses this type of surgery (Maughan et al., 2010).

- **Simple Mastectomy**: It is the process of preserving the pectoral muscles and removing all breast tissue, including the nipple. The fatty tissue around the breast and the skin over it are removed, and the armpit lymph nodes are usually removed at the same time. Simple mastectomy is generally used as primary treatment for early and operable breast cancers (Maughan et al., 2010).
- **Radical Mastectomy**: It is the removal of the pectoral muscles, axillary lymph nodes, nipple and the entire breast. Breast chest muscles and armpit lymph nodes together is taken. Significant improvements have been made in terms of survival in breast cancer with radical mastectomy (Maughan et al., 2010).
- **Removing the Armpit Lymph Nodes:** Cleaning of the axillary lymph nodes is performed to control cancer in the axillary area and to understand the extent of the disease in patients with tumors that have not spread distantly. All lymph nodes and lymph channels under the armpit are removed (Maughan et al., 2010).
- **Senitel Lymph Node Biopsy:** Sentinel lymph node biopsy is the removal of a sentinel lymph node during surgery. Sentinel lymph node is the first lymph node to which the lymph flow of a tumor goes. It is the lymph node where cancer is most likely to spread. Sentinel lymph node is the lymph node where cancer cells that can metastasize from the primary tumor first stop and begin to spread. The status of this lymph node provides vital information about the spread and stage of the cancer, so it is of great importance in determining the correct treatment strategy (Cavalcante et al., 2023).

Breast Conserving Surgery (BCS)

It is suitable for patients with stage I and II breast cancer. For breast-conserving treatment to be administered, the patient must have cancer that can be removed without any clinical or mammographic evidence of residual cancer. The breast is preserved by removing the tumor along with some disease-free breast tissue around it. For such surgery to be performed, the breast must be large and the tumor must be small. Thus, the possibility of breast deformity decreases after the tumor is removed. In order to complete the treatment, radiotherapy is given to the breast where conservative surgery was performed. A high rate of tumor recurrence was observed in patients who were not given radiotherapy (Maughan et al., 2010).

- **Lumpectomy**: Lumpectomy is the removal of only the tumor and surrounding breast tissue. In this procedure, the cancerous tumor and some healthy tissue around it are removed. After surgery, the remaining intact tissue in the area can be treated with radiotherapy. Lumpectomy is one of many methods used to treat breast cancer and is often performed on small breasts with tumors (Maughan et al., 2010).

- **Reconstructive (Plastic) Surgery**

It is a surgery performed using plastic surgery methods during breast treatment. The breast can be reconstructed after mastectomy. This procedure can be performed using an artificial breast prosthesis or the patient's own tissue. Thus, after the tumor in the breast is removed, the breast is reconstructed aesthetically (Maughan et al., 2010).

2.6.5.2. Radiotherapy

Radiotherapy is a type of cancer treatment that uses high-energy x-rays or other types of radiation to kill cancer cells or stop them from multiplying. The main aim of radiotherapy treatment is to destroy the tumor tissue while causing minimal damage to healthy tissues. Radiotherapy applied after BCS is an integral part of the treatment. In this method, high-energy ionizing rays or atomic particles are used and survival is increased primarily by providing local and regional control. In addition, there are two types of radiotherapy. In external radiotherapy, beams are sent to the cancerous area from a machine outside the body. In internal radiotherapy, a needle, wire, core or catheter containing radioactive material is placed directly into or next to the cancerous tissue. The method of administering radiotherapy varies depending on the type and stage of cancer. In addition, some patients receive radiotherapy to destroy cancer cells before surgery to shrink the tumor. This approach may be preferred in cases where the tumor is large or difficult to remove. While chemotherapy or hormone therapy may be applied to some women before surgery, by using radiation techniques and overlapping more than two beams on the tumor volume, the tumor tissue absorbs a greater amount of radiation than the surrounding tissues, thus causing more damage to the tumor. The amount of radiation dose to be applied in radiotherapy is limited to what the patient can tolerate and the expected quality of life after treatment is at an acceptable level (Polgár et al., 2022).

Additionally, in the treatment planning for radiotherapy, the area to be treated is determined in advance and a series of procedures are performed to determine the targeted treatment area before starting radiation therapy, to initiate the treatment in daily treatment practices and to protect healthy tissues from radiation. A team consisting of a medical radiophysicist, radiation oncologist and radiotherapy technician organizes the preparation process called simulation. During the simulation, the patient's skin is marked with a special ink pen and area measurements are taken to determine the treatment area and to define the points determined to give the appropriate position. After the most suitable area is determined, the drawn area is given radiation at the same dose for approximately 1-2 hours in each treatment (Polgár et al., 2022).

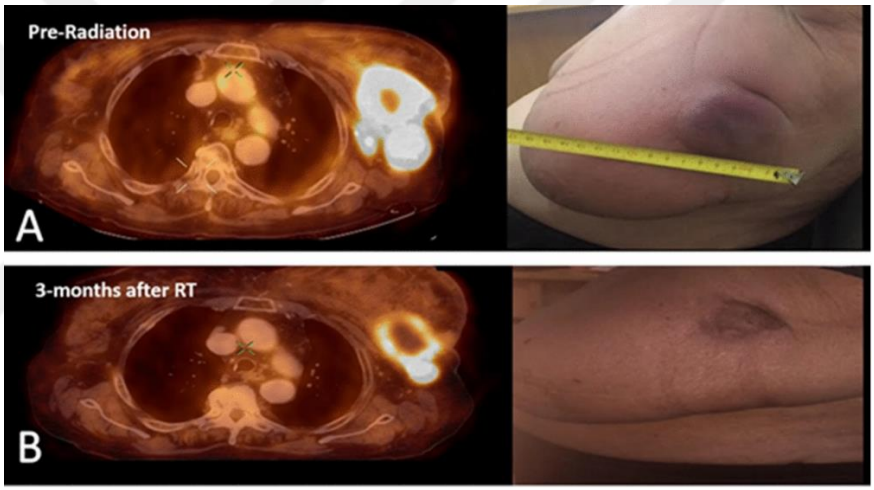


Figure 2.12. PET-CT image of the patient with locally advanced breast cancer before radiation therapy and 3 months after radiation therapy (Jacobson et al., 2021).

2.6.5.3. Chemotherapy

Chemotherapy, also known as drug therapy, is a treatment method applied in addition to surgical treatment. Even if there is no cancer left in any area after surgical treatment, medication can be applied for a while as a precaution. The growth of tumors is controlled by using anticancer drugs. Metastasis is prevented and complaints caused by cancer are reduced through treatment with natural or synthetic chemicals, biological agents and hormones, which have selective lethal effects especially against proliferating cells. In order to understand how chemotherapy works, it is very important to first know the normal life cycle. Chemotherapy drugs kill cancer cells and

is a treatment method planned in 4-6 cures with 3 weeks between cures. The most important difference between a cancer cell and a normal cell is that cancer cells do not have a mechanism that slows the progression of proliferation and are in a non-stop proliferation that leads the organism to death (Zhong et al., 2021).

Chemotherapeutic agents are most effective when the cells are in the proliferative phase. This is of great importance during the development of drugs. Because drugs that specifically affect one phase of the life cycle or drugs that are effective on all phases can be developed. The purpose of chemotherapy is to kill cancer cells and prevent metastasis. Chemotherapy drugs enter the bloodstream and distribute throughout the body, thus providing treatment for cancers that have spread to distant organs. Chemotherapy is a systematic form of treatment and affects all tissues and organs in the body. Therefore, chemotherapy is distinguished from treatments aimed at local treatment, such as surgery and radiotherapy. Chemotherapy also helps the tumor shrink before surgery, making it easier to remove. In this way, doctors understand how the tumor responds to medications. In cases where the tumor does not shrink, different medications are used. This treatment method is performed to complement other treatment methods, and this type of chemotherapy is called adjuvant chemotherapy (Zhong et al., 2021).

Additionally, chemotherapy is usually given through temporary needles into small veins in the hand or arm, or through intrabody devices called ports that are inserted into larger veins. Some chemotherapy drugs can be given orally in pill or syrup form. However, there are also chemotherapy drugs given by direct injection under the muscle or skin or into the tumor area (Zhong et al., 2021).

Chemotherapy may be given before surgery (neoadjuvant) or after surgery (adjuvant). Chemotherapy given before surgery is called "neoadjuvant therapy". There is no difference in survival between chemotherapy given before or after surgery. The main benefit of neoadjuvant chemotherapy is that it can shrink large tumors. Thus, tumor tissues can be removed with a less extensive surgery. Chemotherapy given after surgery is called "adjuvant treatment". While surgery is performed to remove visible cancer tissue, adjuvant chemotherapy is performed to eliminate remaining invisible cancer cells. Adjuvant treatment applied after breast-conserving surgery and

mastectomy reduces the risk of breast cancer recurrence. In the early stages of the disease, cells that can separate from the primary breast tumor and spread through the bloodstream, do not cause symptoms and cannot be seen in imaging systems, can also form new tumors in other parts of the body under appropriate conditions. The goal of adjuvant chemotherapy is to eliminate cancer cells that may cause metastasis. Often, using chemotherapy drug combinations is the most effective chemotherapy treatment method (Zhong et al., 2021). Chemotherapy is generally applied in cancer treatment as follows;

- As an induction agent in the treatment of advanced disease (induction chemotherapy),
- As an adjunct to local treatment methods (surgery, radiotherapy) (adjuvant chemotherapy),
- In the primary treatment of localized cancer (neoadjuvant chemotherapy),
- It is applied directly to the area affected by cancer or to specific areas.

Additionally, breast cancer patients receiving chemotherapy may receive one drug or a mixture of multiple drugs. Most of the time, combination therapy, in which more than one drug is mixed together, gives more effective results than treatment with a single drug. Combination therapy has yielded better results in controlling cancer cells despite taking less of each of the drugs in the mixture. Compared to single-drug chemotherapy treatments given at higher doses, combination chemotherapy not only provides better results but also causes fewer side effects on the patient. Chemotherapy drugs and their administration methods are usually arranged separately for each patient. When planning chemotherapy, parameters such as the patient's age, general health status, stage and grade of the cancer, other health problems, past and future treatments should be taken into consideration (Zhong et al., 2021). Cancer chemotherapy drugs are classified according to their chemical structure and cell activity:

- Alkylating agents: Their cytotoxic effects occur through an irreversible combination of the electrophilic alkyl radical in their structure and the nucleophilic part of the target macromolecules (Michael Colvin, 2003).
- Antimetabolites: Since they are similar to the normal metabolites of the cell, they replace the metabolites they are similar to for enzymes, or they block or

reduce the activity by taking the same role, or they enter into macromolecules and create a non-functional macromolecule (Kasi, 2024).

- Plant alkaloids: These are drugs obtained semisynthetically from podophyllotoxins and vinca alkaloids. They stop cell division during mitosis (González-Burgos & Gómez-Serranillos, 2021).
- Antitumor antibiotics: They stop DNA and RNA transcription in the cell and cause cell death during DNA synthesis as they remain in the tissues for a long time. They increase toxicity when given together with radiation (Kasi, 2024).
- Other drugs: L Asparaginase is in this group. It is an enzyme obtained from Escherichia coli or Erwinia cultures. It inhibits DNA and RNA synthesis in blastic cells. Corticosteroids in this group enter the cell by passive diffusion, bind with glucocorticoid receptors and pass to the nucleus, where they bind with DNA and disrupt the transcription process (Kasi, 2024).
-

Table 2.6. Cancer drugs approved by the Food and Drug Administration (FDA) for breast cancer (Arora et al., 2022).

Abemaciclib	Anastrozole	Aromasin	Capecitabine	Capivasertib
Cyclophosphamide	Docetaxel	Doxorubicin	Elacestrant	Ellence
Ibrance (Palbociclib)	Faslodex (Fulvestrant)	Ibrance (Palbociclib)	Lynparza (Olaparib)	Herceptin (Trastuzumab)
Ribociclib	Letrozole	Olaparib	Paclitaxel	Palbociclib
Pertuzumab	Ribociclib	Tamoxifen	Toremifene	Trastuzumab
Truqap	Tucatinib	Tukysa	Verzenio	Xeloda
Zoladex	Alpelisib	Everolimus	Gemzar	Halaven
Letrozole	Exemestane	Fulsvestrant	Ixabepilone	Pertuzumab

The US Food and Drug Administration (FDA) is an institution that is considered a pioneer worldwide in approving new cancer drugs, and it has also approved drugs for breast cancer. In addition, drugs approved by the FDA are shown in Table 7. Drugs approved to prevent breast cancer are Evista (Raloxifene Hydrochloride, Soltamox (Tamoxifen Citrate) and Tamoxifen Citrate. There are various combinations of chemotherapy drugs used in breast cancer treatment. These combination drugs are shown in Table 8 (Arora et al., 2022).

Table 2.7. Chemotherapy drug combinations used in breast cancer treatment (Fisusi & Akala, 2019).

Treatment Methods	Drugs
AC treatment	A = Doxorubicin Hydrochloride (Adriamycin) C = Cyclophosphamide
AC-T treatment	A = Doxorubicin Hydrochloride (Adriamycin) C = Cyclophosphamide T = Paclitaxel (Taxol)
CAF treatment	C = Cyclophosphamide A = Doxorubicin Hydrochloride (Adriamycin) F = Fluorouracil
CMF treatment	C = Cyclophosphamide M = Methotrexate F = Fluorouracil
FEC treatment	F = Fluorouracil E = Epirubicin Hydrochloride C = Cyclophosphamide
TAC Treatment	T = Docetaxel (Taxotere) A = Doxorubicin Hydrochloride (Adriamycin) C = Cyclophosphamide

2.6.5.4. Hormone Therapy

Hormonotherapy is a form of treatment that aims to stop the growth of cancer cells by blocking the release or effects of hormones. Hormones are substances produced in the body and reaching the organ where they affect through the blood. In some types of breast cancer that are affected by hormones such as estrogen and progesterone, breast cancer cells have receptors (proteins) that bind to estrogen and progesterone, which helps them grow. Treatments that stop the receptors that bind hormones to estrogen and progesterone are called hormone or endocrine therapy. Cancer cells that contain these receptors are called estrogen receptor positive (ER+) or progesterone receptor positive (PR+). Hormone therapy is often used after surgery (as adjuvant therapy) to help reduce the risk of cancer coming back. The most important criterion that determines whether breast cancer patients are sensitive to hormonal treatment is the presence of hormone receptors in the tumor tissue. Estrogen and progesterone receptors play an important role in treatment planning and are the determinants of response to adjuvant endocrine therapy. Hormonal therapy is effective for hormone receptor positive breast cancers. Hormonal therapy blocks the estrogen

hormone, preventing the hormone from becoming active and stimulating the proliferation of breast cancer cells. The aim of hormonal therapy is to eliminate cancer cells that may remain anywhere in the body after initial treatment with surgery, radiotherapy and chemotherapy. In addition, it reduces the risk of cancer recurrence in non-invasive breast cancers (DCIS) and is used to shrink large tumors before any treatment in invasive breast cancers. Hormonal therapy is available for both pre- and post-menopausal women (Drăgănescu & Carmocan, 2017).

Additionally, hormone and hormone antagonist drugs used in cancer treatment have an important place in the treatment of breast cancer. Drugs used in hormonal therapy are examined in various groups, including selective estrogen modulators (SERMs), aromatase inhibitors (AIs), luteinizing hormone-releasing hormone (LHRH) analogues and selective estrogen receptor degraders (SERDs) (Drăgănescu & Carmocan, 2017).

Table 2.8. Classification of drugs used in breast cancer hormone therapy (Drăgănescu & Carmocan, 2017).

Selective Estrogen Modulators (SERMs)	**Aromatase Inhibitors (AIs)**	**Luteinizing Hormone-Releasing Hormone (LHRH)**	**Selective Estrogen Receptor Degraders (SERDs)**
Tamoxifen (Nolvadex)	Anastrozole (Arimidex)	Goserelin (Zoladex)	Fulvestrant (Falsodex)
Toremifene (Fareston)	Letrozole (Femara)	Leuprolide (Lupron)	Elacestrant (Orserdu)

- **Selective Estrogen Modulators (SERMs):** SERMs, which are drugs used in breast cancer, block the estrogen receptor and prevent estrogen from arriving at the estrogen receptor and they come where they are needed. Thus, they cannot bind to the estrogen receptor and stimulate cell proliferation (Drăgănescu & Carmocan, 2017).

- **Aromatase Inhibitors (AIs):** Aromatase inhibitors stop the enzyme called aromatase, which converts androgens into estrogens, thus reducing the amount of estrogen produced outside the ovaries. This means there is less estrogen in the bloodstream, less estrogen reaches estrogen receptors, and less cancer cell proliferation (Drăgănescu & Carmocan, 2017).
- **Luteinizing Hormone-Releasing Hormone (LHRH):** They stop the signal the body sends to the ovaries to make estrogen, causing temporary menopause. The main goal of treatment is to take advantage of the negative feedback and suppression mechanism between LHRH and luteinizing hormone (LH). LHRH analogs act by blocking LH secretion at the anterior pituitary level (Drăgănescu & Carmocan, 2017).
- **Selective Estrogen Receptor Degraders (SERDs):** A selective estrogen receptor degrader (SERD) is a type of medication that binds to the estrogen receptor (ER) and in doing so causes the estrogen receptor to be disrupted and therefore downregulated (Drăgănescu & Carmocan, 2017).

CHAPTER III. MATERIALS AND METHODS

In this theoretical study, breast cancer-causing receptors were modeled in the presence of ligands using high performance computing. Using Molecular Docking (MD), Molecular Dynamic Simulation (MDS) and Density Functional Theory (DFT), the biological activity, IC50 values, affinities and other properties of ER (estrogen receptor) (PDB ID: 7UJM and 1R5K), BRCA1 gene (PDB ID: 4IFI) and BRCA2 gene (PDB ID: 1N0W) receptors, which are most commonly affected by breast cancer, and gallic acid and catechin, which are grape seed flavonoids, were evaluated by PubChem and web-based servers such as ChEMBL were used. As a result of literature research on the polyphenols found in grape seeds that have anticancer properties, the molecular formulas and structures of gallic acid, catechin and epicatechin flavonoids were obtained from the chemical database PubChem. In this study, breast cancer literature research on the RCSB PDB (Research Collaboratory for Structural Bioinformatics Protein Data Bank) site for 7UJM, 1R5K, 4IFI and 1N0W receptors was evaluated. CHEMBL codes of the ligands were obtained by considering the IC50 values of the ligands using UniProtKB (Universal Protein Resource Knowledgebase), a database containing biological and biochemical protein information.

Table 3.1. 3D image of 7UJM (dimer) and 1R5K (dimer) receptors obtained from UCSF Chimera.

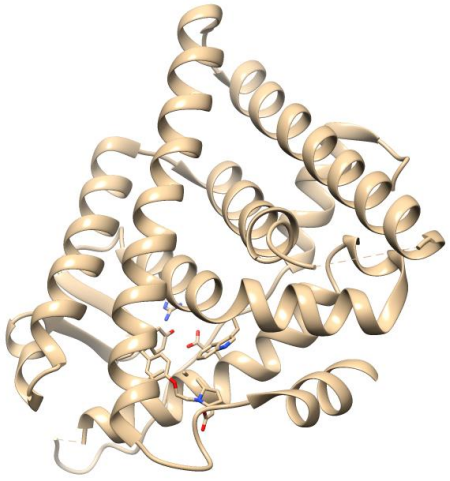
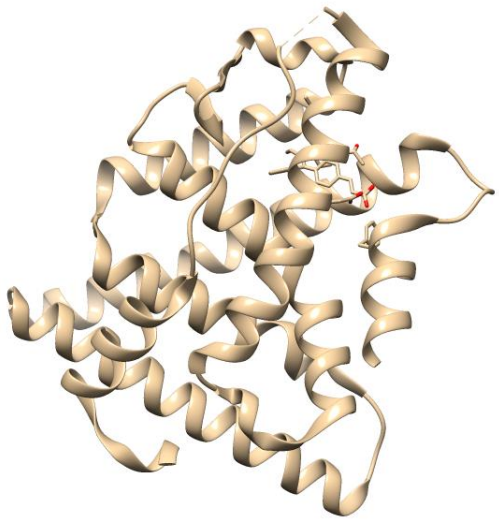
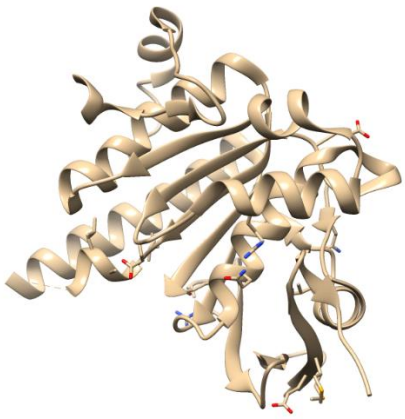
	
7UJM (ESTROGEN RECEPTOR)	1R5K (ESTROGEN RECEPTOR)

Table 3.2. 3D image of 4IFI (dimer) and 1N0W (dimer) receptors obtained from UCSF Chimera.

	
4IFI (BRCA1 Gene)	1N0W (BRCA2 Gene)

Additionally, ligands with low IC50 values were screened in the literature from clinical databases and ChEMBL. It was simulated and modeled in the presence of water and ions. For each receptor, 100 ligands, 100 flavonoids and 21 drugs were docked and the ligands with the lowest binding energies (kcal/mol) were selected. Receptor and ligand were prepared with the help of Autodock Vina. Molecular Dynamics Simulation was performed with the help of GROMACS. The National Center for High Performance Computing (UHeM) was used as the database for calculations. The results were recorded with RMSD, RMSF, Hydrogen Bond, Gyration Radius (Rg), Interaction Energy of Protein Ligand, SASA (Solvent Accessible Surface Area), Radial Distribution Functions (RDF) and MSD (Mean Square Displacement) analysis performed after MDS. Finally, by determining the electron distributions and chemical reactivity indices of the ligands with Density Functional theory, how the ligands interact with proteins was obtained at the molecular level.

3.1. Grape Seed Flavonoids

Gallic acid, catechin and epicatechin flavonoids, found in grape seeds and known for their anticancer properties, were selected as the main ligands and their molecular structures and formulas were obtained from the PubChem database. CHEMBL codes of receptor and ligand chemical structures are shown in Table 3.3.

Table 3.3. Grape seed flavonoids, receptor-ligand and drug CHEMBL codes.

RECEPTOR	7UJM (ER)	1R5K (ER)	4IFI (BRCA1)	1N0W (BRCA2)
LIGAND	CHEMBL4576272	CHEMBL4164133	CHEMBL1784770	CHEMBL3823654
LIGAND (FLAVONOID)	476783 (Epicathecin)	107905 (Epicathecin)	476783 (Epicathecin)	5276454 (Cathecin)
DRUG	CHEMBL2355051 (Clomiphene)	CHEMBL83 (Tamoxifen)	CHEMBL1200973 (Estradiol Cypionate)	CHEMBL1200973 (Estradiol Cypionate)

3.2. Protein Preparation

Protein receptors directly associated with breast cancer, ER (estrogen receptor) (7UJM and 1R5K), BRCA1 (4IFI) and BRCA2 (1N0W), were saved from the RCSB PDB database in PDF format into folders for pre-processing in AutoDock Tools. In order to be able to use these proteins in docking studies and to prepare for Molecular Dynamics Simulations, water molecules were removed, polar hydrogen atoms were added and protein structures were optimized for docking in the AutoDock Tools program.

3.3. Ligand Preparation

Ligands, flavonoid structures gallic acid, catechin and epicatechin and drugs were prepared with 3D structures downloaded from the PubChem database. It is ready for molecular docking with AutoDock Tools and AutoDock Vina.

3.4. Protein-Ligand Docking

The prepared proteins and ligands were energetically optimized by molecular docking using AutoDock Tools and AutoDock Vina programs. Thus, energetic modeling was carried out. A grid box was set around the active regions of 7UJM, 1R5K, 4IFI and 1N0W receptors and docking parameters were defined. Docking simulations were performed to determine the interaction modes and binding affinities of the ligands with target proteins.

3.5. Molecular Docking Procedure with AutoDock Tools and AutoDock Vina

AutoDock Tools and AutoDock Vina are a software program used for protein-ligand docking studies, which is a molecular modeling simulation. In addition, it is a program that contributed to the discovery of many drugs using inhibitors. In this way, this software can also be used to calculate molecular surfaces, view secondary structure strands, and calculate hydrogen bonds. The binding affinities of the ligands to breast cancer receptors were calculated in an aqueous environment using AutoDock Tools and AutoDock Vina.

PDBQT formats of receptors and ligands have been prepared to perform molecular docking studies. For this, first, target protein receptors (7UJM, 1R5K, 4IFI and 1N0W) were downloaded in PDB format from the RCSB PDB database. By removing water molecules and other small molecules in the protein structure, polar hydrogen atoms were added and the 7UJM, 1R5K, 4IFI and 1N0W protein structure was appropriately optimized. Kollman charges are charged after the polar hydrogen atoms. These charges, based on the orbital electron density of the molecule, are calculated using the AMBER force field. These charges, which will reveal how the ligand binds to the 7UJM, 1R5K, 4IFI and 1N0W receptors and which conformations are most energetically favorable, help to accurately model intermolecular interactions.

In order to optimize the ligands (flavonoids, drugs) selected from the PubChem database and bring them to the correct geometric structure, their 3D structures were downloaded from the PubChem database and the grid box was adjusted around the active regions. In order to accurately identify the active sites on 7UJM, 1R5K, 4IFI and 1N0W receptors and to increase the accuracy of docking results, the active regions in the structure of the receptors where the ligand is expected to bind were determined by adjusting the grid box. The dimensions and center of the grid box are adjusted to cover the entire active region. Thus, the center coordinates were determined.

Additionally, taking into account the coordinates containing the active region, the grid box dimensions were determined as shown in the table below:

Table 3.4. Dimensions and centers of the Grid box of the receptors with AutoDock Vina.

RECEPTORS	7UJM (ER)	1R5K (ER)	4IFI (BRCA1)	1N0W (BRCA2)
center_x	-0.639	-20.278	-10.778	34.518
center_y	1.778	40.968	-0.303	26.562
center_z	18	363.661	-15.361	2.214
size_x	80	80	100	80
size_y	80	80	78	80
size_z	80	80	120	80
num modes = 100 energy range = 4				

Definition of docking parameters was carried out using AutoDock Tools and AutoDock Vina also, to see the lowest energy and most probable binding modes, 100 binding modes were set and the energy range was determined as 4. In summary, the 7UJM, 1R5K, 4IFI and 1N0W receptors and ligand files prepared for molecular docking were added to the folder, respectively, then the Grid box and parameters were adjusted and finally the binding modes determined for Docking were calculated. Calculated binding energies (in kcal/mol) for each binding mode were examined for all receptors. The modes with the lowest binding energy, representing the most probable binding conformations, and the binding sites and important interactions such as hydrogen bonds, hydrophobic interactions in these regions were analyzed.

3.6. Molecular Dynamics Simulations Procedure with GROMACS Tutorial

GROMACS was used to perform molecular dynamics simulations, and the stability and behavior of selected protein-ligand complexes in a simulated biological environment were examined after docking analyzes. Ligand-receptor complexes with the lowest binding energy according to docking analysis results, which will help understand how ligands bind to target receptors and the energetic and structural properties of these bindings, have been selected as shown in Table 3.5.

Table 3.5. Ligand-receptor complexes and CHEMBL ID.

Receptors	Ligands and drugs with CHEMBL ID	Selected ligand and drug
7UJM (ER)	CHEMBL4576272	ligand82
	476783 (Epicathecin)	ligand100
	CHEMBL2355051 (Clomiphene)	drug5
1R5K (ER)	CHEMBL4164133	ligand29
	107905 (Epicathecin)	ligand93
	CHEMBL83 (Tamoxifen)	drug16
4IFI (BRCA1)	CHEMBL1784770	ligand12
	476783 (Epicathecin)	ligand100
	CHEMBL1200973 (Estradiol Cypionate)	drug17
1N0W (BRCA2)	CHEMBL3823654	ligand70
	5276454 (Cathecin)	ligand83
	CHEMBL1200973 (Estradiol Cypionate)	drug17

The PDB file of the protein-ligand complex with the best binding mode obtained from AutoDock Vina was saved in the GROMACS compatible format (PDB or GRO) by checking the correct atom types and structural integrity of the complex in the UCSF Chimera. Ligands identified from the PubChem database were minimized using UCSF Chimera. The prepared receptor and minimized ligand files were selected together, the coordinates were determined and the executable location was applied in UCSF Chimera, thus docking files were created in the folder. The lowest scoring protein-ligand complexes were created as Pose1.pdb files.

LIG files (LIG.mol2, LIG.rtf, LIG.pdb, LIG.crd, LIG.itp, LIG.par, LIG,psf) created with UCSF Chimera were obtained using SwissParam for the topology and parameters of small organic molecules compatible with the force atom whose force field is CHARMM. Files (EM.mdp, ions.mdp, MD.mdp, NPT.mdp, NVT.mdp, sort_mol2_bonds.pl) obtained from GitHub GROMACS Protein-Ligand were also downloaded and transferred to the UHeM database to be used in molecular dynamics simulation. It is important for force fields to be polarized in order to carry out accurate simulations in environments with changing dielectricity, and in this study, the appropriate force field for GROMACS (for example, CHARMM, AMBER or OPLS-AA) was chosen as CHARMM37.

Table 3.6. Ligand files and GitHub GROMACS Protein-Ligand files and descriptions

File Name	File Type	Description
LIG.mol2	mol2 file	It contains bonds, coordinates and atoms of the molecular structure.
LIG.rtf	rtf (residue topology file)	It defines ligand topology (such as atom angle and bond).
LIG.pdb	pdb (protein data bank file)	It contains three-dimensional coordinates of molecular structures.
LIG.crd	crd (coordinate file)	It contains the atomic coordinates of molecules.
LIG.itp	itp (include topology parameter file)	It contains the bonding parameters and topology of the molecule.
LIG.par	par (parameter file)	It contains force field parameters used in MDS.
LIG.psf	psf (protein structure file)	The dihedral angles of the protein structure include atoms and bonds.
EM.mdp	EM (Energy Minimization) file	It is used for energy minimization in GROMACS.
ions.mdp	ion addition file	It is used to add ions to the simulation process in GROMACS.
NVT.mdp	Constant Number, Volume, and Temperature file	It is used in GROMACS for temperature equilibrium.
NPT.mdp	Constant Number, Pressure, and Temperature file	It is used in GROMACS for pressure equilibrium.
MD.mdp	Molecular Dynmacis file	It contains the necessary simulation conditions for MDS in GROMACS, such as time, step, temperature and pressure.
sort_mol2_bonds.pl	Perl Script	It is used in modeling to sort or manipulate bond information in MOL2 files.

In addition, after topology files, which are appropriate force field parameter files for the protein and ligand, were created, the protein-ligand complex was placed in a box with dimensions to contain all the atoms of the complex and a sufficient amount of water molecules for the solvatation. The TIP3P water model was chosen for its computational efficiency and water molecules were added.

The edges of the box, which was determined as cubic, were run as 2 nm. In the part after solvation and ion addition, GROMACS' mdrun command was run in the UHeM database with the sbatch file for energy minimization in order to neutralize the system and remove high-energy conformations and while EM.mdp was used as the

parameter file, PME (Particle Mesh Ewald) for long-range electrostatic interaction was used. NVT balancing simulation at constant temperature was performed to stabilize the temperature of the system, and NPT balancing simulation was performed at constant pressure to stabilize the pressure of the system, with a duration of 100 ps with NVT.mdp and NPT.mdp files. In the balancing process carried out with NVT (constant number, volume, and temperature, the temperature thermostat parameter tcoupl = V-rescale in the NVT.mdp file was used. NPT.mdp pcoupl = Parrinello-Rahman parameter was used in the balancing process with NPT (constant number, pressure and temperature).

After NVT and NPT equilibration to reach the appropriate conformation and reference energy, MD process was carried out with the MD.mdp file before the analysis. Each simulation was performed with 0.002 femtoseconds and 25.000.000 steps, thus obtaining 50 ns using the MD.mdp file.

3.7. Molecular Dynamics Simulations Procedure Tutorial 5 (Protein-Ligand Complex) Analysis

3.7.1. Clustering Analysis (Frame Selection)

Cluster analysis was performed to determine different conformational states of the protein-ligand complex. Thus, the determined conformational states were grouped. An RMSD matrix was obtained by calculating the RMSD values of each structure, and representative frames of each cluster were selected according to the cluster analysis results.

3.7.2. RMSD (Root Mean Square Deviation)

Structural changes of the protein-ligand complex over time were observed by measuring the difference between the RMSD value and the actually observed value and the predicted value after the created model. RMSD value is calculated with the formula below (Sargsyan, Grauffel, & Lim, 2017).

di is the distance between atom ,

i in the two structures, and

N is the total number of equivalent atoms,

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N} di^2}{N}}$$

The RMSD.xvg result was made ready for analysis with molecular dynamics simulation by running the "gmx rms -s MD.tpr -f MD_noPBC.xtc -o rmsd.xvg -tu ns" code in the GROMACS Codes file.

3.7.3. RMSF (Root Mean Square Fluctuations)

The RMSF value evaluates the flexibility of different regions of the protein by showing how mobile the atoms are in fixed or variable regions and provides information about how much they deviate from their reference positions. RMSF is calculated atom i which is a group of T structures (Pitera, 2014).

$$RMSF = \sqrt{\frac{1}{T}\sum_{j=1}^{T}\left[\vec{r}_{j,i} - \frac{1}{T}\sum_{k=1}^{T}\vec{r}_{k,i}\right]^2}$$

The RMSF.xvg result was made ready for analysis with molecular dynamics simulation by running the "gmx rmsf -s MD.tpr -f nojump.xtc -o rmsf.xvg" code in the GROMACS Codes file.

3.7.4. Hydrogen Bond Analysis (H-Bond)

The position and number of hydrogen bonds formed between the protein and the ligand within 0.35 nm during the MDS period were determined by hydrogen bond analysis (Prasad & Chakraborty, 2022).

r_{DA} is the donor-acceptor distance ,

θ_{DHA} is the donor-hydrogen-acceptor angle,

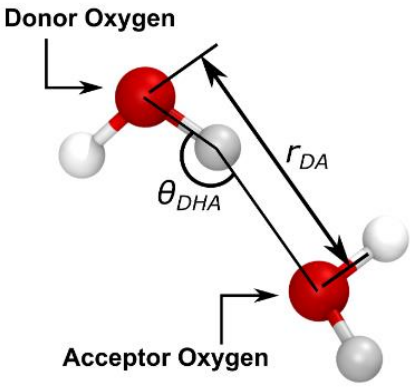


Figure 3.1. Image showing the identification of hydrogen bonding with acceptor and donor atoms in a molecular dynamics simulation (Guide, 2022).

The hb00.xvg results was made ready for analysis with molecular dynamics simulation by running from the "gmx hbond -s MD.tpr -f nojump.xtc -num hb04.xvg -b 0 -e 4 -tu ns" to "gmx hbond -s MD.tpr -f nojump.xtc -num hb4650.xvg -b 46.01 -e 50 -tu ns" code in the GROMACS Codes file.

3.7.5. Radius of Gyration (Rg)

Rg, defined as the root mean square distance of the atoms in a molecule from their center of mass, has been used to monitor the compactness of protein structure over time. Conformational changes over time were evaluated. Rg value is calculated with the formula below (Sadiku-Agboola & Sadiku, 2014).

m_i is identical mass,

r_i is variable distance,

$$Rg^2 = \frac{\sum m_i {r_i}^2}{\sum m_i}$$

The gyrate1.xvg result was made ready for analysis with molecular dynamics simulation by running the "gmx gyrate -s MD.tpr -f nojump.xtc -o gyrate1.xvg" code in the GROMACS Codes file.

3.7.6. Interaction Energy of Protein Ligand (IE)

There should be an IE.mdp file in the folder during the protein-ligand binding energy process, which defines the changes such as binding stability and affinity that occur when the ligand binds to the protein. Total protein-ligand interaction energy was obtained by van der Waals (LJ-SR) and electrostatic (Coul-SR) interactions.

While Van der Waals Interactions (Lennard-Jones Potential) arise from the attractive and repulsive forces between the ligand and protein atoms, Electrostatic Interactions (Coulomb Potential) arise from the mutual interactions of the charges on the ligand and protein (Bitencourt-Ferreira & de Azevedo Junior, 2021).

The protein-ligand interaction energy result was made ready for analysis with molecular dynamics simulation by running the "gmx grompp -f IE.mdp -c NPT.gro -t NPT.cpt -p topol.top -n index.ndx -o IE.tpr" and "gmx mdrun -deffnm IE -rerun MD.xtc -nb cpu" codes in the GROMACS Codes file. Then, the interaction_energy.xvg and Coul-SR:Protein-LIG and LJ-SR:Protein-LIG results were obtained with "gmx energy -f IE.edr -o interaction_energy.xvg".

3.7.7. SASA (Solvent Accessible Surface Area)

SASA, which gives the solvent accessible surface area, provided information on molecular stability to determine the exposure of the protein and ligand to the solvent. After determining the radius of the solvent molecule, the van der Waals radius of the water molecule is used (~1.4 Å) for SASA calculations (Durham, Dorr, Woetzel, Staritzbichler, & Meiler, 2009).

The sasapl.xvg result was made ready for analysis with molecular dynamics simulation by running the "gmx sasa -f nojump.xtc -s MD.tpr -n index.ndx -o sasapl.xvg -tu ns" code in the GROMACS Codes file and then, the sasap.xvg result was made ready for analysis with molecular dynamics simulation by running the "gmx sasa -f nojump.xtc -s MD.tpr -n index.ndx -o sasap.xvg -tu ns" code.

3.7.8. Radial Distribution Functions (RDF)

RDF, which is used to understand molecular structure and interactions, the distance distributions between atoms or molecules are examined. RDF, g(r) value is calculated with the formula below (Hangboce, 2019).

$\rho_{(r)}$, system density (average particle density),

ρ_0 and $\rho_{(r)}$ represent the local and average atomic numberdensity, respectively,

$$RDF(r) = 4\pi r^2 \rho_{(r)} = 4\pi r^2 \rho_0 + rG_{(r)}$$

$$G_{(r)} = \frac{2}{\pi}\int_0^{q_{max}} q(S_{(q)} - 1)\, sin_{(rq)}\, d_q$$

The rdf.xvg result was made ready for analysis with molecular dynamics simulation by running the "gmx rdf -s MD.tpr -f nojump.xtc -o rdf.xvg" code in the GROMACS Codes file.

3.7.9. MSD (Mean Square Displacement)

Which is used to measure the displacement of atoms or molecules over time, mobilities are detected and diffusion behaviors are examined with MSD. The MSD formula that calculates the diffusion coefficient is shown below (Wang, Jeong, Jhiang, Yu, & Menq, 2014).

$$MSD = \bar{\rho}_{A,N} = \frac{\sum_{i=1}^{N-n} I_{i,i+n}{}^2}{N-n}$$

$$\bar{\rho}_{A,1} = (I_{1,2} + I_{2,3} + .. + I_{12,13})/12$$

$$\bar{\rho}_{A,2} = (I_{1,3} + I_{2,4} + .. + I_{11,13})/11$$

$$\bar{\rho}_{A,3} = (I_{1,4} + I_{2,5} + .. + I_{10,13})/10$$

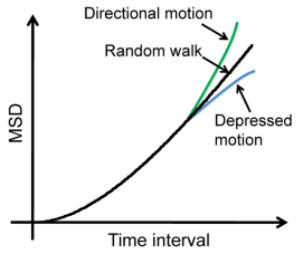


Figure 3.2. Graph showing curves over different time periods for MSD calculation (Wang et al., 2014).

The msd.xvg result was made ready for analysis with molecular dynamics simulation by running the “gmx msd -s MD.tpr -f nojump.xtc -o msd.xvg” code in the GROMACS Codes file.

3.8. Density Functional Theory (DFT) Analysis with ORCA 5.0

The complexes with the lowest binding energy were selected according to the docking results obtained with Autodock Tools for flavonoids. It is shown in Table 3.7. Pdb extension files were created by determining H-bonding amino acids using Pymol. Then, the 3D chemical structures of amino acids and ligand were analyzed from the literature using Avagadro, and Hydrogen and Oxygen addition and removal operations were carried out. Input files for Orca were created with the help of Avagadro. Files with inp extension were created by selecting DFT, B3LYP/def2-SVP basis set for geometry optimization. The created files are ready for ORCA (Forum, 2024).

Table 3.7. Flavonoid conformations with the lowest binding energy selected for ORCA analysis after Autodock Tools.

Receptors	Models	# of Run/ Population Size	Lowest Binding Energy [kcal/mol]	Amino-acid	Conformation
7UJM (ER)	Epicathecin_ ligand100	100/300	-7.16	LYS/GLU	conf84_flav100_dft.pdb/inp
1R5K (ER)	Epicathecin_ ligand93	100/300	-8.32	GLY/LEU	conf28_flav93_dft.pdb/inp
4IFI (BRCA1)	Epicathecin_ ligand100	25/150	-5.95	GLU/ASN	conf23_flav100_dft.pdb/inp
1N0W (BRCA2)	Cathecin_ ligand83	5/50	-6.46	GLU/LYS	conf3_flav83_dft.pdb/inp

CHAPTER IV. RESEARCH FINDINGS AND DISCUSSION

In this study on breast cancer, conformational changes were examined with molecular docking, molecular dynamics simulation and density functional theory. After a detailed literature research, 4 receptors were selected for estrogen receptors (7UJM and 1R5K), BRCA1 gene (4IFI) and BRCA2 gene (1N0W), which were selected and had the most direct effect on breast cancer, and docking studies were carried out in the active regions. A total of 377 ligands, 100 flavonoids and 21 drugs found in the literature for receptors were studied (The docking results of the receptors with Autodock Vina are shown in APPENDIX A). According to the docking results obtained, 4 ligands (CHAINA_ligand82, CHAIN_ligand29, CHAINA_ligand12, CHAINB_ligand70), 4 flavonoids (Epicathecin_ligand100, Epicathecin_ligand93, Epicathecin_ligand100, Cathecin_ligand83) and 4 drugs (Drug_ligand5, Drug_ligand16, Drug_ligand100, Drug_ligand83) with the lowest binding energy and the highest binding interaction energy were determined and protein-ligand simulations were carried out with molecular dynamics simulations. According to the docking results, protein-ligand pairs with the lowest binding energy and the highest binding interaction energy were selected and made ready for molecular dynamics simulation. Molecular Dynamic Simulation and Density Functional Theory methods with by using selected receptors and ligands together , a specific modeling study was carried out with grape seed flavonoids, which are supported by studies in the literature that show anticancer properties in the treatment of breast cancer.

4.1. Molecular Docking Findings

Parameters such as binding energies of molecules, physicochemical properties, interactions with each other and biological activities were obtained with molecular docking. While molecular docking was performed, docking was carried out using two different programs: Autodock Tools and Autodock Vina. The aim of making Autodock Tools was to support the low binding energies obtained from Autodock Vina and to see the inhibition constant that Autodock Vina could not provide. Molecular docking studies were conducted using estrogen receptors (7UJM and 1R5K), BRCA1 gene (4IFI) receptor and BRCA2 gene (1N0W) receptors and grape seed flavonoids selected as a result of literature research for breast cancer. While the receptor structures (7UJM, 1R5K, 1N0W and 4IFI) to be used for docking were obtained from the Protein Data Bank (PDB), the ligand structures (flavonoids and drugs) were obtained from PubChem. The docking grid box is sized large enough to include binding site residues, centered on the active site of each receptor.

Table 4.1. Ligand-receptor complexes with lowest binding energy with AutoDock Vina.

Receptors	Ligands and drugs with CHEMBL ID	Lowest Binding Energy (kcal/mol)
7UJM (ER)	CHEMBL4576272 “ligand82”	-11.4
	476783 (Epicathecin) “ligand100”	-9.5
	CHEMBL2355051 (Clomiphene) “drug_ligand5”	-9.6
1R5K (ER)	CHEMBL4164133 “ligand29”	-12.2
	107905 (Epicathecin) “ligand93”	-10.6
	CHEMBL83 (Tamoxifen) “drug_ligand16”	-9.4
4IFI (BRCA1)	CHEMBL1784770 “ligand12”	-4.4
	476783 (Epicathecin) “ligand100”	-4.5
	CHEMBL1200973 (Estradiol Cypionate) “drug_ligand17”	-4.8
1N0W (BRCA2)	CHEMBL3823654 “ligand70”	-7.6
	5276454 (Cathecin) “ligand83”	-8.2
	CHEMBL1200973 (Estradiol Cypionate) “drug_ligand17”	-8.5

Binding affinity values (kcal/mol) for the ligand-receptor pair selected among 100 ligands, 100 flavonoids, and 100 drugs are shown in Table 4.1 (The docking results of the receptors with Autodock Vina are shown in APPENDIX A). Among the grape seed flavonoids, Epicatecin with the lowest binding energy was selected for

7UJM and 1R5K estrogen receptors and 4IFI BRCA1, while Cathecin flavonoid was selected for 1N0W BRCA2.

4.1.1. Molecular Docking Studies with 7UJM (ER), Ligands and Drugs

Estrogen receptors (ER), which are proteins that work by binding to the estrogen hormone that regulates the growth and proliferation of cells, are divided into two groups: ERα and ERβ, and when ERα binds to the estrogen hormone, it has a triggering effect on tumor growth and cell proliferation.

7UJM, the estrogen receptor, has an important place in breast cancer, and ligands and drugs that interact with this receptor prevent estrogen from binding to this receptor and stop the growth of cancer cells by modulating the receptor activity. Therefore, it is of great importance to discover potential new therapeutic agents in the treatment of breast cancer by examining molecules that bind to the 7UJM receptor. In this thesis study, grape seed flavonoids were used as organic compounds. According to RCSB data, 7UJM is provided together with ligands from RL4 molecule. While docking studies in which the structure of the ligands and receptors were flexible were carried out with AutoDock Vina, rigid docking and blind docking studies were carried out with AutoDock Tools. Suitable grid box dimensions and grid box coordinates for rigid docking studies are shown in Table 4.2 for the 7UJM receptor and docking parameters are shown in Table 4.3. After completing the necessary steps in the AutoDock program for the 7UJM receptor, it was run in **5 steps**, giving the population **size 50, 25 steps**, giving the population **size 150 and 100 steps**, giving the population **size 300.**

Table 4.2. Grid parameter file of 7UJM receptor and ligands with AutuDock Tools.

GRID PARAMETER FILE			
Macromolecule	**7UJM (ER)**		
Models	CHAINA_ligand82	Epicathecin_ligand100	Drug_ligand5
# of grind points in xyz	104 100 98	106 126 126	126 126 126
Spacing (Å)	0.375	0.375	0.375
xyz grid-center coordinates	-10.033	0.659	-1.425
	9.267	2.097	1.655
	13.917	21.996	18.883
# of torsional degrees of freedom	5	5	5
Temperature (K)	298.15		
Charges	Kollman		

The docking parameters obtained accordingly are shown in Table 4.3 below.

Table 4.3. Docking parameter file of 7UJM receptor and ligands with AutoDock Tools.

DOCKING PARAMETERS	
Program	Autodock
Search Parameters	Lamarckian
	Genetic Algorithm
# of Run	5 / 25 / 100
# of Generations	27000
maximum # of energy evulations (short level)	250000
Population Size	50 / 150 / 300

Additionally, grid box dimensions and grid box coordinates for rigid and blind docking results are shown in Table 4.4 for the 7UJM receptor.

Table 4.4. Rigid and blind docking results of 7UJM receptor with the best score.

Models	# of Run/ Population Size	Cluster Rank	Lowest Binding Energy [kcal/mol]	Number of Run	Number in Clusters	Inhibition Constant [μM]
CHAINA_ligand82	5/50	1	-8.98	5	4	0.2613
	25/150	1	-11.79	19	3	0.0023
	100/300	1	-11.63	16	20	0.0030
Epicathecin_ligand100	5/50	1	-5.26	1	1	138.87
	25/150	1	-5.14	12	1	170.45
	100/300	1	-7.16	84	13	5.65
Drug_ligand5	5/50	1	-8.09	3	1	1.17
	25/150	1	-10.25	20	3	0.0305
	100/300	1	-10.43	76	5	0.0226

CHAINA_ligand82 and Drug_ligand5 shows lower binding energies and lower inhibition constants at larger population sizes, implying stronger binding and a more effective inhibition potential. Epicatechin_ligand100 showed the lowest binding energy (-7.16 kcal/mol) and lower inhibition constant (5.65 μM), especially at the population size of 100/300, suggesting that there may be stronger binding and more effective inhibition potential under certain conditions. In addition, the clustering numbers in different conditions indicate that the ligands can bind to the target protein in different conformations.

As a result, according to the docking results, CHAINA_ligand82 generally showed the lowest binding energies and lowest inhibition constants, indicating that the CHAINA_ligand82 model may have a stronger binding and inhibition potential than other ligands.

According to the Autodock Tools results, the conformations with the lowest binding energy were selected and their images were obtained with Discovery Studio. Figure 4.1, Figure 4.2 and Figure 4.3 shows the structure of amino acids in the conformation obtained after docking for the 7UJM receptor and ligand82, flav100 and drug5.

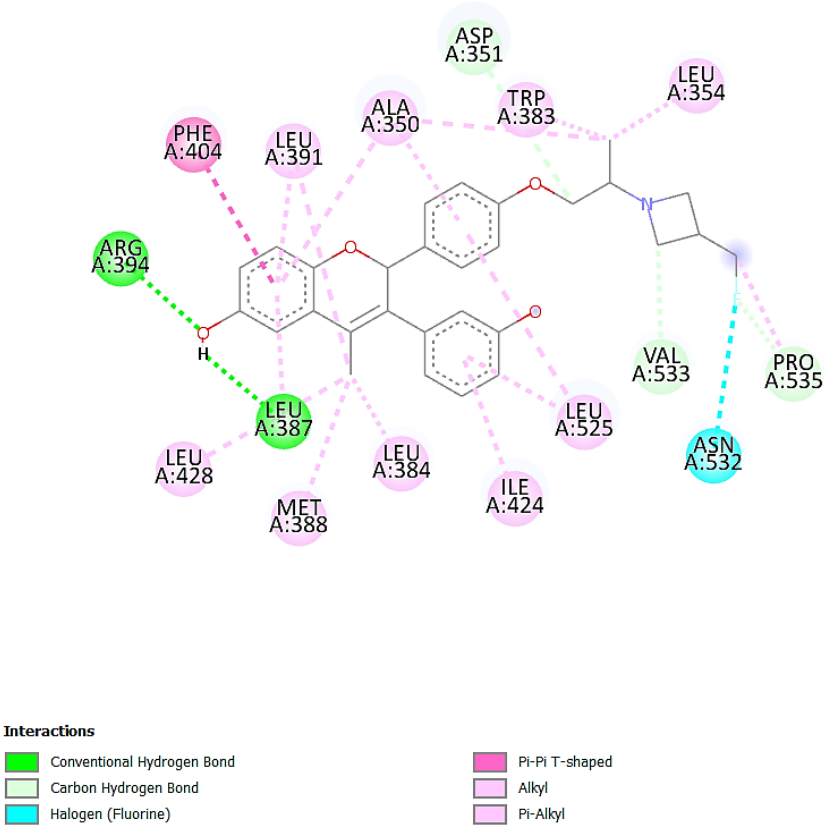


Figure 4.1. Discovery Studio 2D ligand-protein interaction diagram showing ligand82 (conf19_liga82.pdbqt) amino acid interactions with the 7UJM receptor.

According to the figure 4.1, there are hydrogen bonds, halogen bonds, Pi-Pi T-Shaped and Pi-Alkyl interactions. Hydrogen bonds are the strongest specific interactions and help the ligand stay in the correct position. Halogen bonds are a special type of interaction and increase the effectiveness of the ligand by contributing to the binding energy. Pi-Pi interactions and alkyl interactions ensure the stabilization of the ligand in the binding site (Vinod et al., 2023; Yang et al., 2016).

Ligand82 binds strongly to the 7UJM receptor through hydrogen bonds with amino acids ARG394 and ASN532. This interaction increases the potential of ligand82

to function as an effective inhibitor. Ligand82 contributed to the binding energy by forming a halogen bond with ASN532. Halogen bonds are especially important for fluorine-containing ligands. Ligand82 forming a halogen bond with ASN532 causes this molecule to provide additional stabilization in the binding region. Alkyl interactions of ligand82 with TRP383 and other hydrophobic amino acids indicate that a significant hydrophobic attraction is achieved at the binding site.

According to the figure 4.2, GLU380, LYS520, ASN519, GLU523: These amino acids increased the binding affinity by interacting with the hydrogen bonds formed by the flav100. MET522, which has a Pi-Sulfur interaction, the Pi-Sulfur interaction of the ligand with the sulfur-containing methionine strengthens the hydrophobic interactions. TYR526, VAL534, TRP383 interactions are hydrophobic interactions with the alkyl groups of the aromatic rings of the ligand and strengthen the binding of flav100 to the receptor.

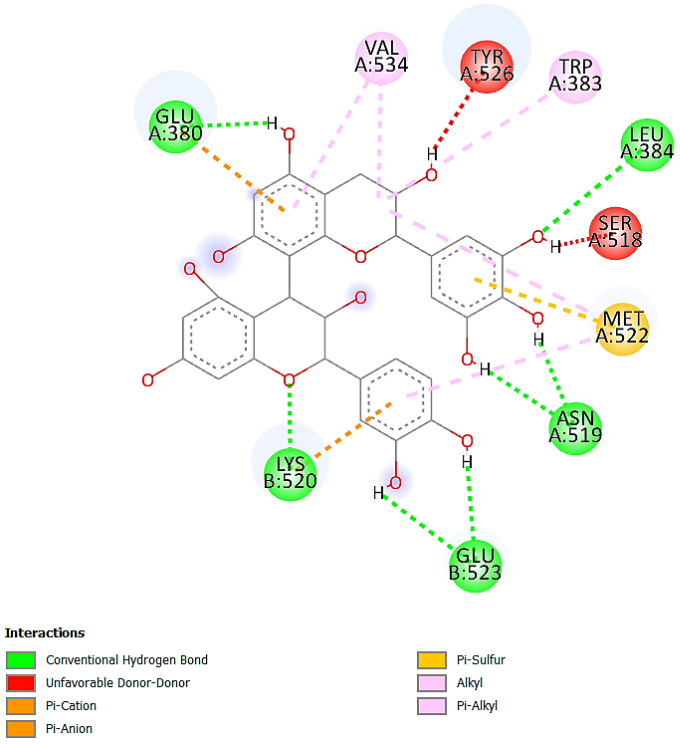
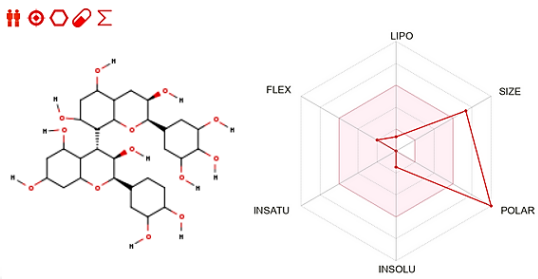


Figure 4.2. Discovery Studio 2D ligand-protein interaction diagram showing flav100 (conf84_flav100.pdbqt) amino acid interactions with the 7UJM receptor.

SwissADME is a web tool that predicts properties of a molecule such as pharmacokinetics (how it acts in the body), bioavailability (how effective it is in the body), lipophilicity (fat solubility), water solubility, etc. These features, which are important in drug design, are used to evaluate whether the molecule could be a potential drug candidate (Daina, Michielin, & Zoete, 2017).



SMILES OC1CC(O)C2C(C1)O[C@@H]([C@@H]([C@H]2C1C(O)CC(C2C1O[C@@H]([C@@H](C2)O)C1CC(O)C(C(C1)O)O)O)O)C1CCC(C(C1)O)O

Physicochemical Properties	
Formula	C30H50O13
Molecular weight	618.71 g/mol
Num. heavy atoms	43
Num. arom. heavy atoms	0
Fraction Csp3	1.00
Num. rotatable bonds	3
Num. H-bond acceptors	13
Num. H-bond donors	11
Molar Refractivity	148.59
TPSA	240.99 Å²
Lipophilicity	
Log $P_{o/w}$ (iLOGP)	2.30
Log $P_{o/w}$ (XLOGP3)	-3.20
Log $P_{o/w}$ (WLOGP)	-3.25
Log $P_{o/w}$ (MLOGP)	-2.63
Log $P_{o/w}$ (SILICOS-IT)	-3.91
Consensus Log $P_{o/w}$	-2.14

Water Solubility	
Log *S* (ESOL)	-1.46
Solubility	2.14e+01 mg/ml ; 3.45e-02 mol/l
Class	Very soluble
Log *S* (Ali)	-1.29
Solubility	3.16e+01 mg/ml ; 5.11e-02 mol/l
Class	Very soluble
Log *S* (SILICOS-IT)	3.80
Solubility	3.93e+06 mg/ml ; 6.35e+03 mol/l
Class	Soluble
Pharmacokinetics	
GI absorption	Low
BBB permeant	No
P-gp substrate	No
CYP1A2 inhibitor	No
CYP2C19 inhibitor	No
CYP2C9 inhibitor	No
CYP2D6 inhibitor	No
CYP3A4 inhibitor	No
Log K_p (skin permeation)	-12.35 cm/s
Druglikeness	
Lipinski	No; 3 violations: MW>500, NorO>10, NHorOH>5
Ghose	No; 4 violations: MW>480, WLOGP<-0.4, MR>130, #atoms>70
Veber	No; 1 violation: TPSA>140
Egan	No; 1 violation: TPSA>131.6
Muegge	No; 5 violations: MW>600, XLOGP3<-2, TPSA>150, H-acc>10, H-don>5
Bioavailability Score	0.17
Medicinal Chemistry	
PAINS	0 alert
Brenk	0 alert
Leadlikeness	No; 1 violation: MW>350
Synthetic accessibility	7.31

Figure 4.3. SwissADME results for potential drug candidacy of flav100 amino acid with 7UJM receptor.

According to the figure 4.3, the molecular weight of the flav100 ligand is 618.71 g/mol. When water solubility was evaluated, flav100 was classified as very soluble according to Log S (ESOL), Log S +(Ali), Log S (SILICOS-IT) data. High solubility is a positive feature for bioavailability as it allows the drug to be absorbed more easily in the body. Low GI Absorption means that the molecule may have low absorption in the gastrointestinal tract. Log P values (ranging from -2.14 to -3.91) indicate that flav100 is not lipophilic. This supports that flav100 is not hydrophobic

and has high solubility. Also, flav100 has no PAINS warnings, indicating that the drug is unlikely to produce "false positive" effects.

As a result, although the flav100 ligand has some challenges with its high molecular weight, low GI absorption, and BBB passage, its high water solubility and positive results in the PAINS test indicate that flav100 may be a usable drug candidate under certain conditions.

According to the figure 4.4, ASP351 formed by carbon hydrogen bonding formed a weak interaction and made small contributions to the binding energy, but could increase the stability of the ligand. It ensured a strong binding by ensuring the correct positioning of the drug5 ligand in the active site of the 7UJM receptor by forming both Pi-Sigma and Pi-Sulfur interactions with MET343. Alkyl and pi-alkyl interactions occurred between amino acids LEU346, ILE424, LEU354, PRO535, LEU525, TRP383 and hydrophobic domains, contributing to the binding stability of the drug5 ligand to the 7UJM receptor.

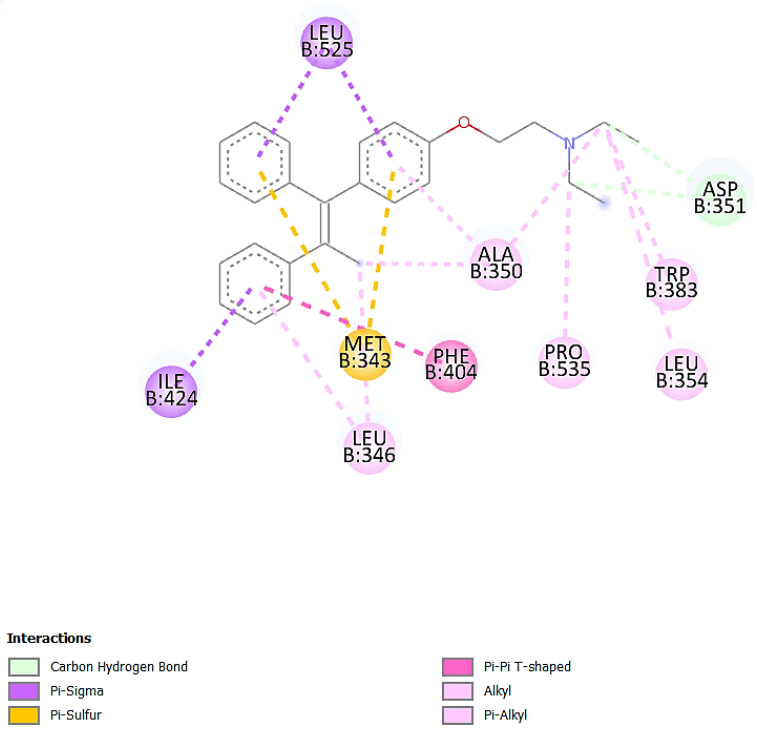


Figure 4.4. Discovery Studio 2D ligand-protein interaction diagram showing drug5 (conf76_drug5.pdbqt) amino acid interactions with the 7UJM receptor.

Additionally, Table 4.5 shows the flexible docking results of three different ligands for the 7UJM receptor. CHAINA_ligand82 has the lowest (strongest) binding energy of -11.4 kcal/mol. This suggests that CHAINA_ligand82 binds most strongly to the 7UJM receptor. Epicatechin_ligand100 and Drug_ligand5 have similar binding energies of -9.5 and -9.6 kcal/mol, respectively, and are slightly weaker bound than CHAINA_ligand82.

In conclusion, according to flexible docking results, the most effective ligand for the 7UJM receptor appears to be CHAINA_ligand82 because it has the lowest binding energy value. The other two ligands, Epicatechin_ligand100 and Drug_ligand5, have similar interaction potential with close binding energies.

Table 4.5. Flexible docking results of 7UJM with the best score after AutoDock Vina.

Receptor	Models	Docking Results (kcal/mol)
7UJM (ER)	CHAINA_ligand82	-11.4
	Epicathecin_ligand100	-9.5
	Drug_ligand5	-9.6

4.1.2. Molecular Docking Studies with 1R5K (ER), Ligands and Drugs

The 1R5K receptor, which is the estrogen receptor, is also a breast cancer receptor, like the 7UJM receptor. Protein-ligand models were created for the 1R5K receptor and flexible docking, blind and rigid docking results were recorded. According to RCSB data, 1R5K is provided together with ligands from GW5 molecule.

Suitable grid box dimensions and grid box coordinates for rigid docking studies are shown in Table 4.6 for the 1R5K receptor and docking parameters are shown in Table 4.7. After completing the necessary steps in the AutoDock program for the 7UJM receptor, it was run in **5 steps**, giving the population **size 50, 25 steps**, giving the population **size 150 and 100 steps**, giving the population **size 300.**

Table 4.6. Grid parameter file of 1R5K receptor and ligands with AutuDock Tools.

GRID PARAMETER FILE			
Macromolecule	**1R5K (ER)**		
Models	CHAINA_ligand29	Epicathecin_ligand93	Drug_ligand16
# of grind points in xyz	126 126 126	106 126 126	126 126 126
Spacing (Å)	0.375	0.375	0.375
xyz grid-center coordinates	-12.971	-13.316	-12.197
	45.484	46.541	48.213
	340.894	349.767	344.836
# of torsional degrees of freedom	5	5	5
Temperature (K)	298.15		
Charges	Kollman		

The docking parameters obtained accordingly are shown in Table 4.7 below.

Table 4.7. Docking parameter file of 1R5K receptor and ligands with AutuDock Tools.

DOCKING PARAMETERS	
Program	Autodock
Search Parameters	Lamarckian
	Genetic Algorithm
# of Run	5 / 25 / 100
# of Generations	27000
maximum # of energy evulations (short level)	250000
Population Size	50 / 150 / 300

Additionally, grid box dimensions and grid box coordinates for rigid and blind docking results are shown in Table 4.8 for the 1R5K receptor.

Table 4.8. Rigid and blind docking results of 1R5K receptor with the best score.

Models	# of Run/Population Size	Cluster Rank	Lowest Binding Energy [kcal/mol]	Number of Run	Number in Clusters	Inhibition Constant [μM]
CHAINA_ligand29	5/50	1	-8.29	1	1	0.8438
	25/150	1	-12.35	15	3	0.0009
	100/300	1	-12.42	31	14	0.0008
Epicathecin_ligand93	5/50	1	-5.61	2	1	76.78
	25/150	1	-6.23	22	1	27.17
	100/300	1	-8.32	28	2	0.7927
Drug_ligand16	5/50	1	-7.20	3	1	5.24
	25/150	1	-9.71	4	5	0.0766
	100/300	1	-9.94	70	19	0.0520

CHAINA_ligand29 and Drug_ligand16 shows lower binding energies and lower inhibition constants at larger population sizes, implying stronger binding and a more effective inhibition potential. Epicatechin_ligand93 showed the lowest binding energy (-8.32 kcal/mol) and lower inhibition constant (0.7927 μM), especially at the population size of 100/300, suggesting that there may be stronger binding and more effective inhibition potential under certain conditions. In addition, the clustering numbers in different conditions indicate that the ligands can bind to the target protein in different conformations.

As a result, according to the docking results, CHAINA_ligand29 generally showed the lowest binding energies and lowest inhibition constants, indicating that the CHAINA_ligand29 model may have a stronger binding and inhibition potential than other ligands.

Figure 4.5 shows the structure of amino acids in the conformation obtained after docking for the 1R5K receptor and ligand29.

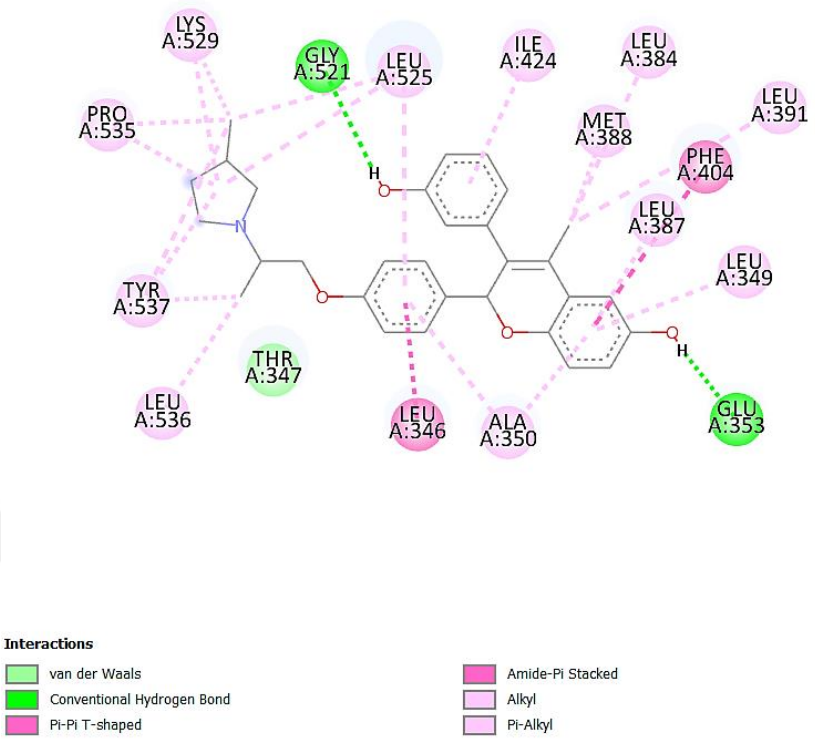


Figure 4.5. Discovery Studio 2D ligand-protein interaction diagram showing ligand29 (conf31_liga29.pdbqt) amino acid interactions with the 1R5K receptor.

According to the figure 4.5, GLY521 and GLU353 formed hydrogen bonds, providing a strong and specific interaction. LEU346, which shows amide-Pi stacking interactions occurred between aromatic rings and amide groups, enabling ligand29 to form a stable conformation. Pi-Alkyl interactions have been observed with LEU525 and ILE424, which contain alkyl groups. These interactions are weaker than hydrogen bonds but stabilized the position of the ligand29.

Figure 4.6 shows the structure of amino acids in the conformation obtained after docking for the 1R5K receptor and flav93.

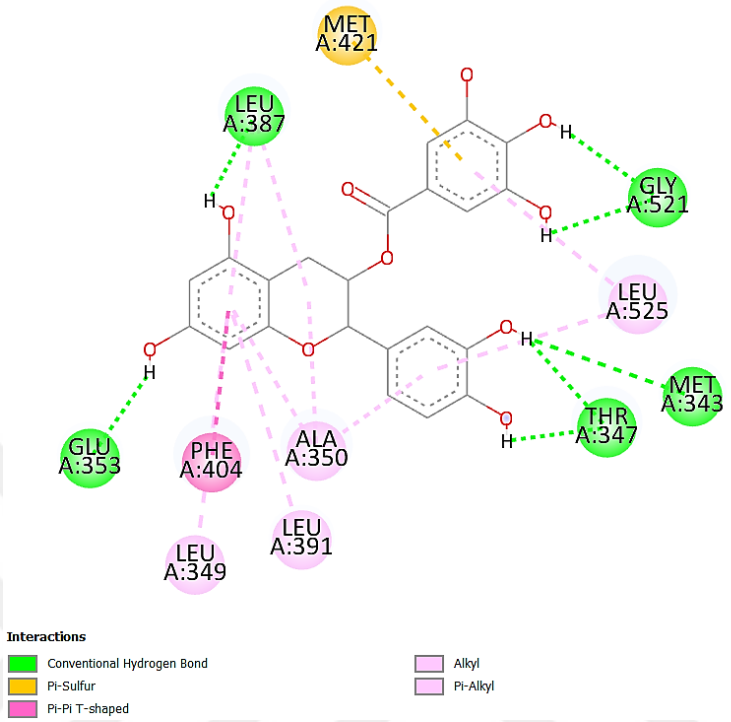
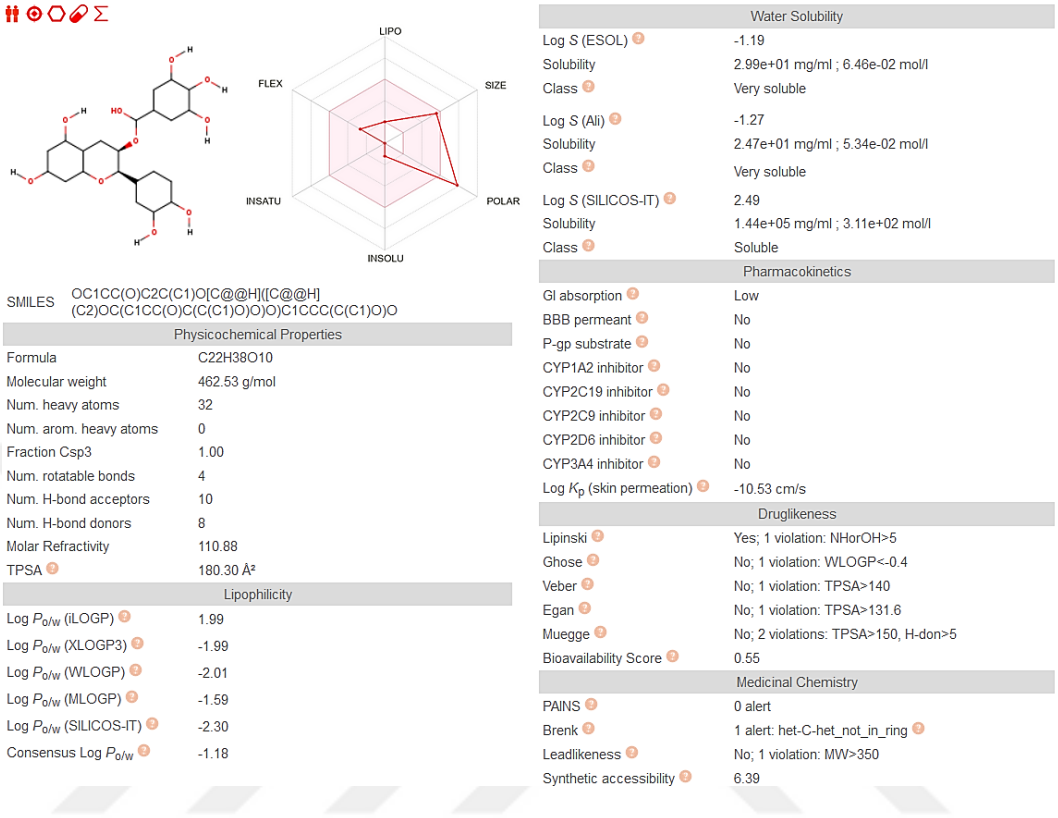


Figure 4.6. Discovery Studio 2D ligand-protein interaction diagram showing flav93 (conf28_flav93.pdbqt) amino acid interactions with the 1R5K receptor.

According to the figure 4.6, GLY521**,** LEU387**,** THR347**,** GLY521, and GLU353 formed hydrogen bonds, providing a strong and specific interaction. Pi-Sulfur interaction with the MET421 residue formed between the sulfur atom in the methionine residue and the aromatic ring of the flavonoid stabilized the binding conformation. Pi-Alkyl interactions have been observed with ALA350**,** LEU349**,** ve LEU391, which contain alkyl groups.

Figure 4.7 shows the SwissADME image obtained with the 28th conformation of the 1R5K receptor and the flav93 ligand after molecular docking.



SMILES OC1CC(O)C2C(C1)O[C@@H]([C@@H](C2)OC(C1CC(O)C(C(C1)O)O)O)C1CCC(C(C1)O)O

Physicochemical Properties	
Formula	C22H38O10
Molecular weight	462.53 g/mol
Num. heavy atoms	32
Num. arom. heavy atoms	0
Fraction Csp3	1.00
Num. rotatable bonds	4
Num. H-bond acceptors	10
Num. H-bond donors	8
Molar Refractivity	110.88
TPSA	180.30 Å²

Lipophilicity	
Log $P_{o/w}$ (iLOGP)	1.99
Log $P_{o/w}$ (XLOGP3)	-1.99
Log $P_{o/w}$ (WLOGP)	-2.01
Log $P_{o/w}$ (MLOGP)	-1.59
Log $P_{o/w}$ (SILICOS-IT)	-2.30
Consensus Log $P_{o/w}$	-1.18

Water Solubility	
Log *S* (ESOL)	-1.19
Solubility	2.99e+01 mg/ml ; 6.46e-02 mol/l
Class	Very soluble
Log *S* (Ali)	-1.27
Solubility	2.47e+01 mg/ml ; 5.34e-02 mol/l
Class	Very soluble
Log *S* (SILICOS-IT)	2.49
Solubility	1.44e+05 mg/ml ; 3.11e+02 mol/l
Class	Soluble

Pharmacokinetics	
GI absorption	Low
BBB permeant	No
P-gp substrate	No
CYP1A2 inhibitor	No
CYP2C19 inhibitor	No
CYP2C9 inhibitor	No
CYP2D6 inhibitor	No
CYP3A4 inhibitor	No
Log K_p (skin permeation)	-10.53 cm/s

Druglikeness	
Lipinski	Yes; 1 violation: NHorOH>5
Ghose	No; 1 violation: WLOGP<-0.4
Veber	No; 1 violation: TPSA>140
Egan	No; 1 violation: TPSA>131.6
Muegge	No; 2 violations: TPSA>150, H-don>5
Bioavailability Score	0.55

Medicinal Chemistry	
PAINS	0 alert
Brenk	1 alert: het-C-het_not_in_ring
Leadlikeness	No; 1 violation: MW>350
Synthetic accessibility	6.39

Figure 4.7. SwissADME results for potential drug candidacy of flav93 ligand with 1R5K receptor.

According to the figure 4.7, the molecular weight of the flav93 ligand is 462.53 g/mol. The values of Log S (ESOL): -1.19 and Log S (Ali): -1.27 indicate that flavonoid93 is highly soluble in water. This is a positive feature for flav93 in terms of its bioavailability and facilitates absorption and distribution in the body. Low GI Absorption means that the molecule may have low absorption in the gastrointestinal tract. Log P values (ranging from -1.99 to -2.01) shows that flavonoid93 is moderately lipophilic. This balances both the water solubility of the drug and its ability to pass through lipid membranes. Also, flav93 has no PAINS warnings, indicating that the drug is unlikely to produce "false positive" effects.

As a result, although flav93 presents some challenges in terms of its pharmacokinetic properties and drug similarity, the water solubility of the molecule

and low GI absorption indicate that flav93 can be made a more effective drug candidate with structural optimization and formulation development. In particular, increasing oral bioavailability and optimizing lipophilicity may increase its potential pharmacological activity.

Figure 4.8 shows the structure of amino acids in the conformation obtained after docking for the 1R5K receptor and drug16.

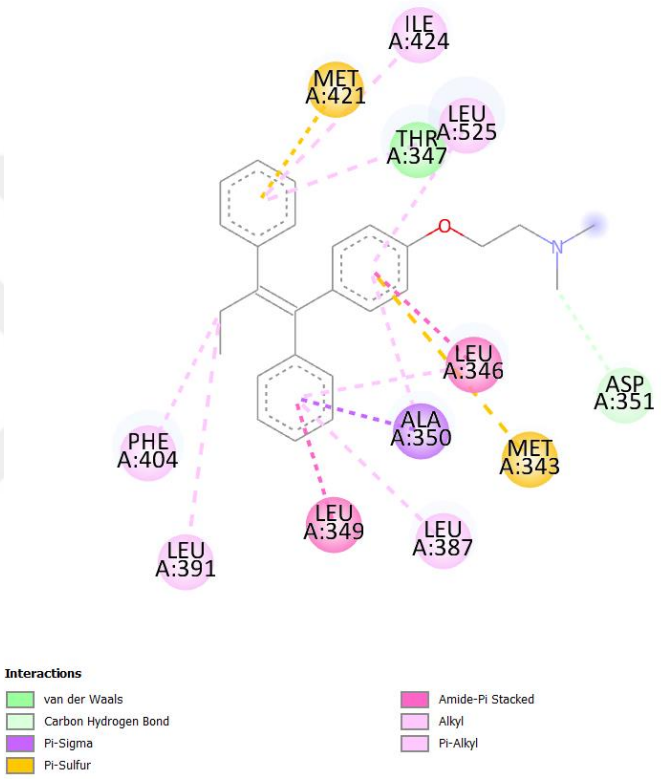


Figure 4.8. Discovery Studio 2D ligand-protein interaction diagram showing drug16 (conf70_drug16.pdbqt) amino acid interactions with the 1R5K receptor.

According to the Figure 4.8, Pi-Sulfur interaction with the MET421 and MET343 residue formed between the sulfur atom in the methionine residue and the aromatic ring of the drug stabilized the binding conformation. Pi-Alkyl interactions have been observed with, LEU349**,** LEU346**,** LEU387, LEU525 and ILE424 which contain alkyl groups.

Additionally, Table 4.9 shows the flexible docking results of three different ligands for the 1R5K receptor. CHAINA_ligand29 has the lowest (strongest) binding energy of -12.2 kcal/mol. This suggests that CHAINA_ligand29 binds most strongly to the 1R5K receptor. Epicatechin_ligand93 and Drug_ligand16 have similar binding energies of -10.6 and -9.4 kcal/mol, respectively, and are slightly weaker bound than CHAINA_ligand29.

In conclusion, according to flexible docking results, the most effective ligand for the 1R5K receptor appears to be CHAINA_ligand29 because it has the lowest binding energy value. The other two ligands, Epicatechin_ligand93 and Drug_ligand16, have similar interaction potential with close binding energies.

Table 4.9. Flexible docking results of 1R5K with the best score after AutoDock Vina.

Receptor	Models	Docking Results (kcal/mol)
1R5K (ER)	CHAINA_ligand29	-12.2
	Epicathecin_ligand93	-10.6
	Drug_ligand16	-9.4

4.1.3. Molecular Docking Studies with 4IFI (BRCA1), Ligands and Drugs

BRCA1 (Breast Cancer Type 1 susceptibility protein), which is especially involved in the repair of double-stranded DNA breaks through homologous recombination, is a tumor suppressor gene that plays an important role in DNA repair. Mutations in the BRCA1 gene, which prevents cancer development by preserving the genetic integrity of cells after the breaks are repaired, may cause the DNA repair mechanism to deteriorate and, as a result, increase the risk of cancer. Mutations in the BRCA1 gene increase the risk of breast cancer by disrupting cellular DNA repair mechanisms. BRCA1 mutations are an important factor in the development of breast and ovarian cancers in particular.

Also, the 4IFI structure is used to examine the structural and functional properties of BRCA1, leading to an understanding of which regions of the BRCA1 protein are critical and what effects mutations cause in these regions. This structure was used to identify potential therapeutic agents that can bind to the BRCA1 protein, especially in the research subject to this thesis, such as molecular docking and dynamic simulation studies. According to RCSB data, 4IFI is provided together with ligands from ACT molecule. While docking studies in which the structure of the ligands and receptors were flexible were carried out with AutoDock Vina, rigid docking and blind docking studies were carried out with AutoDock Tools. Suitable grid box dimensions and grid box coordinates for rigid docking studies are shown in Table 4.10 for the 4IFI receptor and docking parameters are shown in Table 4.11. After completing the necessary steps in the AutoDock program for the 4IFI receptor, it was run in 5 steps, giving the population size 50, 25 steps, giving the population size 150 and 100 steps, giving the population size 300.

Table 4.10. Grid parameter file of 4IFI receptor and ligands with AutuDock Tools.

GRID PARAMETER FILE		
Macromolecule	**4IFI (BRCA1)**	
Models	Epicathecin_ligand100	Drug_ligand17
# of grind points in xyz	126 126 126	126 126 126
Spacing (Å)	0.375	0.375
xyz grid-center coordinates	-18,058	-16,943
	16,178	19
	-11,954	-16,816
# of torsional degrees of freedom	5	5
Temperature (K)	298.15	
Charges	Kollman	

The docking parameters obtained accordingly are shown in Table 4.11 below.

Table 4.11. Docking parameter file of 4IFI receptor and ligands with AutuDock Tools.

DOCKING PARAMETERS	
Program	Autodock
Search Parameters	Lamarckian
	Genetic Algorithm
# of Run	5 / 25 / 100
# of Generations	27000
maximum # of energy evulations (short level)	250000
Population Size	50 / 150 / 300

Additionally, grid box dimensions and grid box coordinates for rigid and blind docking results are shown in Table 4.12 for the 4IFI receptor.

Table 4.12. Rigid and blind docking results of 4IFI receptor with the best score.

Models	# of Run/Population Size	Cluster Rank	Lowest Binding Energy [kcal/mol]	Number of Run	Number in Clusters	Inhibition Constant [μM]
Epicathecin_ligand100	5/50	1	-4.11	2	1	978.82
	25/150	1	-5.95	23	1	43.49
	100/300	1	-4.99	26	1	219.5000
Drug_ligand17	5/50	1	-7.68	2	1	2.34
	25/150	1	-8.0	4	1	1.3600
	100/300	1	-8.14	54	3	1.0800

Table 4.12 shows the results of three different ligands (Epicathecin_ligand100 and Drug_ligand17) for the molecular docking study with the 4IFI receptor. The lowest binding energy of Epicathecin_ligand100 was calculated as -5.95 kcal/mol (at a population size of 25/150). This value indicates that the ligand exhibits a relatively strong interaction with the 4IFI receptor. The lowest inhibition constant is 43.49 μM, indicating that this ligand has a moderate inhibitory effect. It had a population size of 25/150, was repeated 23 times and was located in 1 different cluster. This indicates that the binding sites of the ligand are relatively consistent. In addition, the lowest Drug_ligand17 binding energy was calculated as -8.14 kcal/mol (100/300 population size). This value indicates that the ligand exhibits a reasonable level of interaction with the 4IFI receptor. The lowest inhibition constant is 1.0800 μM, indicating that this

ligand has a low inhibitory effect. It had a population size of 100/300, was repeated 54 times and was located in 3 different clusters. This shows that the binding sites of the ligand are somewhat variable.

In conclusion, among the ligands that bind to the 4IFI receptor, Drug_ligand17 stands out as the most effective ligand with the lowest binding energy and lowest inhibition constant. Epicathecin_ligand100 can be considered relatively less effective ligands with higher binding energies and inhibition constants.

Figure 4.9 shows the structure of amino acids in the conformation obtained after docking for the 4IFI receptor and flav100.

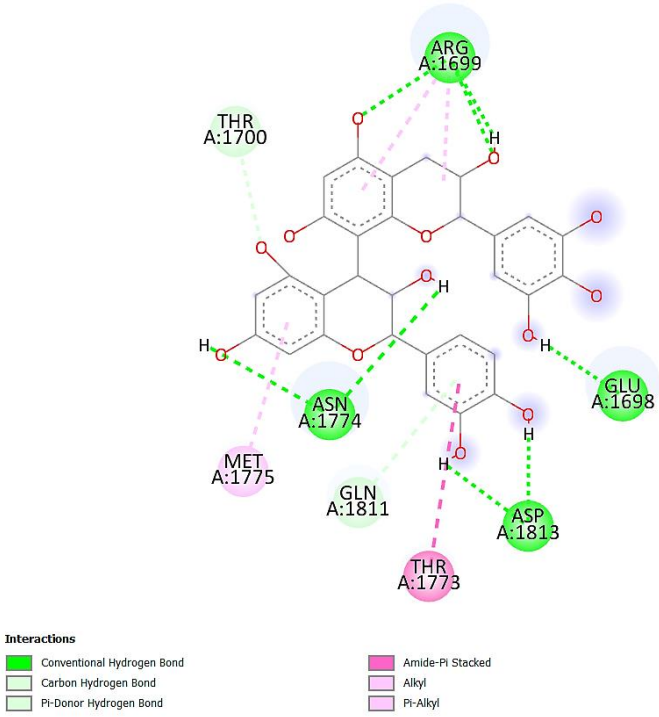
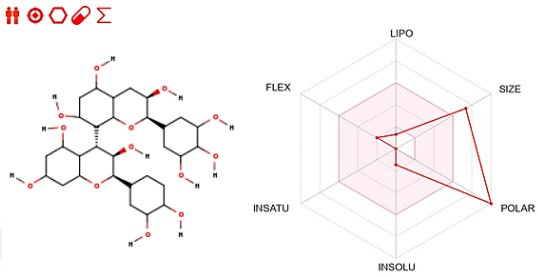


Figure 4.9. Discovery Studio 2D ligand-protein interaction diagram showing flav100 (conf23_flav100.pdbqt) amino acid interactions with the 4IFI receptor.

According to the Figure 4.9, hydrogen bonds were formed with amino acids ARG1699, ASN1774, GLU1698, ASP1813, GLN1811. Thus, flav100 is tightly bound to the 4IFI receptor**.** Pi-Alkyl interaction has been observed with, MET1775 which contain alkyl groups. The pi-donor hydrogen bond with THR1773 ensures the binding

of aromatic rings in this region, indicating that the aromatic part of the flav100 ligand binds more strongly to the 4IFI receptor and that this region plays a critical role in the placement of the flav100 ligand.

Figure 4.10 shows the SwissADME image obtained with the 23th conformation of the 4IFI receptor and the flav100 ligand after molecular docking.



SMILES OC1CC(O)C2C(C1)O[C@@H]([C@@H]([C@@H]2C1C(O)CC(C2C1O[C@@H]([C@@H](C2)O)C1CC(O)C(C(C1)O)O)O)O)C1CCC(C(C1)O)O

Physicochemical Properties	
Formula	C30H50O13
Molecular weight	618.71 g/mol
Num. heavy atoms	43
Num. arom. heavy atoms	0
Fraction Csp3	1.00
Num. rotatable bonds	3
Num. H-bond acceptors	13
Num. H-bond donors	11
Molar Refractivity	148.59
TPSA	240.99 Å²
Lipophilicity	
Log $P_{o/w}$ (iLOGP)	2.30
Log $P_{o/w}$ (XLOGP3)	-3.20
Log $P_{o/w}$ (WLOGP)	-3.25
Log $P_{o/w}$ (MLOGP)	-2.63
Log $P_{o/w}$ (SILICOS-IT)	-3.91
Consensus Log $P_{o/w}$	-2.14

Water Solubility	
Log *S* (ESOL)	-1.46
Solubility	2.14e+01 mg/ml ; 3.45e-02 mol/l
Class	Very soluble
Log *S* (Ali)	-1.29
Solubility	3.16e+01 mg/ml ; 5.11e-02 mol/l
Class	Very soluble
Log *S* (SILICOS-IT)	3.80
Solubility	3.93e+06 mg/ml ; 6.35e+03 mol/l
Class	Soluble
Pharmacokinetics	
GI absorption	Low
BBB permeant	No
P-gp substrate	No
CYP1A2 inhibitor	No
CYP2C19 inhibitor	No
CYP2C9 inhibitor	No
CYP2D6 inhibitor	No
CYP3A4 inhibitor	No
Log K_p (skin permeation)	-12.35 cm/s
Druglikeness	
Lipinski	No; 3 violations: MW>500, NorO>10, NHorOH>5
Ghose	No; 4 violations: MW>480, WLOGP<-0.4, MR>130, #atoms>70
Veber	No; 1 violation: TPSA>140
Egan	No; 1 violation: TPSA>131.6
Muegge	No; 5 violations: MW>600, XLOGP3<-2, TPSA>150, H-acc>10, H-don>5
Bioavailability Score	0.17
Medicinal Chemistry	
PAINS	0 alert
Brenk	0 alert
Leadlikeness	No; 1 violation: MW>350
Synthetic accessibility	7.31

Figure 4.10. SwissADME results for potential drug candidacy of flav100 ligand with 4IFI receptor.

According to the Figure 4.10, the molecular weight of the flav100 ligand is 618.71 g/mol. The values of Log S (ESOL): -1.46 and Log S (Ali): -1.29 indicate that flavonoid100 is highly soluble in water. This is a positive feature for flav100 in terms of its bioavailability and facilitates absorption and distribution in the body. Low GI Absorption means that the molecule may have low absorption in the gastrointestinal tract. The flav100 ligand has high solubility. This provides an advantage in terms of drug formulation. Good solubility in aqueous media generally ensures fast and

effective absorption of the drug, helping its effect to appear more quickly. Also, flav100 has no PAINS warnings, indicating that the drug is unlikely to produce "false positive" effects.

As a result, flav100 may cause some difficulties in its development as a drug when evaluated for its pharmacokinetic properties. However, high resolution may be advantageous for certain therapeutic applications. Improvements in the drug development process, especially in terms of formulation and administration routes, will facilitate its use.

Figure 4.11 shows the structure of amino acids in the conformation obtained after docking for the 4IFI receptor and drug17.

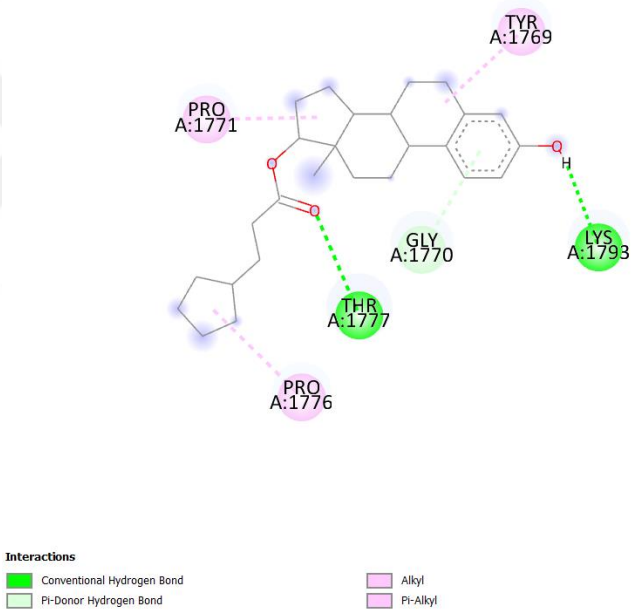


Figure 4.11. Discovery Studio 2D ligand-protein interaction diagram showing drug17 (conf54_drug17.pdbqt) amino acid interactions with the 4IFI receptor.

According to the Figure 4.11, hydrogen bonds were formed with amino acids THR1777 and LYS1793. Thus, drug17 is tightly bound to the 4IFI receptor. Pi-Alkyl and alkyl interaction has been observed with, PRO1771, PRO1776 and TYR1769 which contain alkyl groups. The pi-donor hydrogen bond with GLY1770 ensures the binding of aromatic rings in this region, indicating that the aromatic part of the drug17

ligand binds more strongly to the 4IFI receptor and that this region plays a critical role in the placement of the drug17 ligand.

Additionally, Table 4.13 shows the flexible docking results of three different ligands for the 4IFI receptor. Drug_ligand17 has the lowest (strongest) binding energy of -4.8 kcal/mol. This suggests that Drug_ligand17 binds most strongly to the 4IFI receptor. Epicatechin_ligand100 and CHAINA_ligand12 have similar binding energies of -4.5 and -4.4 kcal/mol, respectively, and are slightly weaker bound than Drug_ligand17.

In conclusion, according to flexible docking results, the most effective ligand for the 4IFI receptor appears to be Drug_ligand17 because it has the lowest binding energy value. The other two ligands, Epicatechin_ligand100 and CHAINA_ligand12, have similar interaction potential with close binding energies.

Table 4.13. Flexible docking results of 4IFI with the best score after AutoDock Vina.

Receptor	Models	Docking Results (kcal/mol)
4IFI (BRCA1)	CHAINA_ligand12	-4.4
	Epicathecin_ligand100	-4.5
	Drug_ligand17	-4.8

4.1.4. Molecular Docking Studies with 1N0W (BRCA2), Ligands and Drugs

BRCA2 (Breast Cancer Type 2 susceptibility protein), which is especially involved in the repair of double-stranded DNA breaks through homologous recombination, is a tumor suppressor gene that plays an important role in DNA repair. Mutations in the BRCA2 gene, which prevents cancer development by preserving the genetic integrity of cells after the breaks are repaired, may cause the DNA repair mechanism to deteriorate and, as a result, increase the risk of cancer. BRCA2 mutations are an important factor in the development of breast and ovarian cancers in particular.

Also, the 1N0W receptor, used to study the functional and structural properties of the BRCA2 protein, is a model used to observe how the gene interacts with DNA and is involved in the homologous recombination process. The 1N0W receptor directs DNA repair by interacting with the BRCA2 protein and other DNA repair proteins such as RAD51, and disruptions in the function of BRCA2 lead to the accumulation of DNA damages and destabilization of the cellular genome. This situation triggers tumor formation by gaining uncontrolled growth and division in cells. According to RCSB data, 1N0W is provided together with ligands from EDO, CL and MG molecules. While docking studies in which the structure of the ligands and receptors were flexible were carried out with AutoDock Vina, rigid docking and blind docking studies were carried out with AutoDock Tools. Suitable grid box dimensions and grid box coordinates for rigid docking studies are shown in Table 4.14 for the 1N0W receptor and docking parameters are shown in Table 4.15. After completing the necessary steps in the AutoDock program for the 1N0W receptor, it was run in 5 steps, giving the population size 50, 25 steps, giving the population size 150 and 100 steps, giving the population size 300.

Table 4.14. Grid parameter file of 1N0W receptor and ligands with AutuDock Tools.

GRID PARAMETER FILE			
Macromolecule	**1N0W (BRCA2)**		
Models	CHAINB_ligand70	Cathecin_ligand83	Drug_ligand17
# of grind points in xyz	126 126 126	126 126 126	126 124 126
Spacing (Å)	0.375	0.375	0.375
xyz grid-center coordinates	16.575	16.476	13.888
	14.426	19.543	21.509
	-0,129	-4,43	-3,466
# of torsional degrees of freedom	5	5	5
Temperature (K)	298.15		
Charges	Kollman		

The docking parameters obtained accordingly are shown in Table 4.15 below.

Table 4.15. Docking parameter file of 1N0W receptor and ligands with AutuDock Tools.

DOCKING PARAMETERS	
Program	Autodock
Search Parameters	Lamarckian
	Genetic Algorithm
# of Run	5 / 25 / 100
# of Generations	27000
maximum # of energy evulations (short level)	250000
Population Size	50 / 150 / 300

Additionally, grid box dimensions and grid box coordinates for rigid and blind docking results are shown in Table 4.16 for the 1N0W receptor.

Table 4.16. Rigid and blind docking results of 1N0W receptor with the best score.

Models	# of Run/Population Size	Cluster Rank	Lowest Binding Energy [kcal/mol]	Number of Run	Number in Clusters	Inhibition Constant [μM]
CHAINB_ligand70	5/50	1	-6.41	4	3	19.9000
	25/150	1	-6.63	5	3	13.8700
	100/300	1	-6.66	29	15	13.0400
Cathecin_ligand83	5/50	1	-6.46	3	1	18.55
	25/150	1	-4.96	16	5	229.58
	100/300	1	-6.00	12	4	39.8500
Drug_ligand17	5/50	1	-7.54	5	2	2.96
	25/150	1	-7.51	9	1	3.1100
	100/300	1	-7.38	76	9	3.9100

Table 4.16 shows the results of three different ligands (CHAINB_ligand70, Cathecin_ligand83 and Drug_ligand17) for the molecular docking study with the 1N0W receptor. The lowest binding energy of CHAINB_ligand70 was calculated as -6.66 kcal/mol (at a population size of 100/300). This value indicates that the ligand exhibits a relatively strong interaction with the 1N0W receptor. The lowest inhibition constant is 13.04 μM, indicating that this ligand has a moderate inhibitory effect. It had a population size of 100/300, was repeated 15 times and was located in 3 different clusters. This indicates that the binding sites of the ligand are relatively consistent. In

addition, the lowest Cathecin_ligand83 binding energy was calculated as -6.46 kcal/mol (5/50 population size). This value indicates that the ligand exhibits a reasonable level of interaction with the 1N0W receptor. The lowest inhibition constant is 18.55 μM, indicating that this ligand has a low inhibitory effect. It had a population size of 100/300, was repeated 6 times and was located in 5 different clusters. This shows that the binding sites of the ligand are somewhat variable. Also, the lowest binding energy for drug_ligand17 was calculated as -7.51 kcal/mol (25/150 population size). This value indicates that the ligand exhibits a very strong interaction with the 1N0W receptor. The lowest inhibition constant is 3.11 mM, indicating that this ligand has a high inhibitory effect. It had a population size of 100/300, was repeated 76 times and was located in 9 different clusters. This shows that the binding sites of the ligand are quite consistent.

Figure 4.12 shows the structure of amino acids in the conformation obtained after docking for the 1N0W receptor and ligand70.

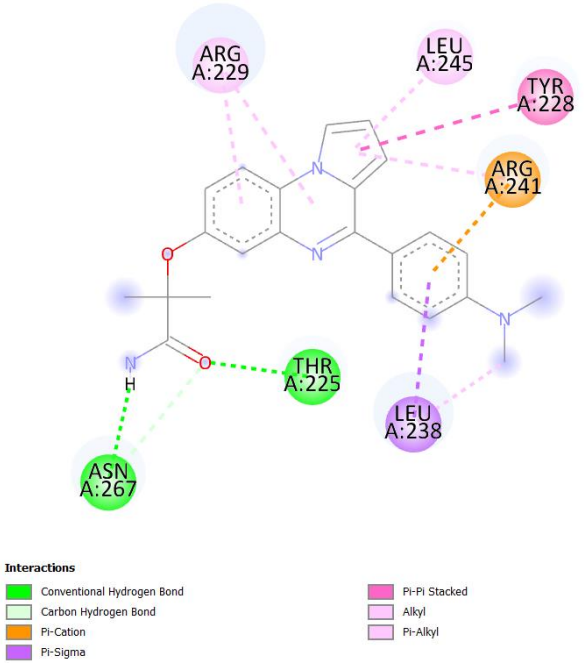


Figure 4.12. Discovery Studio 2D ligand-protein interaction diagram showing ligand70 (conf29_liga70.pdbqt) amino acid interactions with the 1N0W receptor.

According to the Figure 4.12, ASN267 and THR225 formed hydrogen bonds, providing a strong and specific interaction. Pi-Alkyl interactions have been observed with LEU245 and ARG229, which contain alkyl groups. These interactions are weaker than hydrogen bonds but stabilized the position of the ligand70. Pi-Interactions indicate that the aromatic rings of ligand70 in the protein are ARG229, TYR228, LEU245, ARG241 and LEU238 although their interactions with amino acids were not as strong as hydrogen bonds, they ensured the correct positioning of ligand70 and contributed to stability.

Figure 4.13 shows the structure of amino acids in the conformation obtained after docking for the 4IFI receptor and flav83.

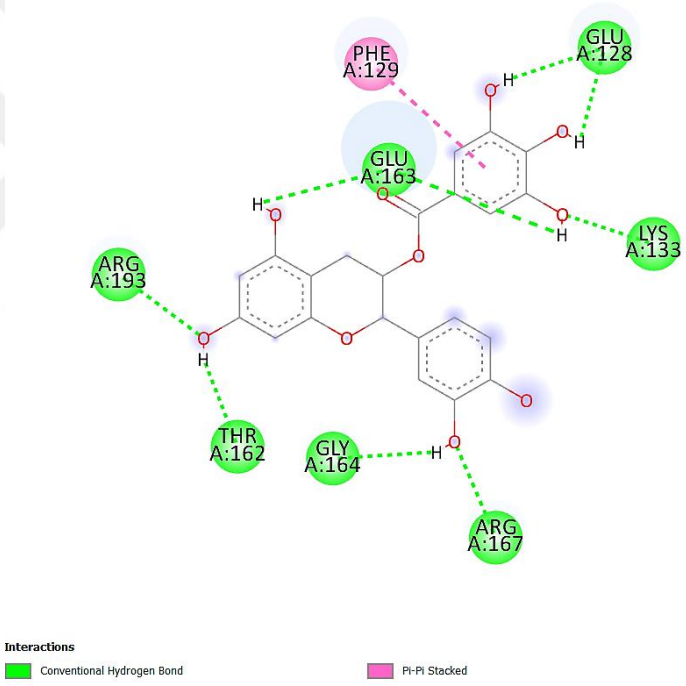
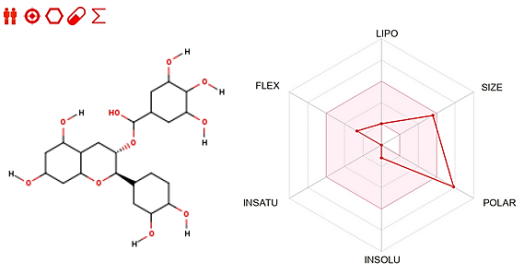


Figure 4.13. Discovery Studio 2D ligand-protein interaction diagram showing flav83 (conf3_flav83.pdbqt) amino acid interactions with the 1N0W receptor.

According to the Figure 2.27, The hydroxyl and carbonyl groups of flav83 are GLU128, GLU163, LYS133, ARG193, THR162, GLY164, ARG167 formed conventional hydrogen bonds. These bonds enabled strong and specific binding of flav83 to the 1N0W receptor, increasing the stability of flav83 at the binding site.

The aromatic ring of flav83 increased its binding affinity with the 4IFI receptor by making Pi-Pi stacked interaction with PHE129.

Figure 4.14 shows the SwissADME image obtained with the 3th conformation of the 1N0W receptor and the flav83 ligand after molecular docking.



SMILES	OC1CC(O)C2C(C1)O[C@@H]([C@H](C2)OC(C1CC(O)C(C(C1)O)O)O)C1CCC(C(C1)O)O

Physicochemical Properties	
Formula	C22H38O10
Molecular weight	462.53 g/mol
Num. heavy atoms	32
Num. arom. heavy atoms	0
Fraction Csp3	1.00
Num. rotatable bonds	4
Num. H-bond acceptors	10
Num. H-bond donors	8
Molar Refractivity	110.88
TPSA	180.30 Å²

Lipophilicity	
Log $P_{o/w}$ (iLOGP)	2.20
Log $P_{o/w}$ (XLOGP3)	-1.99
Log $P_{o/w}$ (WLOGP)	-2.01
Log $P_{o/w}$ (MLOGP)	-1.59
Log $P_{o/w}$ (SILICOS-IT)	-2.30
Consensus Log $P_{o/w}$	-1.14

Water Solubility	
Log *S* (ESOL)	-1.19
Solubility	2.99e+01 mg/ml ; 6.46e-02 mol/l
Class	Very soluble
Log *S* (Ali)	-1.27
Solubility	2.47e+01 mg/ml ; 5.34e-02 mol/l
Class	Very soluble
Log *S* (SILICOS-IT)	2.49
Solubility	1.44e+05 mg/ml ; 3.11e+02 mol/l
Class	Soluble

Pharmacokinetics	
GI absorption	Low
BBB permeant	No
P-gp substrate	No
CYP1A2 inhibitor	No
CYP2C19 inhibitor	No
CYP2C9 inhibitor	No
CYP2D6 inhibitor	No
CYP3A4 inhibitor	No
Log K_p (skin permeation)	-10.53 cm/s

Druglikeness	
Lipinski	Yes; 1 violation: NHorOH>5
Ghose	No; 1 violation: WLOGP<-0.4
Veber	No; 1 violation: TPSA>140
Egan	No; 1 violation: TPSA>131.6
Muegge	No; 2 violations: TPSA>150, H-don>5
Bioavailability Score	0.55

Medicinal Chemistry	
PAINS	0 alert
Brenk	1 alert: het-C-het_not_in_ring
Leadlikeness	No; 1 violation: MW>350
Synthetic accessibility	6.39

Figure 4.14. SwissADME results for potential drug candidacy of flav83 ligand with 1N0W receptor.

According to the Figure 4.14, the molecular weight of the flav83 ligand is 462.53 g/mol. The values of Log S (ESOL): -1.19 and Log S (Ali): -1.27 indicate that flavonoid100 is highly soluble in water. This is a positive feature for flav83 in terms of its bioavailability and facilitates absorption and distribution in the body. Low GI Absorption means that the molecule may have low absorption in the gastrointestinal tract. Also, flav100 has no PAINS warnings, indicating that the drug is unlikely to produce "false positive" effects.

As a result, although flav83 presents some challenges in terms of its pharmacokinetic properties and drug similarity, the water solubility of the molecule and low GI absorption indicate that flav83 can be made a more effective drug candidate with structural optimization and formulation development. In particular, increasing oral bioavailability and optimizing lipophilicity may increase its potential pharmacological activity.

Figure 4.15 shows the structure of amino acids in the conformation obtained after docking for the 1N0W receptor and drug17.

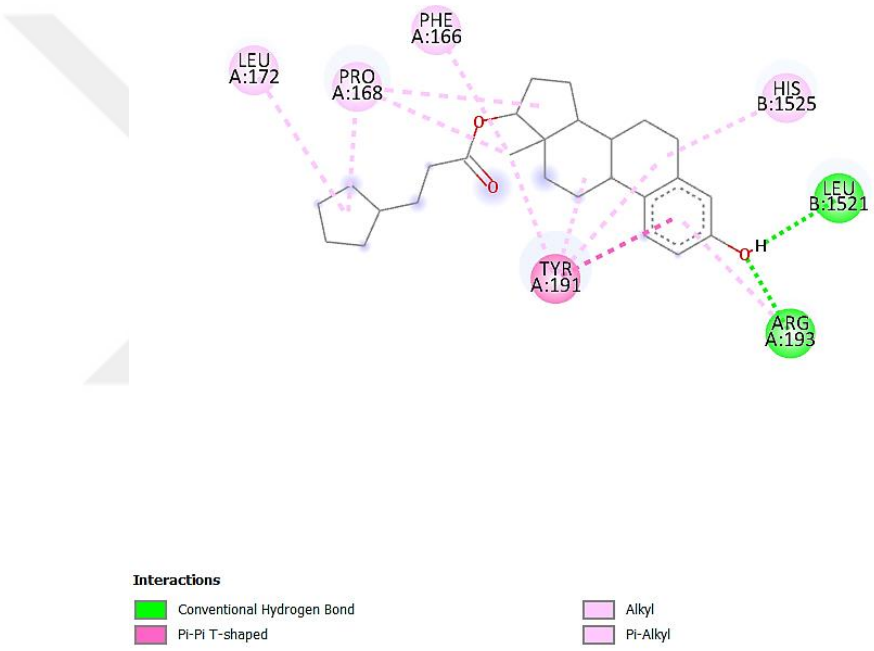


Figure 4.15. Discovery Studio 2D ligand-protein interaction diagram showing drug17 (conf5_drug17.pdbqt) amino acid interactions with the 1N0W receptor.

According to the Figure 2.29, conventional hydrogen bonds were formed with amino acids LEU1521 and ARG193. Thus, drug17 is tightly bound to the 1N0W **receptor.** Pi-Alkyl and alkyl interaction has been observed with, LEU172, PRO168, PHE166 and HIS1525 which contain alkyl groups. Pi-Pi T-shaped interactions were observed with TYR191 residue. This interaction between the aromatic rings helped the drug17 ligand to remain in a stable position**.**

In conclusion, among the ligands that bind to the 1N0W receptor, Drug_ligand17 stands out as the most effective ligand with the lowest binding energy and lowest inhibition constant. CHAINB_ligand70 and Cathecin_ligand83 can be considered relatively less effective ligands with higher binding energies and inhibition constants.

Table 4.17. Flexible docking results of 1N0W with the best score after AutoDock Vina.

Receptor	Models	Docking Results (kcal/mol)
1N0W (BRCA2)	CHAINB_ligand70	-7.6
	Cathecin_ligand83	-8.2
	Drug_ligand17	-8.5

4.2. Molecular Dynamics Simulations Results (Stability of Ligand Receptor Complexes)

Molecular dynamics studies were used to observe the dynamic shape changes occurring in the structure in order to understand how the ligand changes conformation relative to the protein by determining the kinetic energy change in order to calculate the average potential energy change and then the force values acting on each atom according to the energy values. An environment closer to biological systems has been achieved by modeling the movement of protein-ligand pairs in water molecules over time with molecular dynamics simulations. Thus, it has been observed that efficient dynamic results can be achieved if the necessary parameters are added to polarizable force fields in modeling the dynamic structures of disordered and rapidly folding proteins. Additionally, the parameters used in all MD simulations are shown in Table 4.18.

Table 4.18. MD simulations parameters.

Temperature	300 K
Force Field	charmm36
Solvent	TIP3P
Cubic Box Boundary	2.0 nm

MD simulations are performed by providing thermodynamic conditions has been carried out after NVT balancing with 100 ps in order to adjust the thermostat and balance the system depending on the constant particle number, constant volume,

constant temperature, and NPT balancing with 100 ps in order to select the barostat and control the density based on the constant particle number, constant pressure, constant temperature,

The protocol for the codes used to perform MD simulations is shown in APPENDEX A. Also, for the simulations performed with these codes used for each protein-ligand complex, RMSD (root mean square deviation), RMSF (root meand square fluctuations), hydrogen bond analysis, interaction energy of protein-ligand (IE), SASA (solvent accessible surface area), radial distribution functions (RDF) and MSD (mean square displacement) analyzes were performed and 50 ns simulations were recorded for each ligand and drug with UHeM (The National Center for High Performance Computing).

4.2.1. Molecular Dynamics Simulations with 7UJM (ER), Ligands and Drugs

First, the RMSD value, which stands for root mean square deviation, was plotted to measure the conformational changes of CHAINA_ligand82, Epicathecin_ligand100 and the Drug_ligand5 depending on the 7UJM (ER) receptor throughout the simulation, to understand how it changes and stabilizes relative to the initial structure.

Additionally, Figure 4.16 shows the time-dependent root-mean-square deviation (RMSD) during molecular dynamics (MD) simulations of three different ligands bound to the 7UJM (ER) receptor. Lower RMSD values indicate that the system is more stable, while higher RMSD values indicate more mobility and possible structural changes. The RMSD values for CHAINA_ligand82 remained constant between approximately 0.1-0.2 nm, and this ligand with the lowest RMSD values exhibited a very stable structure throughout the simulation. This shows that CHAINA_ligand82 binds tightly to the 7UJM (ER) receptor and maintains its conformation throughout the simulation. RMSD values for Epicathecin_ligand100 were initially between 0.2-0.3 nm and stabilized around 0.3-0.4 nm over time. Epicatechin_ligand100 had lower RMSD values initially but showed slightly greater mobility over time. This shows that there is a certain flexibility and mobility in the conformation of the flavonoid, but it still exhibits a relatively stable structure. The

RMSD value for Drug_ligand5 was initially between 0.2-0.3 nm and fluctuated between 0.3-0.5 nm over time. Drug_ligand5 had higher RMSD values than the other two ligands and showed greater mobility over time. This suggests that the drug undergoes more conformational changes over the course of the simulation and its binding stability with the receptor may be lower than other ligands.

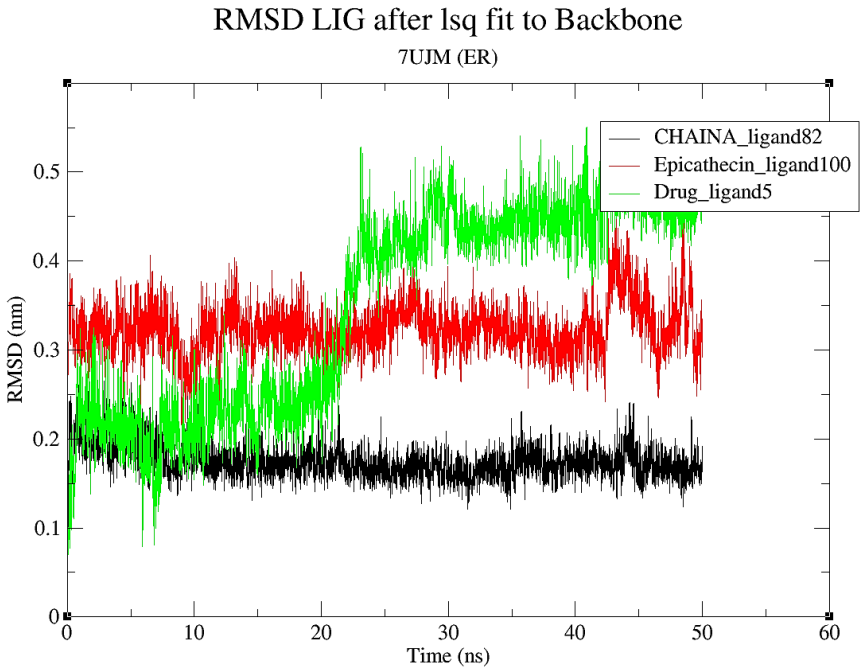


Figure 4.16. RMSD analysis results graph of 7UJM receptor-bound ligands in MD simulation with apo state.

As a result, CHAINA_ligand82 shows the most stable and tight binding to the 7UJM (ER) receptor. Epicatechin_ligand100 shows moderately stable binding, while Drug_ligand5 shows the least stable binding state.

RMSF provided information about the elasticity and dynamic behavior of specific regions by measuring how much each atom deviated from its average position throughout the simulation.

Additionally, Figure 4.17 shows the atomic root-mean-square deviations (RMSF) of ligands and drug bound to the 7UJM (ER) receptor. RMSF values generally remain below 0.2 nm for CHAINA_ligand82 and do not exceed 0.5 nm.

CHAINA_ligand82 has low RMSF values, indicating that it is quite stable and does not exhibit much flexibility when bound to the receptor. RMSF values for Epicathecin_ligand100 are generally between 0.2-0.4 nm, approaching 0.5 nm in certain ranges. Epicatechin_ligand100 has slightly higher RMSF values than CHAINA_ligand82, indicating greater flexibility at certain atoms. RMSF values for Drug_ligand5 are between 0.2-0.4 nm and exceed 0.5 nm at certain atoms. Drug_ligand5 has the highest RMSF values and shows greater flexibility and mobility in certain regions.

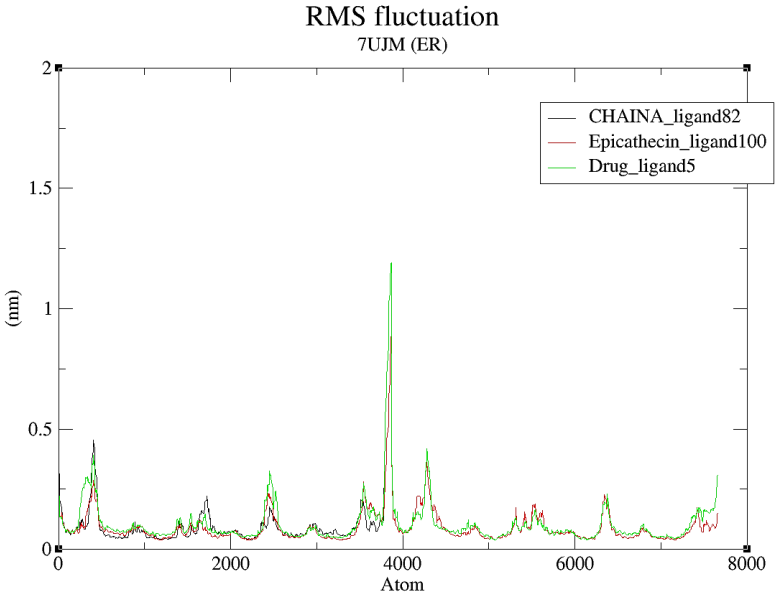


Figure 4.17. RMSF analysis results graph of 7UJM receptor-bound ligands in MD simulation with apo state.

As a result, RMSF peaks are observed in similar regions for all three ligands. These regions indicate more flexible or mobile regions of the 7UJM receptor. Additionally, CHAINA_ligand82 has the lowest RMSF values, indicating that this ligand shows the most stable binding and causes the least conformational change in the 7UJM receptor. Epicatechin_ligand100 and Drug_ligand5 show greater flexibility with higher RMSF values. Results based on the RMSF plot are consistent with the RMSD analyses, showing that CHAINA_ligand82 binds most stably to the 7UJM (ER) receptor, Epicatechin_ligand100 is moderately stable, and Drug_ligand5 is the least stable.

Radius of gyration, R_g values during the simulation are shown in figure 4.18 The radius ratio provides information about the compactness and stability of the protein-ligand complex. The lower R_g value of CHAINA_ligand82 indicates that this complex has a more compact structure than the others. The similar R_g values of Epicathecin_ligand100 and Drug_ligand5 indicate that these two complexes have a structurally similar compactness. CHAINA_ligand82 has a more compact structure than the other two ligand complexes and remained stable throughout the simulation. Epicathecin_ligand100 and Drug_ligand5 have similar structural compactness and remained stable throughout the simulation.

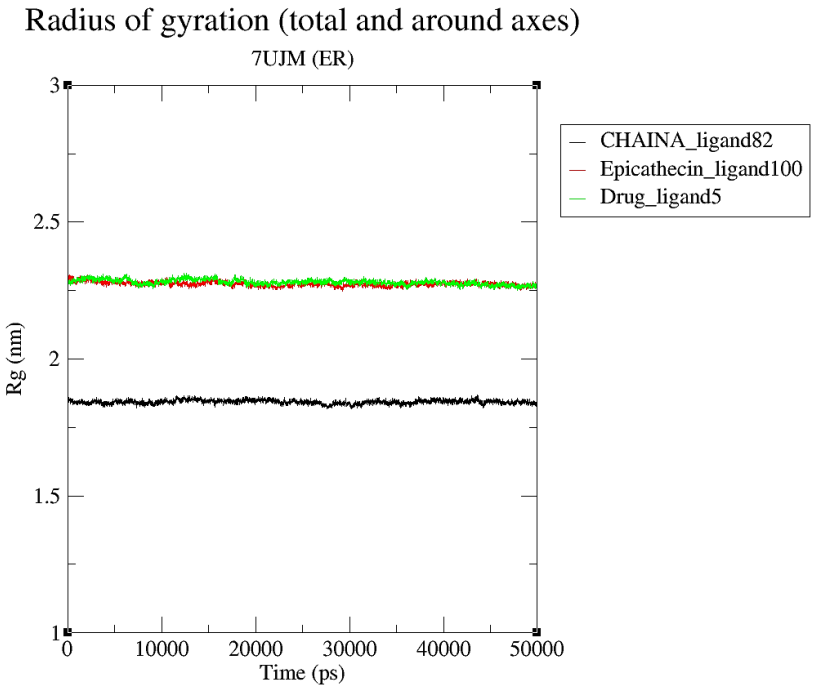


Figure 4.18. R_g analysis results graph of 7UJM receptor-bound ligands in MD simulation with apo state.

As a result, according to these results, the R_g values of the complexes remained stable over time and did not show large fluctuations. This suggests that the complexes are stable throughout the simulation period and do not experience large conformational changes.

Hydrogen bonding is used to understand the stability, structural integrity and interaction strength of protein-ligand complexes by providing strong non-covalent

interactions between protein and ligand. The high number and duration (persistence) of hydrogen bonds indicate that the complex is stable and the ligand has a strong interaction with the protein. However, the formation and breaking of hydrogen bonds are important for understanding the dynamic behavior of the ligand-protein complex. The constantly forming and breaking hydrogen bonds show that the complex has a dynamic and flexible structure.

Figure 4.19 shows that hydrogen bonds generally change between 1 and 4 over time. This shows that the binding of the CHAINA_ligand82 to the 7UJM receptor is stable and hydrogen bonds are constantly formed between 0-50 ns. As a result, the number of hydrogen bonds generally varies between 1 and 4, indicating strong and persistent binding of the CHAINA_ligand82 to the 7UJM receptor.

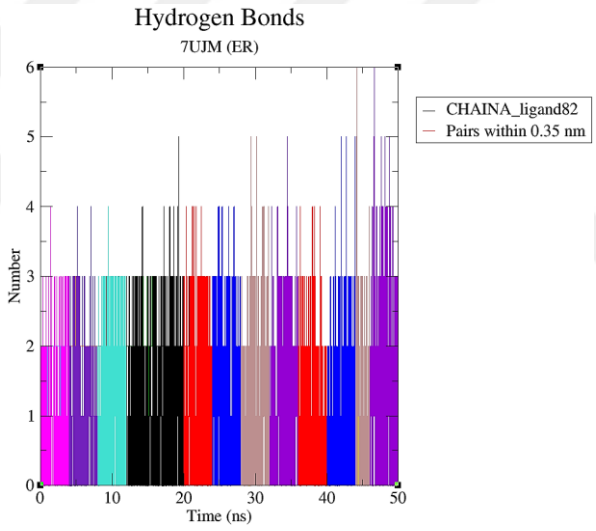


Figure 4.19. Hydrogen bonding analysis results graph of 7UJM receptor-bound CHAINA_ligand82 in MD simulation.

Figure 4.20 shows that hydrogen bonds generally vary from 5 to 15 over time and Figure 4.21 shows that hydrogen bonds generally vary from 1 to 3 over time. Considering the mobility and dynamic structure of the ligand in the binding region, it can be seen that hydrogen bonds are constantly formed and lost between 0-50 ns. As a result, the number of hydrogen bonds generally varies between 5 and 15 for Epicathecin_ligand100 and between 1 and 3 for Drug_ligand5 indicating strong and

persistent binding of the Epicathecin_ligand100 and Drug_ligand5 to the 7UJM receptor.

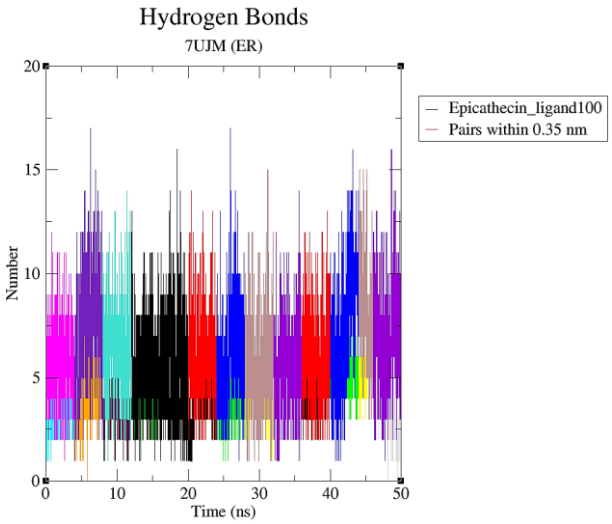


Figure 4.20. Hydrogen bonding analysis results graph of 7UJM receptor-bound Epicathecin_ligand100 in MD simulation.

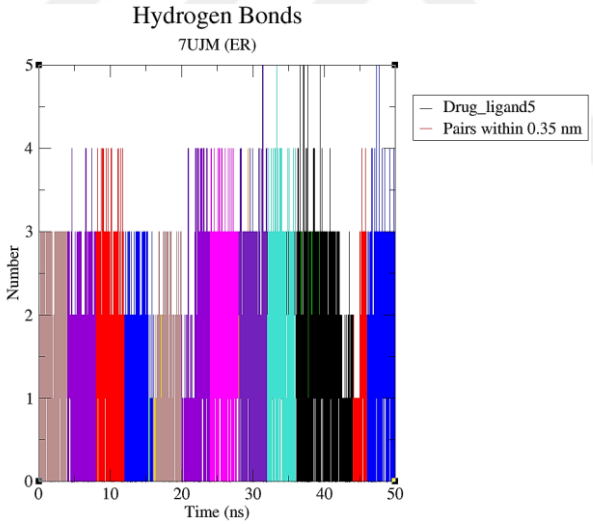


Figure 4.21. Hydrogen bonding analysis results graph of 7UJM receptor-bound Drug_ligand5 in MD simulation.

Protein-ligand interaction energy (IE) provides detailed information about ligand binding stability, dynamics and binding affinity in molecular dynamics simulations. Table 4.19 shows the protein-ligand interaction energies between the 7UJM (ER) receptor and CHAINA_ligand82, Epicathecin_ligand100 and Drug_ligand5. These values include the average values of Coulomb (Coul-SR) and Lenard-Jones (LJ-SR) interaction energies.

Table 4.19. Protein-ligand interaction energy results of 7UJM receptor-bound ligands in MD simulation.

MODELS	ENERGY	AMOUNT (Average)
CHAINA_ligand82	Coul-SR: Protein-LIG	-106.702
	LJ-SR: Protein-LIG	-235.711
Epicathecin_ligand100	Coul-SR: Protein-LIG	-152.996
	LJ-SR: Protein-LIG	-120.876
Drug_ligand5	Coul-SR: Protein-LIG	-94.151
	LJ-SR: Protein-LIG	-167.878

Additionally, according to Table 4.19, the strongest binding in terms of coulomb interactions was provided by Epicathecin_ligand100 (-152.996). This ligand shows the strongest electrostatic interactions with the protein. Also, CHAINA_ligand82 shows the strongest Van der Waals interactions with the lowest LJ-SR energy (-235.711). As a result, Epicathecin_ligand100 shows the strongest binding in terms of Coulomb interactions, and CHAINA_ligand82 shows the strongest binding in terms of Van der Waals interactions. Drug_ligand5 does not bind as strongly as the other two ligands in terms of either type of energy.

RDF (Radial Distribution Functions), which shows radial distribution functions and is expressed as g (r) , shows the density of other atoms around a reference atom, depending on the distance.

Additionally, the height of the RDF shows the density of other atoms around the reference atom, and this height indicates that the atoms are more clustered at this distance. The fact that RDF approaches zero over time with distance shows that the atoms are randomly distributed at a certain distance and that there is no regular structure at long distances.

According to Figure 4.22, the highest peak and wide distribution for CHAINA_ligand82 indicate that this ligand has a strong and broad interaction with the reference atom, meaning that the ligand is frequently found in regions close to the reference atom and concentrated there. Epicathecin_ligand100 has an intermediate peak and a narrower distribution and is concentrated at a distance from the reference atom, but is not as widely distributed as CHAINA_ligand82. Drug_ligand5 is in lower

density and narrower distribution, indicating that it has weaker and less diffuse interactions with the reference atom.

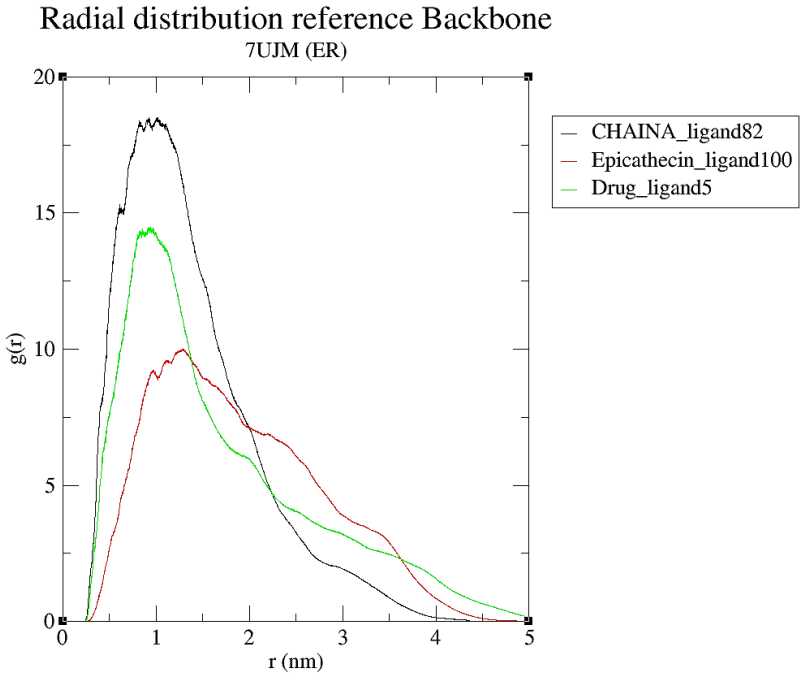


Figure 4.22. RDF (Radial Distribution Functions) analysis results graph of 7UJM receptor-bound ligands in MD simulation with apo state.

As a result, among the CHAINA_ligand82, Epicathecin_ligand100, Drug_ligand5 ligands, the ones that bind most strongly to the reference atom are CHAINA_ligand82, Epicathecin_ligand100, Drug_ligand5 respectively.

4.2.2. Molecular Dynamics Simulations with 1R5K (ER), Ligands and Drugs

First, the RMSD value, which stands for root mean square deviation, was plotted to measure the conformational changes of CHAINA_ligand29, Epicathecin_ligand93 and the Drug_ligand16 depending on the 1R5K (ER) receptor throughout the simulation, to understand how it changes and stabilizes relative to the initial structure.

Additionally, Figure 4.23 shows the time-dependent root-mean-square deviation (RMSD) during molecular dynamics (MD) simulations of three different

ligands bound to the 1R5K (ER) receptor. The RMSD value of CHAINA_ligand29 remained nearly constant over time, undergoing very little structural change and remaining fairly stable. The RMSD values of epicathecin_ligan93 fluctuated from 0 ns to 50 ns, varying between 10 nm and 25 nm, and it underwent significant structural changes throughout the simulation and did not remain stable. Drug16 remained very low in terms of RMSD values, showing values close to zero, and remained stable with the ER.

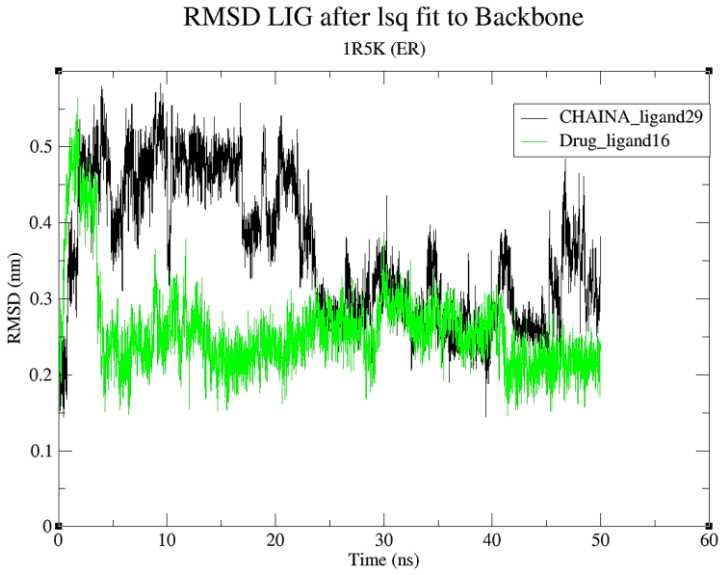


Figure 4.23. RMSD analysis results graph of 1R5K receptor-bound ligands in MD simulation with apo state.

RMSF analysis, Figure 4.24 shows the atomic root-mean-square deviations (RMSF) of ligands and drug bound to the 1R5K (ER). RMSF values of CHAINA_ligand29 are generally low and consistent, although slight increases are seen at certain atoms. While Epicatechin_ligand93 and Drug_ligand16 show higher RMSF values in some regions, CHAINA_ligand29 generally exhibits a more stable structure with lower RMSF values. High RMSF values indicate potentially more mobile or flexible regions.

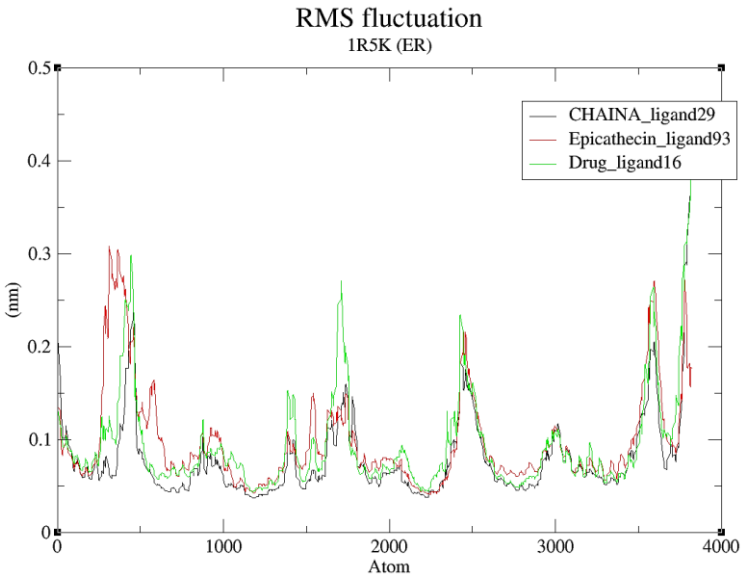


Figure 4.24. RMSF analysis results graph of 1R5K receptor-bound ligands in MD simulation with apo state.

The Rg value of CHAINA_ligand29 has remained fairly constant over time and is around 1.8 nm. The Rg value of epicathecin_ligand93 was similarly fairly constant over time, but fluctuated slightly compared to CHAINA_ligand29. The Rg value of drug_ligand16 remains quite constant and has a value very close to the other CHANA_ligand29 and Epicathecin_ligand93.

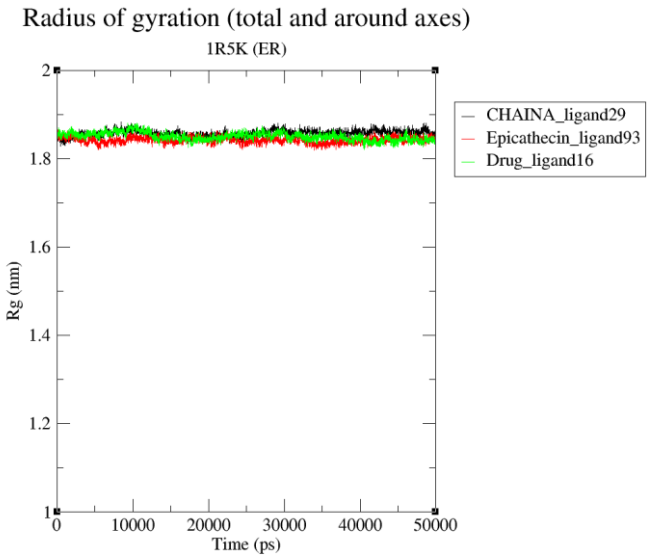


Figure 4.25. R_g analysis results graph of 1R5K receptor-bound ligands in MD simulation with apo state.

The gyration radius of CHAINA_ligand29, Epicathecin_ligand93, and Drug_ligand16 remained quite stable over time, indicating that the molecules largely maintained their compact structure throughout the simulation. Ligands showing an Rg value around 1.8 nm have a similar level of compactness and did not shrink significantly throug-hout the simulation.

Figure 4.26 shows that hydrogen bonds generally change between 1 and 5 over time. This shows that the binding of the CHAINA_ligand29 to the 1R5K receptor is stable and hydrogen bonds are constantly formed between 0-50 ns. As a result, the number of hydrogen bonds generally varies between 1 and 5, indicating strong and persistent binding of the CHAINA_ligand29 to the 1R5K receptor. The maximum number of hydrogen bonds is around 5-6. This represents specific moments when CHAINA_ligand29 binds across multiple nodes.

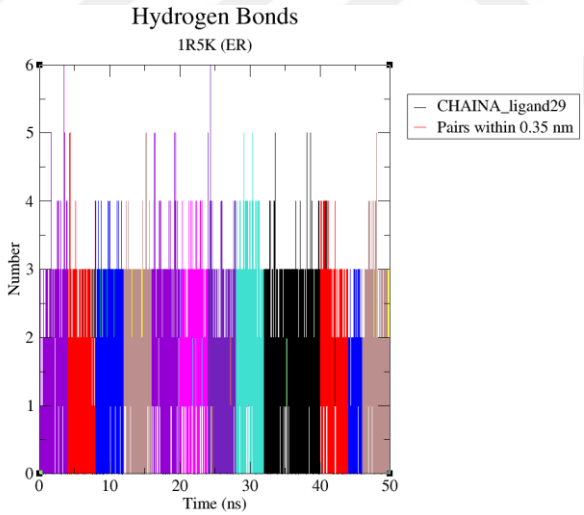


Figure 4.26. Hydrogen bonding analysis results graph of 1R5K receptor-bound CHAINA_ligand29 in MD simulation.

Figure 4.27 shows that hydrogen bonds in the graph of Drug_ligand16, the number of hydrogen bonds was observed to change over time, but this number never increased above 2. This suggests that drug_ligand16 generally forms a single hydrogen bond.

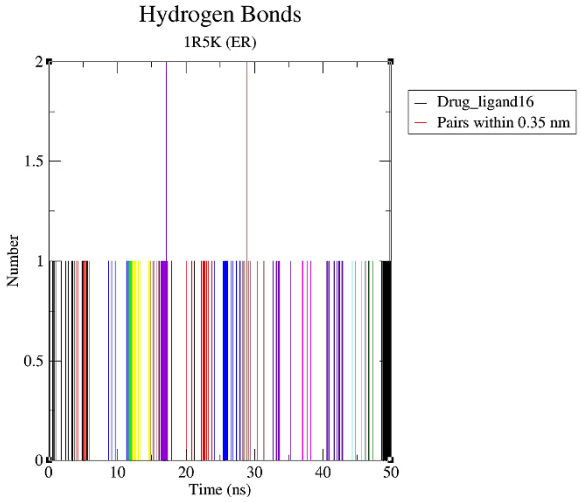


Figure 4.27. Hydrogen bonding analysis results graph of 1R5K receptor-bound Drug_ligand16 in MD simulation.

Table 4.20 shows the protein-ligand interaction energies between the 1R5K (ER) and CHAINA_ligand29, Epicathecin_ligand93 and Drug_ligand16. These values include the average values of Coulomb (Coul-SR) and Lenard-Jones (LJ-SR) interaction energies.

Table 4.20. Protein-ligand interaction energy results of 1R5K receptor-bound ligands in MD simulation.

MODELS	ENERGY	AMOUNT (Average)
CHAINA_ligand29	Coul-SR: Protein-LIG	-90.5774
	LJ-SR: Protein-LIG	-199.833
Epicathecin_ligand93	Coul-SR: Protein-LIG	0
	LJ-SR: Protein-LIG	0
Drug_ligand16	Coul-SR: Protein-LIG	-6.45903
	LJ-SR: Protein-LIG	-187.098

According to Table 4.20, CHAINA_ligand29 has the lowest Coulomb (-90.5774) and Lennard-Jones energy (-199.833) values with the 1R5K receptor. CHAINA_ligand29 bound strongly to the 1R5K receptor via both electrostatic and van der Waals interactions. This demonstrated that CHAINA_ligand29 has high binding affinity and therefore may be effective as a potential inhibitor. **The molecular dynamics simulation will be repeated for Epicatechin_ligand93.** Drug_ligand16 showed very strong binding in terms of Van der Waals interactions, but weak

electrostatic interactions. In conclusion, CHAINA_ligand29 showed the strongest binding and therefore may be an ideal candidate for further studies. Drug_ligand16 may be effective in certain situations, but it needs to be optimized for electrostatic interactions.

RDF (Radial Distribution Functions), according to Figure 4.28, the main peak for CHAINA_ligand29 and Drug_ligand16 was located approximately between 0.5-1.0 nm for the RDF value. It was observed that drug_ligand16 had a higher atomic density and showed a slightly stronger and more regular interaction than CHAINA_ligand29.

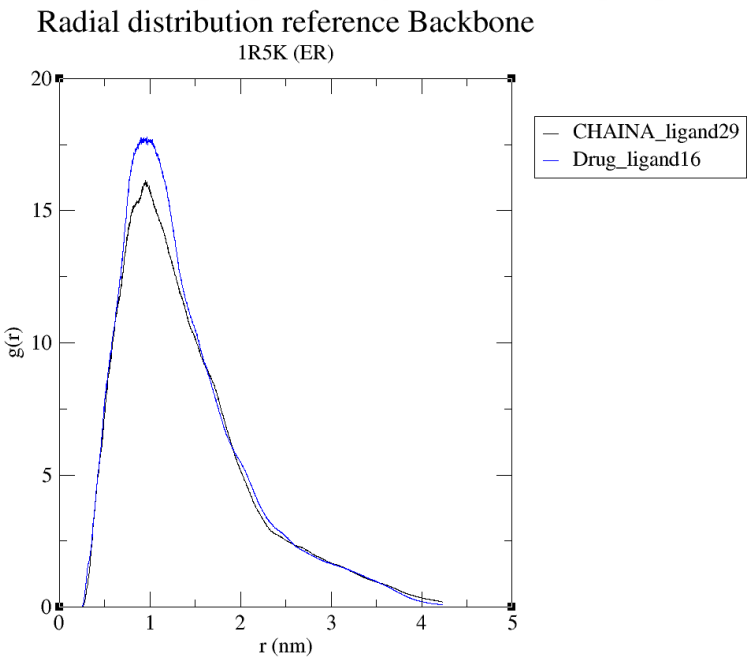


Figure 4.28. RDF (Radial Distribution Functions) analysis results graph of 1R5K receptor-bound ligands in MD simulation with apo state.

4.2.3. Molecular Dynamics Simulations with 4IFI (BRCA1), Ligands and Drugs

First, the RMSD value, which stands for root mean square deviation, was plotted to measure the conformational changes of Epicathecin_ligand10 and the Drug_ligand17 depending on the 4IFI (BRCA1) receptor throughout the simulation, to understand how it changes and stabilizes relative to the initial structure.

Additionally, Figure 4.29 shows the time-dependent root-mean-square deviation (RMSD) during molecular dynamics (MD) simulations of three different ligands bound to the 4IFI (BRCA1) receptor. Epicathecin_ligand100 showed a very low RMSD value throughout the simulation period (around 0.5-1.0 nm) and did not undergo major structural changes, binding quite stably to the 4IFI receptor. The RMSD value of drug_ligand17, although low at the beginning of the simulation, showed a sudden increase at approximately 10 ns and stabilized in the 2-3 nm range. This indicated that Drug_ligand17 experienced a significant change in its binding position or structure and subsequently formed a new stable structure.

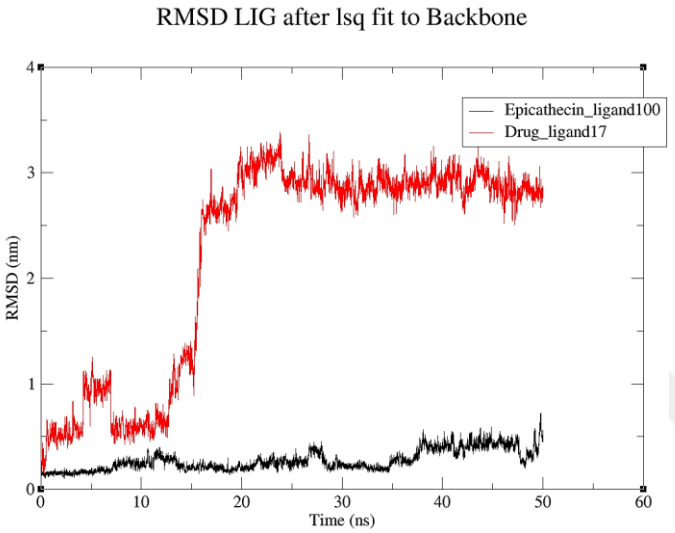


Figure 4.29. RMSD analysis results graph of 4IFI receptor-bound ligands in MD simulation with apo state.

As a result, it has been observed that Epicatechin_ligand100 has a more stable interaction with the 4IFI receptor, and Drug_ligand17 undergoes more structural changes during the binding process.

RMSF analysis, figure 4.30 shows the atomic root-mean-square deviations (RMSF) of ligands and drug bound to the 4IFI (BRCA1) receptor. According to Figure 2.44, it was observed that the deviations were higher in certain regions (1000-1500 atom range or regions after 3000 atoms). These areas are more flexible and move more. Although there were minor differences in some regions between

Epicathecin_ligand100 and Drug_ligand17, they generally exhibited similar fluctuation profiles.

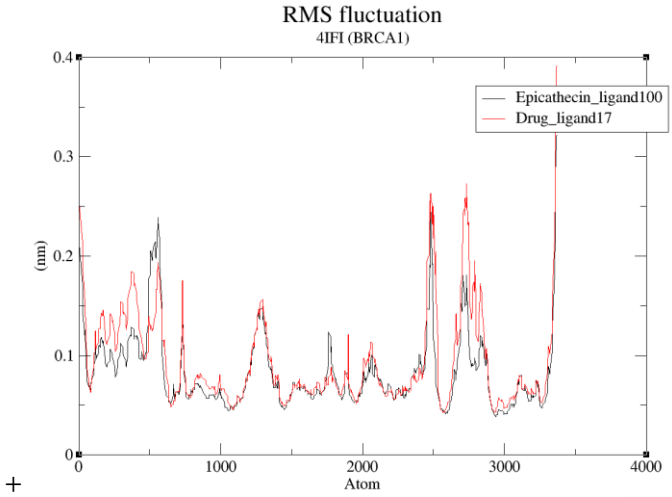


Figure 4.30. RMSF analysis results graph of 4IFI receptor-bound ligands in MD simulation with apo state.

The Rg value for Epicatechin_ligand100 and Drug_ligand17 show a fairly constant Rg value throughout the simulation. The Rg value for both ligands is fixed at around 2.0 nm. This constant Rg value demonstrated that both ligands maintained the overall structural integrity of the protein throughout the simulation and no major structural changes occurred.

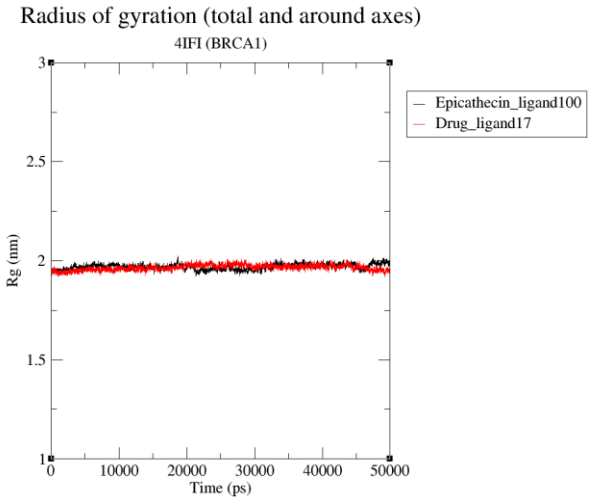


Figure 4.31. R_g analysis results graph of 4IFI receptor-bound ligands in MD simulation with apo state.

Hydrogen bonding for Epicathecin_ligand100 formed many (approximately 10-15) hydrogen bonds at the beginning of the simulation. However, there is a decrease in this number over time. Epicatechin_ligand100 formed more hydrogen bonds throughout the simulation, indicating a stronger binding tendency to the 4IFI receptor. The hydrogen bonds formed by Drug_ligand17 are less than Epicathecin_ligand100 (approximately 2-5). However, over time, this number has experienced some changes and has stabilized at a certain point. Drug_ligand17 formed fewer hydrogen bonds and showed weaker binding to the 4IFI receptor than Epicathecin_ligand100.

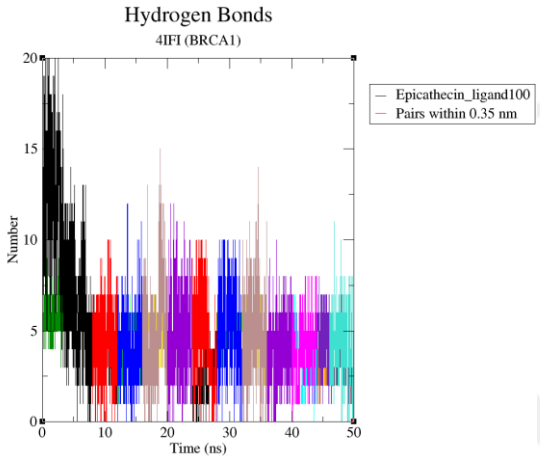


Figure 4.32. Hydrogen bonding analysis results graph of 4IFI receptor-bound Epicathecin_ligand100 in MD simulation.

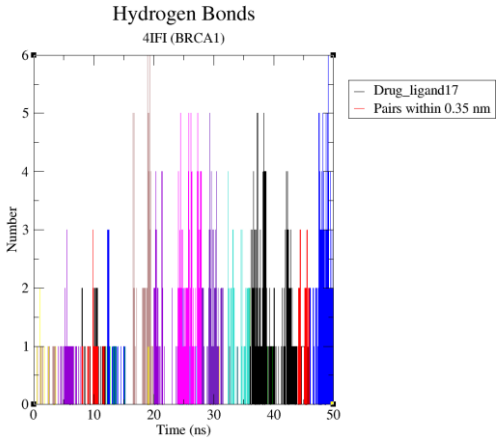


Figure 4.33. Hydrogen bonding analysis results graph of 4IFI receptor-bound Drug_ligand17 in MD simulation.

Table 4.21 shows the protein-ligand interaction energies between the 4IFI (BRCA1) receptor and Epicathecin_ligand10 and Drug_ligand17. These values include the average values of Coulomb (Coul-SR) and Lenard-Jones (LJ-SR) interaction energies.

Table 4.21. Protein-ligand interaction energy results of 4IFI receptor-bound ligands in MD simulation.

MODELS	ENERGY	AMOUNT (Average)
Epicathecin_ligand100	Coul-SR: Protein-LIG	-278.905
	LJ-SR: Protein-LIG	-79.4532
Drug_ligand17	Coul-SR: Protein-LIG	-14,422
	LJ-SR: Protein-LIG	-91,6227

According to Table 4.21, the strongest binding in terms of coulomb interactions was provided by Epicathecin_ligand100 (-278.905). This ligand shows the strongest electrostatic interactions with the protein. As a result, Epicathecin_ligand100 shows the strongest binding in terms of Coulomb interactions. Drug_ligand17 showed very strong binding in terms of Van der Waals interactions, but weak electrostatic interactions. In conclusion, Epicathecin_ligand100 showed the strongest binding and therefore may be an ideal candidate for further studies. Drug_ligand17 may be effective in certain situations, but it needs to be optimized for electrostatic interactions.

RDF (Radial Distribution Functions), according to Figure 4.34, Epicathecin_ligand100 has a peak between approximately 0.8-1.2 nm. Thus, the Epicathecin_ligand100 molecule was most densely located at this distance from the BRCA1 protein and exhibited a tighter and more regular interaction with the 4IFI receptor. Drug_ligand17 peaks between approximately 0.6-1.0 nm. This suggests that Drug_ligand17 shows maximum intensity slightly closer to the 4IFI receptor, but is not as intense as Epicathecin_ligand100.

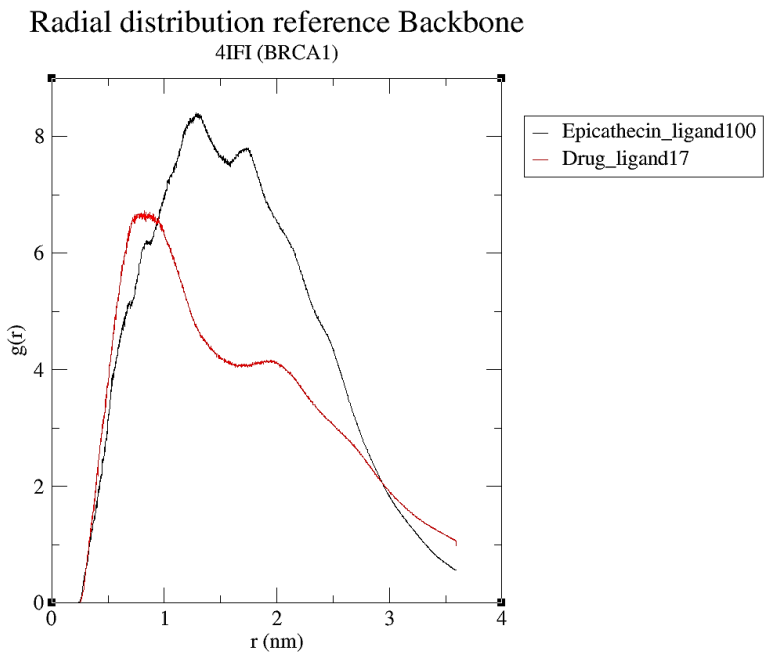


Figure 4.34. RDF (Radial Distribution Functions) analysis results graph of 4IFI receptor-bound ligands in MD simulation with apo state.

As a result, it has been observed that the interactions of Epicathecin_ligand100 with the 4IFI receptor are more intense and regular compared to Drug_ligand17. Epicathecin_ligand100 has a wider range of interactions with the 4IFI receptor and is therefore more stable.

4.2.4. Molecular Dynamics Simulations with 1N0W (BRCA2), Ligands and Drugs

First, the RMSD value, which stands for root mean square deviation, was plotted to measure the conformational changes of CHAINB_ligand70, Cathecin_ligand83 and the Drug_ligand17 depending on the 1N0W (BRCA2) receptor throughout the simulation, to understand how it changes and stabilizes relative to the initial structure.

Additionally, Figure 4.35 shows the time-dependent root-mean-square deviation (RMSD) during molecular dynamics (MD) simulations of three different ligands bound to the 1N0W (BRCA2) receptor. CHAINB_ligand70 shows the highest

RMSD values over time. The average RMSD value is between 3-4 nm, sometimes up to 6 nm. This suggests that ligand70 is highly unstable and undergoes large conformational changes throughout the simulation. Cathecin_ligand83 remained relatively stable relative to its starting structure, showing a constant RMSD value around 2 nm, but showed slight conformational changes. Drug_ligand17 has a low RMSD value and has undergone minimal conformational changes, forming a very stable complex with the BRCA2 receptor.

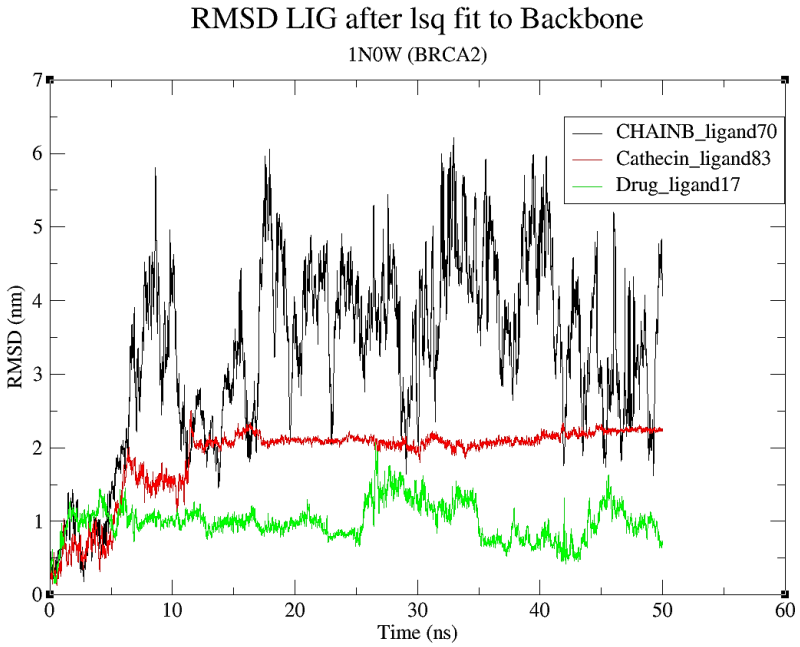


Figure 4.35. RMSD analysis results graph of 1N0W receptor-bound ligands in MD simulation with apo state.

As a result, ccording to the RMSD data of the ligands and the drug with the 1N0W receptor, it shows that Drug_ligand17 has the most stable interaction with the BRCA2 receptor, Cathecin_ligand83 shows an intermediate level of stability, and CHAINB_ligand70 has the most unstable interaction.

RMSF analysis, Figure 4.36 shows the atomic root-mean-square deviations (RMSF) of ligands and drug bound to the 1N0W (BRCA2) receptor. The RMSF value exceeds 1 nm at certain atoms and shows an overall higher release compared to the other Cathecin_ligand83 and Drug_ligand17. This suggests that CHAINB_ligand70 exhibits greater flexibility or instability when bound to protein. Cathecin_ligand83

RMSF value is more constant and stable, usually around 0.5 nm. The RMSF of drug_ligand17 was generally low and moved very little at the atomic level, remaining fairly stable when bound to the 1N0W receptor.

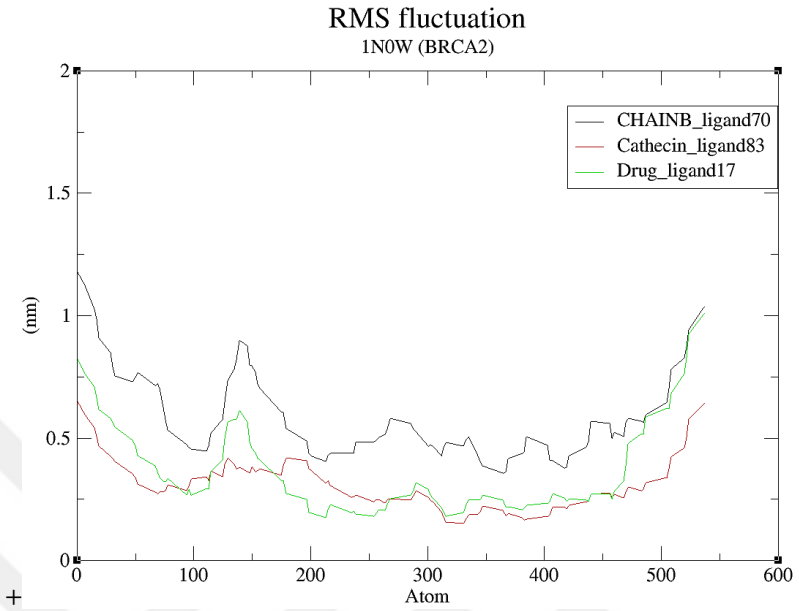


Figure 4.36. RMSF analysis results graph of 1N0W receptor-bound ligands in MD simulation with apo state.

As a result, The RMSF plot shows that CHAINB_ligand70 is the most mobile and flexible ligand, Cathecin_ligand83 shows intermediate stability, and Drug_ligand17 is the most stable ligand.

The Rg value for CHAINB_ligand70 is highly variable over time, ranging between ~1.2 nm and ~1.8 nm. It has been shown that the 1N0W receptor exhibits a dynamic and flexible structure when bound with CHAINB_ligand70, but is generally less stable. Cathecin_ligand83 Rg value showed that it was lower and more stable than CHAINB_ligand70. The Rg value is generally between ~1.1 nm and ~1.3 nm and does not fluctuate much over time. The 1N0W receptor showed a more compact and stable structure with Cathecin_ligand83. When Drug_ligand17 was bound, the 1N0W receptor showed little change over time, remaining between ~1.0 nm and ~1.3 nm. The 1N0W receptor exhibited a very stable and compact structure with Drug_ligan17.

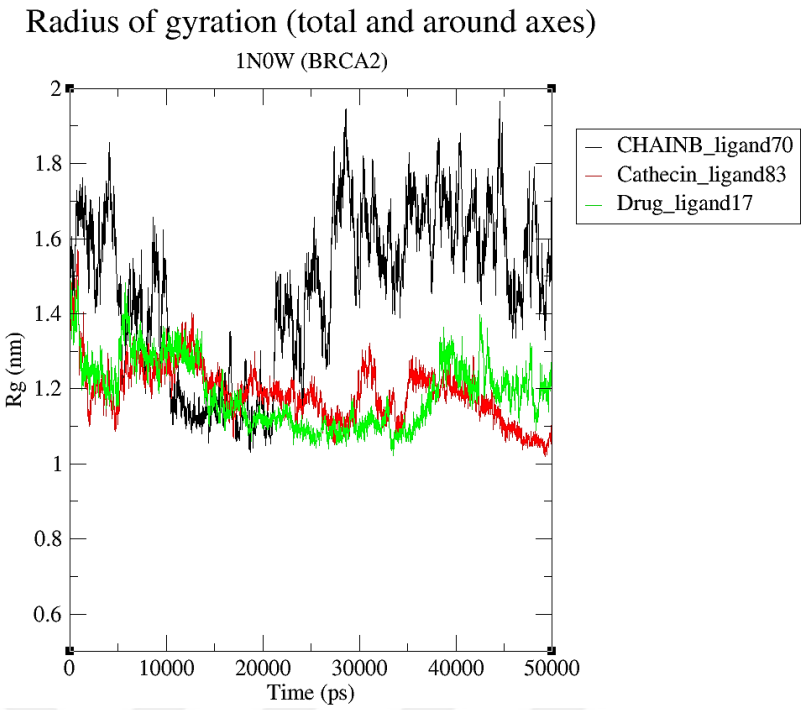


Figure 4.37. R_g analysis results graph of 1N0W receptor-bound ligands in MD simulation with apo state.

As a result, according to these results, CHAINB_ligand70 has greater structural flexibility and a higher caused the value. This demonstrated that the 1N0W receptor displays a larger and potentially more disordered structure. When bound with Cathecin_ligand83 and Drug_ligand17, the 1N0W receptor exhibited a more compact and stable structure. However, Drug_ligand17 enabled the 1N0W receptor to form the most stable and compact structure with the lowest Rg value.

Hydrogen bonding of CHAINB_ligand70, Figure 2.52, the number of CHAINB_ligand70 hydrogen bonds remained quite low over time, and the hydrogen bonds between CHAINB_ligand70 and the 1N0W receptor were generally weak and limited. Also, around 40-50 ns, a few hydrogen bonds were formed again, but these bonds remained in low numbers. These bonds did not remain stable over time and were often short-lived. This shows that the 1N0W receptor exhibits a weak interaction with CHAINB_ligand70.

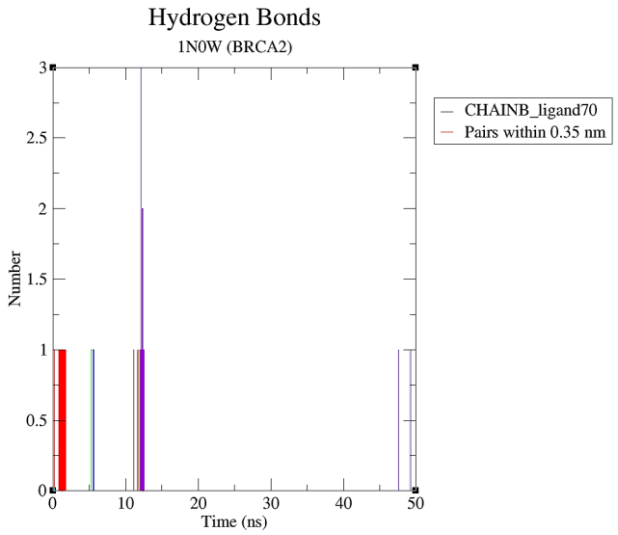


Figure 4.38. Hydrogen bonding analysis results graph of 1N0W receptor-bound CHAINB_ligand70 in MD simulation.

In the graph of cathecin_ligand83 (Figure 4.39), it can be seen that hydrogen bonds generally occur between 0 and 10. This variability in the number of hydrogen bonds indicated that the binding location of Cathecin_ligand83 to the 1N0W receptor and the overall structure of the receptor changed over time. According to Figure 2.20, the binding dynamics of drug17 to the 1N0W receptor varied over time. The last part (between 40-50 ns), where the variability increases, indicates that the binding of drug17 to the 1N0W receptor becomes stable or a significant structure change occurs in the 1N0W receptor-drug17 complex. In summary, the low number of hydrogen bonds indicates that drug17 does not bind tightly to the 1N0W receptor or the binding is weak.

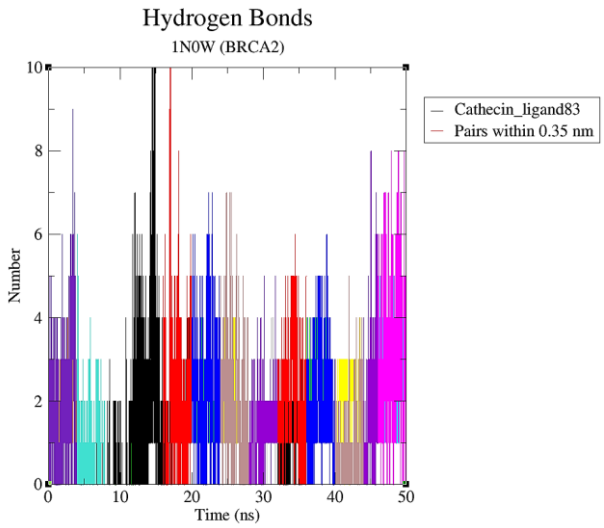


Figure 4.39. Hydrogen bonding analysis results graph of 1N0W receptor-bound Cathecin_ligand83 in MD simulation.

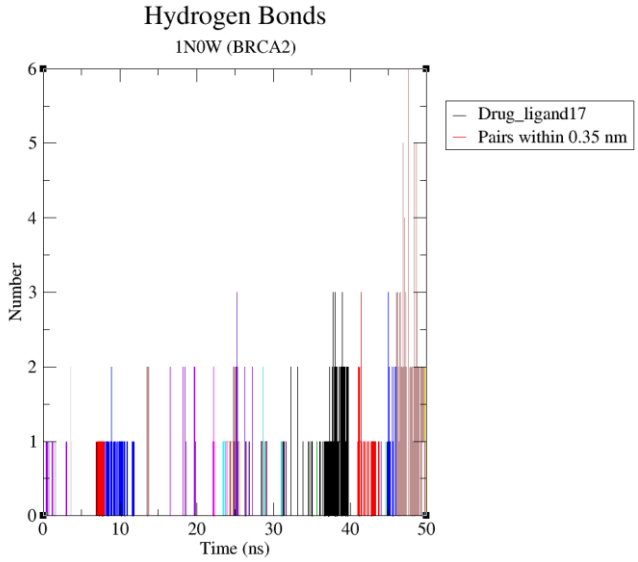


Figure 4.40. Hydrogen bonding analysis results graph of 1N0W receptor-bound Drug_ligand17 in MD simulation.

Table 4.22 shows the protein-ligand interaction energies between the 1N0W (BRCA2) receptor and CHAINB_ligand70, Cathecin_ligand83 and Drug_ligand17. These values include the average values of Coulomb (Coul-SR) and Lenard-Jones (LJ-SR) interaction energies.

Table 4.22. Protein-ligand interaction energy results of 1N0W receptor-bound ligands in MD simulation.

MODELS	ENERGY	AMOUNT (Average)
CHAINB_ligand70	Coul-SR: Protein-LIG	-3.08932
	LJ-SR: Protein-LIG	-9.23314
Cathecin_ligand83	Coul-SR: Protein-LIG	-64.3487
	LJ-SR: Protein-LIG	-117.188
Drug_ligand17	Coul-SR: Protein-LIG	-10.4756
	LJ-SR: Protein-LIG	-99.493

According to Table 4.22, CHAINB_ligand70 has the lowest Coulomb (-3.08932) and Lennard-Jones energy (-9.23314) values with the 1N0W receptor. This showed that CHAINB_ligand70 has a relatively weak binding energy with the 1N0W receptor. Cathecin_ligand83 has the strongest Coulomb (-64.3487) and Lennard-Jones energy (-117.188) values. Thus, Cathecin_ligand83 formed the strongest interaction with the 1N0W receptor. This means that it binds to the 1N0W receptor more strongly than other ligands and forms a more stable complex. Drug_ligand17 has weaker binding energies compared to Cathecin_ligand83, but stronger than CHAINB_ligand70.

RDF (Radial Distribution Functions), according to Figure 4.41, CHAINB_ligand17, shows a more distant and weaker binding to the 1N0W receptor and has the lowest peak. This showed that CHAINB_ligand17 was located farther from the 1N0W receptor and at lower density. Cathecin_ligand83 showed the highest and closest interaction with the 1N0W receptor. It has the highest peak in the graph and is tightly bound to the 1N0W receptor. Although drug_ligand17 was located at a slightly greater distance than Cathecin_ligand83, it still interacted closely and extensively with the 1N0W receptor.

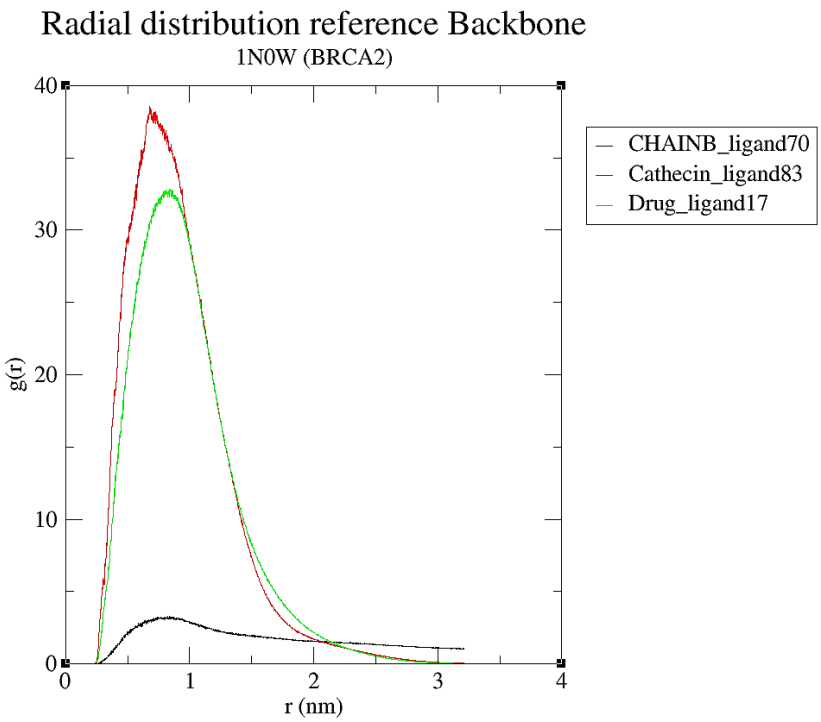


Figure 4.41. RDF (Radial Distribution Functions) analysis results graph of 1N0W receptor-bound ligands in MD simulation with apo state.

As a result, among the CHAINB_ligand70, Cathecin_ligand83, Drug_ligand17 ligands, the ones that bind most strongly to the reference atom are Cathecin_ligand83, Drug_ligand17 and CHAINB_ligand17 respectively.

4.3. Geometry Optimization Results for Selected Conformations

In this thesis, at this stage, from the most stable models obtained as a result of molecular modeling (molecular docking & molecular dynamics simulation), conformation number 84 for flav100 (conf84_flav100), conformation number 28 for flav93 (conf28_flav93), conformation number 23 for flav100 (conf23_flav100), and conformation number 3 for flav83 (conf3_flav83) were selected.

The aim here is to create starting geometries using only the H-bonding amino acid residues of the initial geometries of these ligands in Figure 4.2, Figure 4.6, Figure 4.9 and Figure 4.13. The final geometries obtained as a result of geometry optimization using the B3LYP/def2-SVP set are shown in Figure 4.42, Figure 4.43, Figure 4.44 and Figure 4.45.

Figure 4.42 shows the results of Density Functional Theory (DFT) optimization final in relation to the 7UJM receptor. In terms of molecular modeling, structure shows the effect of positioning and geometry optimization.

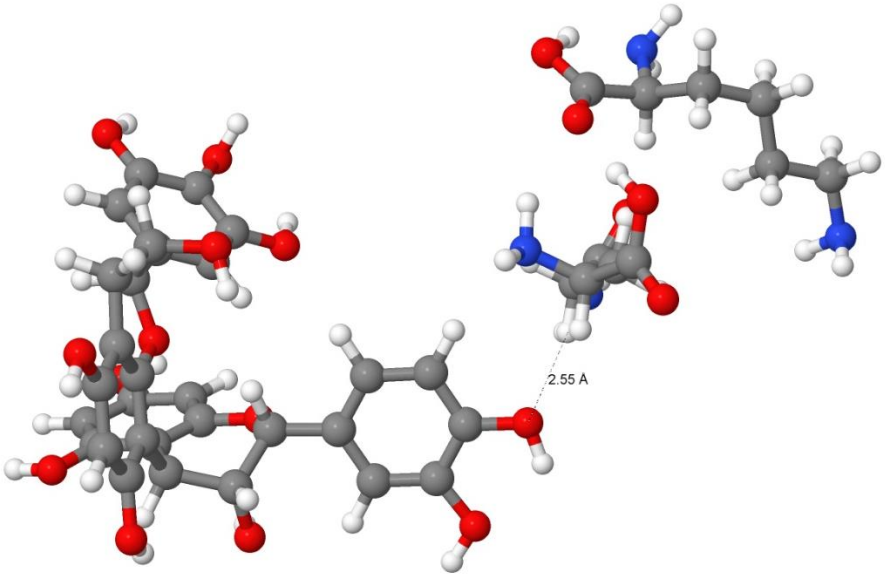


Figure 4.42. Final geometries after DFT calculation for ligand/7UJM residues.

According to Figure 4.42 shows the distances between ligand and residue as 2.55 Å. The distances are important in terms of the possibility of bond formation and intermolecular interaction. Additionally, after DFT optimization, the structure became more stable. H bond interactions provide stability. The results of these energy interactions are shown in Table 4.23.

The decrease in the distance between H and O molecules indicates a redistribution of electron density and intermolecular bonds becoming more prominent. This showed that a more stable binding state was achieved in the interaction of the 7UJM receptor with the flavonoid.

Figure 4.43 shows the results of Density Functional Theory (DFT) optimization final in relation to the 1R5K receptor. In terms of molecular modeling, structure shows the effect of positioning and geometry optimization before DFT.

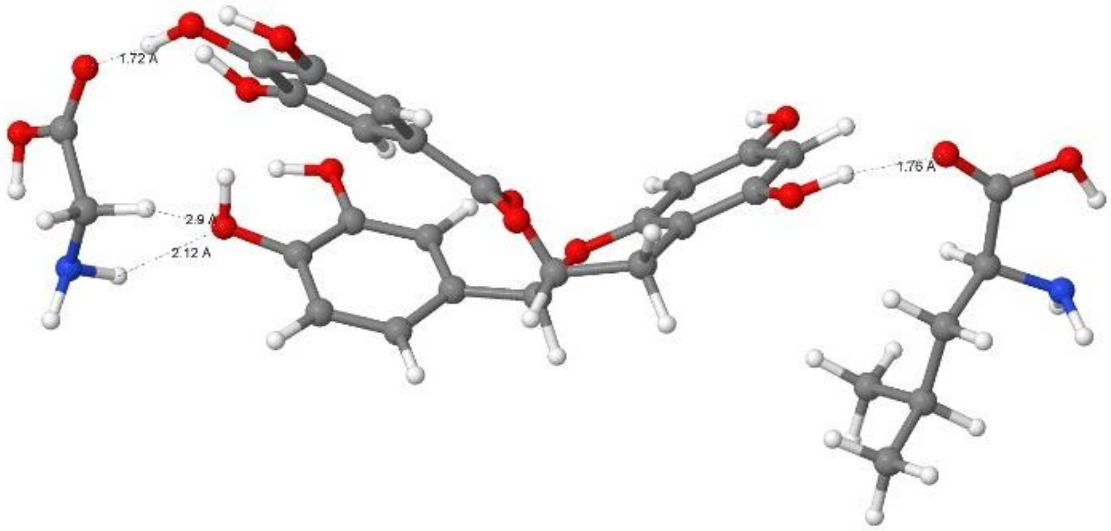


Figure 4.43. Final geometries after DFT calculation for ligand/1R5K residues.

According to Figure 4.6 shows The bond length distances between H and O molecules have values such as 2.08 Å, 3.01 Å, 4.82 Å. The structure generally has a more irregular and non-geometrically optimized shape.

After DFT optimization, the structure became more stable in terms of energy. The bonds between molecules are rearranged and the distances are optimized. The distances between bond lengths decreased to values such as 1.72 Å, 1.76 Å, 2.12 Å, respectively (Figure 4.43) and became shorter and more stable. This showed that a more stable binding state was achieved in the interaction of the 1R5K receptor with the flavonoid.

Figure 4.44 shows the results of Density Functional Theory (DFT) optimization final in relation to the 4IFI receptor. In terms of molecular modeling, structure shows the effect of positioning and geometry optimization.

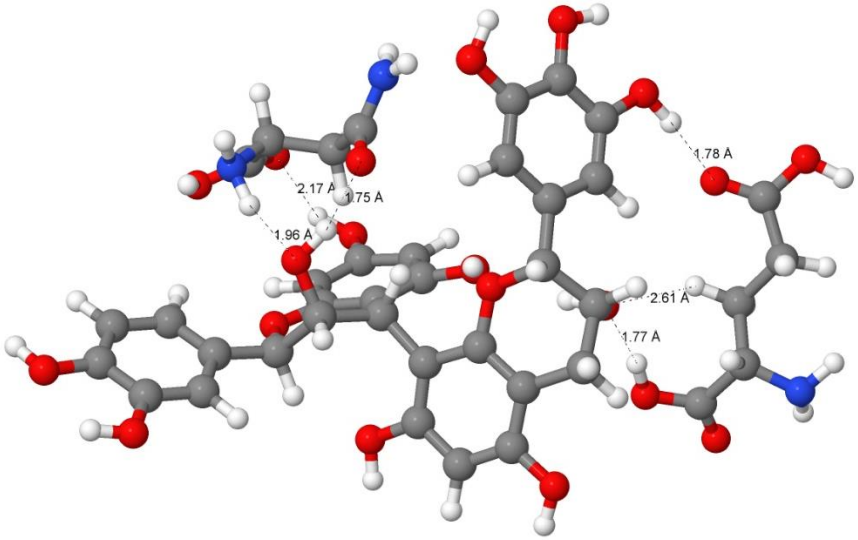


Figure 4.44. Final geometries after DFT calculation for ligand/4IFI residues.

According to Figure 4.9 shows The bond length distances between H and O molecules have values such as 2.12 Å, 2.27 Å, 2.35 Å, 2.85 Å respectively. The structure generally has a more irregular and non-geometrically optimized shape.

After DFT optimization, the structure became more stable in terms of energy. The bonds between molecules are rearranged and the distances are optimized. The distances between bond lengths decreased to values such as 1.98 Å, 1.77 Å, 2.17 Å (Figure 4.44) and became shorter and more stable. This showed that a more stable binding state was achieved in the interaction of the 4IFI receptor with the flavonoid. Also, after optimization, the bonds tightened and the structure became more stable. This suggests that flavonoid and ligand interactions at the 4IFI receptor are stronger and more stable.

Figure 4.45 shows the results of Density Functional Theory (DFT) optimization final in relation to the 1N0W receptor. In terms of molecular modeling, structure shows the effect of positioning and geometry optimization.

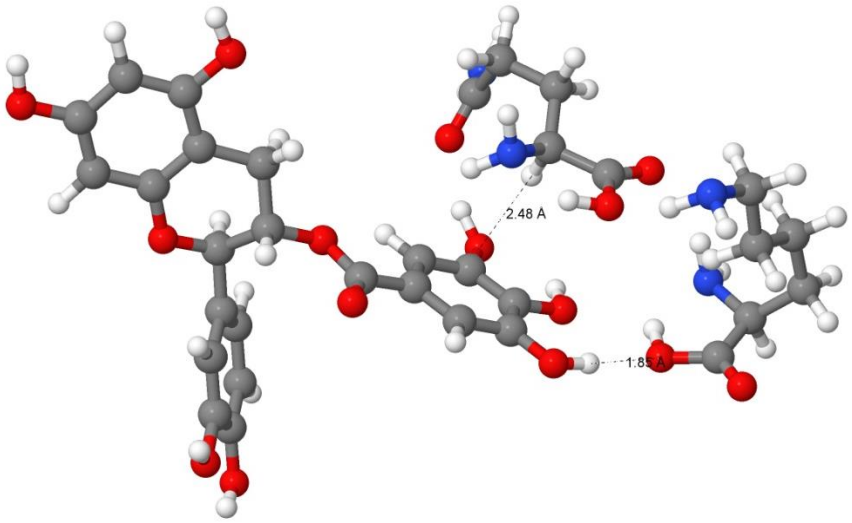


Figure 4.45. Final geometries after DFT calculation for ligand/1N0W residues.

According to Figure 4.13 shows The bond length distances between H and O molecules have values such as 3.64 Å ve 4.05 Å respectively. The structure generally has a more irregular and non-geometrically optimized shape.

After DFT optimization, the structure became more stable in terms of energy. The bonds between molecules are rearranged and the distances are optimized. The distances between bond lengths decreased to values such as 2.48 Å ve 1.85 Å respectively (Figure 4.45) and became shorter and more stable. This showed that a more stable binding state was achieved in the interaction of the 1N0W receptor with the flavonoid. Also, after optimization, the bonds tightened and the structure became more stable. This suggests that flavonoid and ligand interactions at the 1N0W receptor are stronger and more stable.

Table 4.23 shows various interaction energy values based on DFT (Density Functional Theory) calculations and ORCA results. Each calculation provides information about the energies of the receptor-ligand complex. The interaction energy of the 4IFI (BRCA1) receptor-ligand complex is -0.079, indicating the strongest binding by Hartree. The weakest binding is -0.007 Hartree in the 7UJM (ER) complex.

However, when the interaction energy is examined in kcal/mol, it is seen that the 4IFI (BRCA1) receptor shows the strongest binding (-50.02 kcal/mol) against the flav100 ligand. The lowest binding energy belongs to the 7UJM (ER) receptor (-4.46 kcal/mol). According to the results in Table 4.23, it appears that this complex is the most stable structure, as the total energy of the 4IFI (BRCA1) receptor and the flav100 ligand is the lowest and the interaction energy is the highest (most negative). The interaction energy of the 7UJM (ER) receptor-ligand complex is quite low. This indicates that the binding is weaker and the stability of the complex is less.

As a result, considering the energy values and interaction energies, the compatibility between the receptor and the ligand varies. In particular, BRCA1 and BRCA2 receptors appear to provide stronger interactions.

Table 4.23. Orca interaction energy results after DFT geometry optimization.

RECEPTORS	**7UJM (ER)**	**1R5K (ER)**	**4IFI (BRCA1)**	**1N0W (BRCA2)**
CONF.	flav100 (conf84_flav100_ dft.inp)	flav93 (conf28_flav93_ dft.inp)	flav100 (conf23_flav100_ dft_inp)	flav83 (conf3_flav83_ dft.inp)
Total	-3.161,5091060355	-2.324,62179443980	-3.176,8350013767	-2.626,9134679812
Receptor	-1.027,4801303133	-725,1726808020	-2.134,0135583009	-1.027,4746283358
Ligand	-2.134,0218559301	-1.599,38773688540	-1.042,741730784	-1.599,4052057411
IE (Interaction Energy), Hartree	-0.0071197921	-0.061376752	-0.0797122918	-0.0336339043
IE (Interaction Energy) kcal/mol	-4.4677336207	-38.51446452	-50.0201805150	-21.1055776535

CHAPTER V. CONCLUSION AND SUGGESTIONS

In this master's thesis, a comprehensive approach integrating Molecular Docking, Molecular Dynamics Simulations (MDS), and Density Functional Theory (DFT) has been utilized to model breast cancer at the atomistic level, focusing specifically on the potential therapeutic effects of grape seed flavonoids. The findings of this study provide several key insights into the interaction mechanisms between selected flavonoids and breast cancer-related proteins. Firstly, the Molecular Docking results demonstrated the binding affinities and potential inhibitory effects of the flavonoids on the targeted estrogen receptors (7UJM and 1R5K) and BRCA proteins (4IFI and 1N0W). The identified flavonoids showed promising binding energies, suggesting their potential as effective inhibitors in breast cancer treatment. These results provide a basis for the selection of optimal ligand-receptor pairs for further analysis. Subsequently, the Molecular Dynamics Simulations offered valuable information regarding the stability and dynamic behavior of the ligand-receptor complexes in a simulated biological environment. The stability of these complexes, as evidenced by parameters such as the radius of gyration and hydrogen bonding patterns, indicated that the flavonoids maintained stable interactions with the target proteins over the simulation period. This stability is crucial for the therapeutic potential of these compounds. Finally, the Density Functional Theory calculations provided detailed insights into the electronic properties of the ligand-receptor complexes, further validating the interaction mechanisms at a quantum mechanical level. The combination of these methods not only reinforced the reliability of the results but also provided a multi-faceted understanding of the flavonoids' potential role in breast cancer treatment. Overall, this study has successfully demonstrated that grape seed flavonoids, particularly catechins and epicatechins, exhibit strong potential as natural therapeutic agents in the treatment of breast cancer. The integrated computational approach used in this research offers a novel and efficient method for drug discovery and development, paving the way for future studies in this field. Additionally, while this study provides a strong theoretical foundation, it is crucial to validate the findings through in vitro and in vivo experiments. Testing the identified flavonoids in breast cancer cell lines and animal models will help confirm their efficacy and safety as potential therapeutic agents. Following successful experimental validation, clinical trials should be designed to assess the therapeutic potential of these flavonoids in

human subjects. Given the promising results from this study, these compounds could be included in clinical studies targeting specific breast cancer subtypes. Future studies could expand the scope by investigating other flavonoids present in grape seeds or other natural sources. A broader screening of natural compounds might uncover even more potent inhibitors of breast cancer-related proteins. Considering the complex nature of cancer, it may be beneficial to explore the effects of combining grape seed flavonoids with existing chemotherapy drugs. Synergistic effects could enhance treatment efficacy and reduce side effects, offering a more comprehensive approach to breast cancer treatment. Further Molecular Dynamics Simulations could be conducted under varying physiological conditions, such as different pH levels or temperatures, to better understand how environmental factors influence the stability and efficacy of the flavonoid-protein complexes. To further refine the drug discovery process, other computational techniques, such as machine learning algorithms, could be integrated with the current approach. These methods could help predict the behavior of ligand-receptor interactions more accurately and identify novel compounds with high therapeutic potential. By following these suggestions, future research can build on the findings of this thesis, potentially leading to the development of new, effective treatments for breast cancer that harness the power of natural compounds.

REFERENCES

Ana M. Sanchez, H. A., Robert H. Blick, Roberto Raimondi, Rudolf A. Roemer, Vladimir Mihajlovic Fomin. (2023). Density-functional based methods. In *Encyclopedia of Condensed Matter Physics*.

Antoraz, S., Santamaría, R. I., Díaz, M., Sanz, D., & Rodríguez, H. (2015). Toward a new focus in antibiotic and drug discovery from the Streptomyces arsenal. *Front Microbiol, 6*, 461. doi:10.3389/fmicb.2015.00461

Arora, S., Narayan, P., Osgood, C. L., Wedam, S., Prowell, T. M., Gao, J. J., . . . Beaver, J. A. (2022). U.S. FDA Drug Approvals for Breast Cancer: A Decade in Review. *Clin Cancer Res, 28*(6), 1072-1086. doi:10.1158/1078-0432.Ccr-21-2600

Badve, S. S., & Gökmen-Polar, Y. (2019). Ductal carcinoma in situ of breast: update 2019. *Pathology, 51*(6), 563-569. doi:10.1016/j.pathol.2019.07.005

Baines, C. J. (1992). Breast self-examination. *Cancer, 69*(7 Suppl), 1942-1946. doi:10.1002/1097-0142(19920401)69:7+<1942::aid-cncr2820691712>3.0.co;2-k

Barazi, V. C. H. (2024). Inflammatory Breast Cancer. In *StatPearls*.

Barlow, W. E., Lehman, C. D., Zheng, Y., Ballard-Barbash, R., Yankaskas, B. C., Cutter, G. R., . . . Taplin, S. H. (2002). Performance of Diagnostic Mammography for Women With Signs or Symptoms of Breast Cancer. *JNCI: Journal of the National Cancer Institute, 94*(15), 1151-1159. doi:10.1093/jnci/94.15.1151 %J JNCI: Journal of the National Cancer Institute

Barrios, C. H., Reinert, T., & Werutsky, G. (2018). Global Breast Cancer Research: Moving Forward. *Am Soc Clin Oncol Educ Book, 38*, 441-450. doi:10.1200/edbk_209183

Bent, S. F. (2008). Chapter 5 - Semiconductor Surface Chemistry. In A. Nilsson, L. G. M. Pettersson, & J. K. Nørskov (Eds.), *Chemical Bonding at Surfaces and Interfaces* (pp. 323-395). Amsterdam: Elsevier.

Bhushan, A., Gonsalves, A., & Menon, J. U. (2021). Current State of Breast Cancer Diagnosis, Treatment, and Theranostics. *Pharmaceutics, 13*(5). doi:10.3390/pharmaceutics13050723

Bitencourt-Ferreira, G., & de Azevedo Junior, W. F. (2021). Electrostatic Potential Energy in Protein-Drug Complexes. *Curr Med Chem, 28*(24), 4954-4971. doi:10.2174/0929867328666210201150842

Budny, A., Starosławska, E., Budny, B., Wójcik, R., Hys, M., Kozłowski, P., . . . Burdan, F. (2019). [Epidemiology and diagnosis of breast cancer]. *Pol Merkur Lekarski, 46*(275), 195-204.

Cavalcante, F. P., Millen, E. C., Novita, G. G., Zerwes, F. P., Mattar, A., Machado, R. H. S., & Frasson, A. L. (2023). Sentinel lymph node biopsy following neoadjuvant chemotherapy: an evidence-based review and recommendations for current practice. *Chin Clin Oncol, 12*(1), 6. doi:10.21037/cco-22-110

Chan-Yao-Chong, M., Durand, D., & Ha-Duong, T. (2019). Molecular Dynamics Simulations Combined with Nuclear Magnetic Resonance and/or Small-Angle X-ray Scattering Data for Characterizing Intrinsically Disordered Protein Conformational Ensembles.

Cygan, R. (2003). Molecular Models of Metal Sorption on Clay Minerals. In (pp. 143-194).

Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep, 7*, 42717. doi:10.1038/srep42717

de Souza, O. N., & Ornstein, R. L. (1997). Effect of periodic box size on aqueous molecular dynamics simulation of a DNA dodecamer with particle-mesh Ewald method. *Biophysical Journal, 72*(6), 2395-2397. doi:https://doi.org/10.1016/S0006-3495(97)78884-2

Drăgănescu, M., & Carmocan, C. (2017). Hormone Therapy in Breast Cancer. *Chirurgia (Bucur), 112*(4), 413-417. doi:10.21614/chirurgia.112.4.413

Durham, E., Dorr, B., Woetzel, N., Staritzbichler, R., & Meiler, J. (2009). Solvent accessible surface area approximations for rapid and accurate protein structure prediction. *J Mol Model, 15*(9), 1093-1108. doi:10.1007/s00894-009-0454-9

Ferraz da Costa, D. C., Pereira Rangel, L., Quarti, J., Santos, R. A., Silva, J. L., & Fialho, E. (2020). Bioactive Compounds and Metabolites from Grapes and Red Wine in Breast Cancer Chemoprevention and Therapy. *Molecules, 25*(15). doi:10.3390/molecules25153531

Fisusi, F. A., & Akala, E. O. (2019). Drug Combinations in Breast Cancer Therapy. *Pharm Nanotechnol, 7*(1), 3-23. doi:10.2174/2211738507666190122111224

Forum, p. O. (2024). ORCA 6.0.0, Windows, 64bit, Installer. Retrieved from https://orcaforum.kofo.mpg.de/app.php/dlext/?view=detail&df_id=218

Fuleki, T., & Ricardo da Silva, J. M. (1997). Catechin and Procyanidin Composition of Seeds from Grape Cultivars Grown in Ontario. *Journal of Agricultural and Food Chemistry, 45*(4), 1156-1160. doi:10.1021/jf960493k

González-Burgos, E., & Gómez-Serranillos, M. P. (2021). Chapter 4 - Vinca Alkaloids as Chemotherapeutic Agents Against Breast Cancer. In G. Brahmachari (Ed.), *Discovery and Development of Anti-Breast Cancer Agents from Natural Products* (pp. 69-101): Elsevier.

Gu, Y., & Li, M. (2011). Chapter 3 - Molecular Modeling. In H. Ishida & T. Agag (Eds.), *Handbook of Benzoxazine Resins* (pp. 103-110). Amsterdam: Elsevier.

Guide, M. A. U. (2022). Calculating hydrogen bonds: the basics. Retrieved from https://userguide.mdanalysis.org/stable/examples/analysis/hydrogen_bonds/hbonds.html

Guimin, W., & Zhu, W. (2016). Molecular Docking for Drug Discovery and Development: A Widely Used Approach but Far from Perfect. *Future Medicinal Chemistry, 8*. doi:10.4155/fmc-2016-0143

Hangboce, Y. (2019). Atomic structure evolution of high-entropy metallic glass microwires at cryogenic temperature. *Scripta Materialia, 163*, 29-33.

Harbeck, N., & Gnant, M. (2017). Breast cancer. *Lancet, 389*(10074), 1134-1150. doi:10.1016/s0140-6736(16)31891-8

Harbeck, N., Penault-Llorca, F., Cortes, J., Gnant, M., Houssami, N., Poortmans, P., . . . Cardoso, F. (2019). Breast cancer. *Nat Rev Dis Primers, 5*(1), 66. doi:10.1038/s41572-019-0111-2

Hospital, A., Goñi, J. R., Orozco, M., & Gelpí, J. L. (2015). Molecular dynamics simulations: advances and applications. *Adv Appl Bioinform Chem, 8*, 37-47. doi:10.2147/aabc.S70333

Huang, S. Y., & Zou, X. (2010). Advances and challenges in protein-ligand docking. *Int J Mol Sci, 11*(8), 3016-3034. doi:10.3390/ijms11083016

Ilaiyaraja, N., & Khanum, F. (2011). Antioxidant Potential of Tinospora cordifolia Extracts and their Protective Effect on Oxidation of Biomolecules. *Pharmacognosy Journal, 3*(20), 56-62. doi:https://doi.org/10.5530/pj.2011.20.11

Islam, R., Yan, M., Yen, K., Rasol, N., Meng, C., & Lam, D. (2023). Synthesis and biological evaluation of chromone derivatives against triple-negative breast cancer cells. *Medicinal Chemistry Research, 32*. doi:10.1007/s00044-023-03048-4

Jacobson, G., Kaidar-Person, O., Haisraely, O., Galper, S., Rabin, T., Dromi Shahadi, I., . . . Akiva Ben-David, M. (2021). Palliative radiation therapy for symptomatic advance breast cancer. *Scientific Reports, 11*(1), 5282. doi:10.1038/s41598-021-84872-9

Jayaprakasha, G. K., Selvi, T., & Sakariah, K. K. (2003). Antibacterial and antioxidant activities of grape (Vitis vinifera) seed extracts. *Food Research International, 36*(2), 117-122. doi:https://doi.org/10.1016/S0963-9969(02)00116-3

Jesinger, R. A. (2014). Breast anatomy for the interventionalist. *Tech Vasc Interv Radiol, 17*(1), 3-9. doi:10.1053/j.tvir.2013.12.002

Kasi, M. T. A. A. C. A. (2024). Cancer Chemotherapy. In *StatPearls*.

Kim, G., & Bahl, M. (2021). Assessing Risk of Breast Cancer: A Review of Risk Prediction Models. *J Breast Imaging, 3*(2), 144-155. doi:10.1093/jbi/wbab001

Koh, J., & Kim, M. J. (2019). Introduction of a New Staging System of Breast Cancer for Radiologists: An Emphasis on the Prognostic Stage. *Korean J Radiol, 20*(1), 69-82. doi:10.3348/kjr.2018.0231

Kursat M, E. P., Kırbag S. (2009). ANTIMICROBIAL ACTIVITIES OF SOME MEDICINAL PLANTS FROM TURKEY. *Ecological Life Sciences 4 2 42–47*.

Lindahl, E. R. (2008). Molecular dynamics simulations. *Methods Mol Biol, 443*, 3-23. doi:10.1007/978-1-59745-177-2_1

Lopes, D., Dias de Queiros, L., Ávila, A., Monteiro, N., & Macedo, G. (2017). The Importance of Microbial and Enzymatic Bioconversions of Isoflavones in Bioactive Compounds. In (pp. 55-93).

Maloy, S., & Hughes, K. (2013). *Brenner's Encyclopedia of Genetics, Second Edition*.

Massivebio. (2021). Breast Cancer. Retrieved from https://massivebio.com/breast-cancer/

Maughan, K. L., Lutterbie, M. A., & Ham, P. S. (2010). Treatment of breast cancer. *Am Fam Physician, 81*(11), 1339-1346.

Michael Colvin, M. (2003). Alkylating Agents. In P. R. Kufe DW, Weichselbaum RR (Ed.), *Holland-Frei Cancer Medicine. 6th edition.*

Morris, G. M., & Lim-Wilby, M. (2008). Molecular docking. *Methods Mol Biol, 443*, 365-382. doi:10.1007/978-1-59745-177-2_19

Mukherjee, P. K., Kumar, V., & Houghton, P. J. (2007). Screening of Indian medicinal plants for acetylcholinesterase inhibitory activity. *Phytother Res, 21*(12), 1142-1145. doi:10.1002/ptr.2224

Nishiumi, S., Miyamoto, S., Kawabata, K., Ohnishi, K., Mukai, R., Murakami, A., . . . Terao, J. (2011). Dietary flavonoids as cancer-preventive and therapeutic biofactors. *Frontiers in bioscience (Scholar edition), 3*, 1332-1362.

Oldenburg, R. A., Meijers-Heijboer, H., Cornelisse, C. J., & Devilee, P. (2007). Genetic susceptibility for breast cancer: how many more genes to be found? *Crit Rev Oncol Hematol, 63*(2), 125-149. doi:10.1016/j.critrevonc.2006.12.004

Organization, W. H. (2024). Breast Cancer. Retrieved from https://www.who.int/news-room/fact-sheets/detail/breast-cancer

Pandya, S., & Moore, R. G. (2011). Breast development and anatomy. *Clin Obstet Gynecol, 54*(1), 91-95. doi:10.1097/GRF.0b013e318207ffe9

Park, M. Y., Kim, Y., Ha, S. E., Kim, H. H., Bhosale, P. B., Abusaliya, A., . . . Kim, G. S. (2022). Function and Application of Flavonoids in the Breast Cancer. *23*(14), 7732.

Pitera, J. W. (2014). Expected distributions of root-mean-square positional deviations in proteins. *J Phys Chem B, 118*(24), 6526-6530. doi:10.1021/jp412776d

Polgár, C., Kahán, Z., Ivanov, O., Chorváth, M., Ligačová, A., Csejtei, A., . . . Fodor, J. (2022). Radiotherapy of Breast Cancer-Professional Guideline 1st Central-Eastern European Professional Consensus Statement on Breast Cancer. *Pathol Oncol Res, 28*, 1610378. doi:10.3389/pore.2022.1610378

Posysaev, S. (2018). *Applications of density functional theory for modeling metal-semiconductor contacts, reaction pathways, and calculating oxidation states*.

Prasad, M., & Chakraborty, S. N. (2022). Hydrogen bonding in liquid water at 1 GPa : Molecular dynamics simulation study of TIP4P/2005 water model. *Computational and Theoretical Chemistry, 1208*, 113527. doi:https://doi.org/10.1016/j.comptc.2021.113527

Proteomics, C. (2024). Molecular Docking Service. Retrieved from https://www.creative-proteomics.com/services/molecular-docking-service.htm

Provenzano, E., Ulaner, G. A., & Chin, S. F. (2018). Molecular Classification of Breast Cancer. *PET Clin, 13*(3), 325-338. doi:10.1016/j.cpet.2018.02.004

Rakha, E. A., Tse, G. M., & Quinn, C. M. (2023). An update on the pathological classification of breast cancer. *Histopathology, 82*(1), 5-16. doi:10.1111/his.14786

Rocque, G. B., Caston, N. E., Franks, J. A., Williams, C. P., Aswani, M. S., Azuero, A., & Gidwani, R. (2021). Clinical trial representativeness and treatment intensity in a real-world sample of women with early stage breast cancer. *Breast Cancer Res Treat, 190*(3), 531-540. doi:10.1007/s10549-021-06381-7

Rydberg, P., Jørgensen, F. S., & Olsen, L. (2014). Use of density functional theory in drug metabolism studies. *Expert Opin Drug Metab Toxicol, 10*(2), 215-227. doi:10.1517/17425255.2014.864278

Sadiku-Agboola, O., & Sadiku, E. R. (2014). Chapter 3 - Theoretical Modeling of Nanostructured Formation in Polymer Blends. In S. Thomas, R. Shanks, & S. Chandrasekharakurup (Eds.), *Nanostructured Polymer Blends* (pp. 33-99). Oxford: William Andrew Publishing.

Sargsyan, K., Grauffel, C., & Lim, C. (2017). How Molecular Size Impacts RMSD Applications in Molecular Dynamics Simulations. *J Chem Theory Comput, 13*(4), 1518-1524. doi:10.1021/acs.jctc.7b00028

Sheeja, K., & Kuttan, G. (2007). Activation of Cytotoxic T Lymphocyte Responses and Attenuation of Tumor Growth in vivo by Andrographis paniculata Extract and Andrographolide. *Immunopharmacology and immunotoxicology, 29*, 81-93. doi:10.1080/08923970701282726

Shi, J., Yu, J., Pohorly, J. E., & Kakuda, Y. (2003). Polyphenolics in grape seeds-biochemistry and functionality. *J Med Food, 6*(4), 291-299. doi:10.1089/109662003772519831

Shvab, I. (2014). *MOLECULAR DYNAMICS SIMULATION OF WATER AND AQUEOUS SOLUTIONS.* Retrieved from https://researchbank.swinburne.edu.au/file/878b3473-1d10-4e44-bd43-64b907d816b9/1/Igor%20Shvab%20Thesis.pdf

Sun, Y. S., Zhao, Z., Yang, Z. N., Xu, F., Lu, H. J., Zhu, Z. Y., . . . Zhu, H. P. (2017). Risk Factors and Preventions of Breast Cancer. *Int J Biol Sci, 13*(11), 1387-1397. doi:10.7150/ijbs.21635

Tanner, S. Y. M. a. D. E. (2011). Comparing Solvent Models for Molecular Dynamics of Protein.

Toukmaji, A. Y., & Board, J. A. (1996). Ewald summation techniques in perspective: a survey. *Computer Physics Communications, 95*(2), 73-92. doi:https://doi.org/10.1016/0010-4655(96)00016-1

Veronesi, U., Boyle, P., Goldhirsch, A., Orecchia, R., & Viale, G. (2005). Breast cancer. *Lancet, 365*(9472), 1727-1741. doi:10.1016/s0140-6736(05)66546-4

Vinod, S. M., Murugan Sreedevi, S., Krishnan, A., Ravichandran, K., Karthikeyan, P., Kotteswaran, B., & Rajendran, K. (2023). Complexity of the Role of Various Site-Specific and Selective Sudlow Binding Site Drugs in the Energetics and Stability of the Acridinedione Dye-Bovine Serum Albumin Complex: A Molecular Docking Approach. *ACS Omega, 8*(6), 5634-5654. doi:10.1021/acsomega.2c07111

Wang, Y., Jeong, Y., Jhiang, S., Yu, L., & Menq, C.-H. (2014). Quantitative Characterization of Cell Behaviors through Cell Cycle Progression via Automated Cell Tracking. *PloS one, 9*, e98762. doi:10.1371/journal.pone.0098762

Weigelt, B., Geyer, F. C., & Reis-Filho, J. S. (2010). Histological types of breast cancer: how special are they? *Mol Oncol, 4*(3), 192-208. doi:10.1016/j.molonc.2010.04.004

Yang, C.-Y., Phillips, J., Stuckey, J., Bai, L., Sun, H., Delproposto, J., . . . Chinnaswamy, K. (2016). Buried Hydrogen Bond Interactions Contribute to the High Potency of Complement Factor D Inhibitors. *ACS Medicinal Chemistry Letters, 7*. doi:10.1021/acsmedchemlett.6b00299

Yersal, O., & Barutca, S. (2014). Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World J Clin Oncol, 5*(3), 412-424. doi:10.5306/wjco.v5.i3.412

Zhong, L., Li, Y., Xiong, L., Wang, W., Wu, M., Yuan, T., . . . Yang, S. (2021). Small molecules in targeted cancer therapy: advances, challenges, and future perspectives. *Signal Transduct Target Ther, 6*(1), 201. doi:10.1038/s41392-021-00572-w

APPENDICES

APPENDIX A

Table A.1. 7UJM and 1R5K receptors (ER) molecular docking ChEMBL codes and Autodock Vina docking results.

MOLECULAR DOCKING FORMAT		ChEMBL ACTIVITIES			RECEPTORS Docking Results (kcal/mol)	
pdbqt format	**sdf format**	**Order**	**CHEMBL CODE**	**Standard Value (nM)**	**7UJM (ER)**	**1R5K (ER)**
liga	lig	1	CHEMBL42707	0.002	-7.3	-7.3
liga	lig	2	CHEMBL4591338	0.01	-6.7	-6.7
liga	lig	3	CHEMBL4439408	0.021	-8.3	-8.4
liga	lig	4	CHEMBL4447340	0.03	-7.6	-8.9
liga	lig	5	CHEMBL4649161	0.031	-7.2	-11.4
liga	lig	6	CHEMBL4743256	0.03981	-6.8	-7.2
liga	lig	7	CHEMBL4518283	0.04	-7.5	-7.8
liga	lig	8	CHEMBL4464224	0.045	-9.1	-8.9
liga	lig	9	CHEMBL1193539	0.046	-8.3	-8.3
liga	lig	10	CHEMBL4650316	0.05	-7.7	-7.4
liga	lig	11	CHEMBL4455953	0.05	-7.7	-8.2
liga	lig	12	CHEMBL4777139	0.05012	-8.2	-10.0
liga	lig	13	CHEMBL4449709	0.051	-9.2	-9.7
liga	lig	14	CHEMBL4437985	0.055	-9.1	-9.0
liga	lig	15	CHEMBL4464058	0.059	-8.1	-8.7
liga	lig	16	CHEMBL4866043	0.06	-9.0	-7.7
liga	lig	17	CHEMBL4759597	0.0631	-7.5	-9.0
liga	lig	18	CHEMBL4635493	0.065	-7.7	-11.5
liga	lig	19	CHEMBL4463541	0.07	-7.5	-8.8
liga	lig	20	CHEMBL4539802	0.076	-8.3	-11,4
liga	lig	21	CHEMBL4528514	0.07943	-7.7	-7.4
liga	lig	22	CHEMBL4594150	0.08	-7.9	-10.7
liga	lig	23	CHEMBL4646522	0.083	-7.4	-8.0
liga	lig	24	CHEMBL4864829	0.09	-7.4	-7.7
liga	lig	25	CHEMBL4845726	0.09	-7.8	-7.6
liga	lig	26	CHEMBL4648503	0.09	-10.2	-7.2
liga	lig	27	CHEMBL4644486	0.09	-8.2	-11.9
liga	lig	28	CHEMBL4754652	0.1	-7.6	-8.0
liga	lig	29	CHEMBL4164133	0.1	-9.1	**-12.2**
liga	lig	30	CHEMBL4438260	0.1	-7.7	-7.8
liga	lig	31	CHEMBL4871161	0.1	-7.5	-7.6
liga	lig	32	CHEMBL4435067	0.1	-7.3	-7.4
liga	lig	33	CHEMBL4563316	0.1	-8.1	-8.5
liga	lig	34	CHEMBL4164869	0.1	-8.3	-10.9
liga	lig	35	CHEMBL4758424	0.1	-7.1	-7.3

liga	lig	36	CHEMBL4453442	0.1	-8.5	-6.9
liga	lig	37	CHEMBL4756497	0.1259	-7.8	-8.2
liga	lig	38	CHEMBL4650365	0.1259	-9.4	-9.3
liga	lig	39	CHEMBL4752435	0.1259	-7.1	-7.6
liga	lig	40	CHEMBL4799234	0.1259	-7.4	-7.9
liga	lig	41	CHEMBL4780809	0.1259	-7.7	-7.3
liga	lig	42	CHEMBL4857736	0.13	-7.5	-7.5
liga	lig	43	CHEMBL3623004	0.138	-7.7	-7.7
liga	lig	44	CHEMBL4877338	0.15	-7.4	-7.7
liga	lig	45	CHEMBL4756831	0.1585	-8.5	-8.0
liga	lig	46	CHEMBL4785486	0.1585	-7.3	-8.0
liga	lig	47	CHEMBL4795387	0.1585	-7.6	-7.5
liga	lig	48	CHEMBL4786058	0.16	-7.4	-7.5
liga	lig	49	CHEMBL4641732	0.17	-7.3	-7.9
liga	lig	50	CHEMBL4755196	0.17	-7.3	-11.4
liga	lig	51	CHEMBL4447143	0.17	-8.2	-8.6
liga	lig	52	CHEMBL4785851	0.17	-7.0	-10.0
liga	lig	53	CHEMBL4778779	0.17	-7.7	-7.6
liga	lig	54	CHEMBL4640902	0.17	-8.5	-7.5
liga	lig	55	CHEMBL4633886	0.19	-7.7	-7.2
liga	lig	56	CHEMBL4583286	0.1995	-7.3	-7.8
liga	lig	57	CHEMBL4751770	0.1995	-7.4	-7.1
liga	lig	58	CHEMBL4751971	0.1995	-7.1	-7.1
liga	lig	59	CHEMBL4759688	0.1995	-7.6	-8.1
liga	lig	60	CHEMBL4749445	0.1995	-7.2	-7.4
liga	lig	61	CHEMBL4792482	0.1995	-7.2	-7.0
liga	lig	62	CHEMBL4789498	0.1995	-7.5	-7.2
liga	lig	63	CHEMBL4076124	0.2	-10.7	-8.4
liga	lig	64	CHEMBL4438826	0.2	-8.5	-9.1
liga	lig	65	CHEMBL198803	0.2	-8.7	-8.4
liga	lig	66	CHEMBL4167575	0.2	-9.4	-9.1
liga	lig	67	CHEMBL1088485	0.2	-10.7	-8.4
liga	lig	68	CHEMBL4069711	0.2	-7.5	-7.9
liga	lig	69	CHEMBL4160292	0.2	-7.8	-7.5
liga	lig	70	CHEMBL4177307	0.2	-7.5	-7.7
liga	lig	71	CHEMBL4165845	0.2	-7.3	-8.2
liga	lig	72	CHEMBL4174789	0.2	-9.1	-8.8
liga	lig	73	CHEMBL4168008	0.2	-10.9	-11.1
liga	lig	74	CHEMBL4171078	0.2	-9.1	-8.7
liga	lig	75	CHEMBL4176563	0.2	-8.4	-8.5
liga	lig	76	CHEMBL4438992	0.2	-8.6	-8.6
liga	lig	77	CHEMBL4464718	0.2	-7.5	-6.9
liga	lig	78	CHEMBL4164447	0.2	-7.7	-8.7
liga	lig	79	CHEMBL4461269	0.2	-7.6	-8.3
liga	lig	80	CHEMBL4163562	0.2	-8.7	-8.4
liga	lig	81	CHEMBL4171332	0.2	-7.8	-7.9
liga	lig	82	CHEMBL4576272	0.2	**-11.4**	-7.9
liga	lig	83	CHEMBL4175085	0.2	-8.7	-8.6
liga	lig	84	CHEMBL3623004	0.2089	-7.7	-7.7
liga	lig	85	CHEMBL4645364	0.21	-7.9	-8.3
liga	lig	86	CHEMBL4742066	0.22	-7.1	-7.9
liga	lig	87	CHEMBL180707	0.22	-8.6	-9.0
liga	lig	88	CHEMBL178679	0.22	-8.8	-9.4

liga	lig	89	CHEMBL4866232	0.23	-7.6	-7.5
liga	lig	90	CHEMBL4745817	0.24	-7.3	-7.4
liga	lig	91	CHEMBL4443826	0.24	-7.9	-8.4
liga	lig	92	CHEMBL3774584	0.25	-7.6	-7.0
liga	lig	93	CHEMBL4873860	0.25	-7.6	-7.8
liga	lig	94	CHEMBL4789223	0.2512	-7.4	-9.8
liga	lig	95	CHEMBL4591926	0.2512	-7.7	-7.3
liga	lig	96	CHEMBL4448852	0.2512	-7.7	-7.9
liga	lig	97	CHEMBL4777960	0.2512	-7.6	-8.0
liga	lig	98	CHEMBL4527655	0.257	-7.1	-8.4
liga	lig	99	CHEMBL180873	0.27	-8.6	-8.9
liga	lig	100	CHEMBL1087419	0.28	-8.1	-9.5

Table A.2. 4IFI receptor (BRCA1 gene) molecular docking ChEMBL codes and Autodock Vina docking results.

MOLECULAR DOCKING FORMAT		**ChEMBL ACTIVITIES**			**RECEPTORS Docking Results (kcal/mol)**
pdbqt format	**sdf format**	**Order**	**CHEMBL CODE**	**Standard Value (nM)**	**4IFI (BRCA1)**
liga	lig	1	CHEMBL1784784	1000	-2.8
liga	lig	2	CHEMBL1784785	1600	-2.4
liga	lig	3	CHEMBL1784703	3200	-3.2
liga	lig	4	CHEMBL1784774	4.600	-3.0
liga	lig	5	CHEMBL1784776	7.100	-3.1
liga	lig	6	CHEMBL1784783	10.800	-2.9
liga	lig	7	CHEMBL1784779	14.900	-3.4
liga	lig	8	CHEMBL1784704	15000	-3.5
liga	lig	9	CHEMBL2369531	17700	-3.8
liga	lig	10	CHEMBL1784777	18400	-2.9
liga	lig	11	CHEMBL1784775	30100	-3.1
liga	lig	12	CHEMBL1784770	35000	**-4.4**
liga	lig	13	CHEMBL1784772	52800	-3.4
liga	lig	14	CHEMBL1784780	98400	-3.1
liga	lig	15	CHEMBL2369532	171700	-3.8
liga	lig	16	CHEMBL1784773	250000	-3.2
liga	lig	17	CHEMBL1784782	250000	-3.2
liga	lig	18	CHEMBL1784771	250000	-3.2
liga	lig	19	CHEMBL1784781	250000	-3.3

Table A.3. 1N0W receptor (BRCA2 gene) molecular docking ChEMBL codes and Autodock Vina docking results.

MOLECULAR DOCKING FORMAT		ChEMBL ACTIVITIES			RECEPTORS Docking Results (kcal/mol)
pdbqt format	**sdf format**	**Order**	**CHEMBL CODE**	**Standard Value (nM)**	**1N0W (BRCA2)**
liga	lig	1	CHEMBL2316725	250	-5.9
liga	lig	2	CHEMBL2316724	370	-5.8
liga	lig	3	CHEMBL2316723	810	-7.5
liga	lig	4	CHEMBL2316721	1.400	-6.7
liga	lig	5	CHEMBL3823105	3.000	-6.6
liga	lig	6	CHEMBL2316738	3.100	-6.4
liga	lig	7	CHEMBL2316739	5.290	-6.5
liga	lig	8	CHEMBL1271764	6430	-6.4
liga	lig	9	CHEMBL1271993	6820	-6.0
liga	lig	10	CHEMBL2316736	7910	-6.3
liga	lig	11	CHEMBL2316735	8320	-6.0
liga	lig	12	CHEMBL3823002	8400	-7.0
liga	lig	13	CHEMBL3822706	9800	-6.9
liga	lig	14	CHEMBL2316737	9940	-6.3
liga	lig	15	CHEMBL3824267	13000	-6.6
liga	lig	16	CHEMBL1547184	13100	-6.7
liga	lig	17	CHEMBL2316734	13160	-6.1
liga	lig	18	CHEMBL3824107	13200	-6.5
liga	lig	19	CHEMBL1414442	14880	-5.6
liga	lig	20	CHEMBL2316719	15620	-6.7
liga	lig	21	CHEMBL1334552	19700	-5.6
liga	lig	22	CHEMBL3823807	20000	-6.6
liga	lig	23	CHEMBL3824080	22100	-6.5
liga	lig	24	CHEMBL3823899	25000	-6.9
liga	lig	25	CHEMBL3822575	27200	-7.3
liga	lig	26	CHEMBL3822556	27600	-6.8
liga	lig	27	CHEMBL3823875	27600	-7.1
liga	lig	28	CHEMBL3823795	34000	-7.1
liga	lig	29	CHEMBL3823373	35000	-6.4
liga	lig	30	CHEMBL2316733	38170	-5.9
liga	lig	31	CHEMBL3823248	38600	-6.5
liga	lig	32	CHEMBL2316726	44170	-7.1
liga	lig	33	CHEMBL3823178	46100	-6.7
liga	lig	34	CHEMBL2316730	50000	-6.8
liga	lig	35	CHEMBL2316728	50000	-7.1
liga	lig	36	CHEMBL3824216	57800	-7.2
liga	lig	37	CHEMBL3823341	64900	-6.8
liga	lig	38	CHEMBL3823966	64900	-6.6
liga	lig	39	CHEMBL3822942	71700	-6.5
liga	lig	40	CHEMBL3824220	91100	-6.8

liga	lig	41	CHEMBL3823002	96900	-7.0
liga	lig	42	CHEMBL3823248	100000	-6.3
liga	lig	43	CHEMBL224458	100000	-6.5
liga	lig	44	CHEMBL3823934	100000	-7.1
liga	lig	45	CHEMBL3822905	100000	-7.3
liga	lig	46	CHEMBL3823629	100000	-7.6
liga	lig	47	CHEMBL3823737	100000	-6.6
liga	lig	48	CHEMBL3824180	100000	-6.9
liga	lig	49	CHEMBL3823341	100000	-6.7
liga	lig	50	CHEMBL3823677	100000	-7.2
liga	lig	51	CHEMBL3823737	100000	-6.6
liga	lig	52	CHEMBL3824268	100000	-7.6
liga	lig	53	CHEMBL3823840	100000	-7.1
liga	lig	54	CHEMBL3823105	100000	-6.7
liga	lig	55	CHEMBL3823795	100000	-7.0
liga	lig	56	CHEMBL3822797	100000	-6.8
liga	lig	70	CHEMBL3823654	100000	**-7.6**
liga	lig	58	CHEMBL3824107	100000	-6.5

Table A.4. 7UJM (ER), 1R5K (ER), 4IFI (BRCA1) and 1N0W (BRCA2) receptors molecular docking flavonoids and Autodock Vina docking results.

MOLECULAR DOCKING FORMAT		**CHEMICAL ID**		**RECEPTORS Docking Results (kcal/mol)**			
pdbqt	**sdf**	**Order**	**PubChem CID**	**7UJM**	**1R5K**	**4IFI**	**1N0W**
GALLIC ACID							
flav	analog	1	370	-6.3	-5.6	-2.9	-5.2
flav	analog	2	10059574	-6.7	-6.8	-3.0	-5.9
flav	analog	3	4947	-6.1	-6.5	-3.0	-5.7
flav	analog	4	54600902	-7.1	-6.4	-3.2	-5.6
flav	analog	5	7428	-6.3	-5.9	-2.8	-5.2
flav	analog	6	13250	-6.5	-6.2	-2.9	-5.3
flav	analog	7	10742	-6.4	-5.8	-3.0	-5.4
flav	analog	8	8357	-6.1	-5.3	-2.7	-4.9
flav	analog	9	65064	-7.7	-8.6	-4.1	-6.9
flav	analog	10	341	-7.6	-7.4	-4.2	-7.1
flav	analog	11	14128	-6.6	-6.5	-3.1	-5.6
flav	analog	12	78016	-6.3	-6.0	-3.1	-5.1
flav	analog	13	19829	-6.3	-6.0	-2.9	-5.8
flav	analog	14	69256	-6.4	-6.0	-3.0	-5.8
flav	analog	15	75596	-6.9	-6.6	-3.2	-6.0
flav	analog	16	95088	-7.0	-6.2	-3.1	-5.2
flav	analog	17	101653700	-6.6	-6.8	-3.1	-6.0
flav	analog	18	592744	-6.7	-6.8	-3.2	-5.9
flav	analog	19	4628122	-7.1	-6.9	-3.3	-6.6
flav	analog	20	22895377	-6.6	-5.7	-3.5	-6.6

flav	analog	21	69378634	-7.3	-6.6	-2.9	-6.1
flav	analog	22	11140124	-6.3	-5.9	-2.8	-5.7
flav	analog	23	70826	-6.4	-6.3	-3.0	-5.5
flav	analog	24	102137145	-6.0	-6.3	-3.1	-6.1
flav	analog	25	487559	-6.3	-6.8	-3.1	-5.7
flav	analog	26	122372409	-6.6	-6.0	-2.8	-5.2
flav	analog	27	122372411	-6.6	-6.8	-3.2	-5.5
flav	analog	28	73157749	-8.7	-8.2	-4.0	-7.7
flav	analog	29	23423196	-5.9	-5.2	-2.9	-5.1
flav	analog	30	5275540	-7.3	-5.5	-2.8	-5.4
flav	analog	31	122372410	-6.1	-5.4	-2.8	-5.3
flav	analog	32	5276405	-7.8	-8.3	-3.5	-6.3
flav	analog	33	489304	-7.7	-8.4	-4.1	-7.0
flav	analog	34	5270725	-7.4	-6.7	-3.8	-6.4
flav	analog	35	9831030	-6.3	-6.3	-2.8	-6.0
flav	analog	36	87339261	-8,0	-7.0	-3.3	-6.2
flav	analog	37	18688975	-6.5	-6.5	-3.0	-5.5
flav	analog	38	21617649	-6.5	-6.4	-2.9	-5.4
flav	analog	39	73157750	-8.2	-9.1	-3.8	-6.8
flav	analog	40	68372	-7.2	-6.9	-3.5	-6.0
flav	analog	41	5270720	-8.1	-7.9	-4.1	-7.2
flav	analog	42	5270721	-8.8	-8.7	-4.1	-6.9
flav	analog	43	24180539	-7.6	-7.3	-3.6	-6.7
flav	analog	44	24180556	-7.0	-7.0	-2.8	-6.5
flav	analog	45	10266206	-8.2	-8.3	-3.53	-6.4
flav	analog	46	71444368	-6.9	-6.1	-2,9	-5.5
flav	analog	47	101683334	-7.1	-7.1	-3.1	-6.1
flav	analog	48	3047894	-7.1	-7.0	-3.4	-6.4
flav	analog	49	21145076	-7.4	-8.3	-3.3	-6.5
flav	analog	51	140567676	-6.6	-5.7	-3.0	-5.3
flav	analog	52	10362392	-7.7	-6.9	-2.9	-5.9
flav	analog	53	476582	-8.5	-9.2	-4.1	-7.2
flav	analog	54	53893450	-8.0	-8.5	-3.5	-6.2
flav	analog	55	122372408	-6.9	-6.1	-2.9	-5.6
flav	analog	56	476581	-8.4	-8.5	-4.3	-7.1
flav	analog	57	101653699	-6.8	-7.1	-3.0	-5.8
flav	analog	58	15339776	-6.2	-6.4	-3.1	-6.1
flav	analog	59	87719879	-5.8	-7.4	-3.0	-5.9
flav	analog	60	129670024	-7.3	-6.2	-3.3	-5.7
flav	analog	61	137199552	-7.0	-6.0	-3.0	-5.5
flav	analog	62	151157187	-7.1	-6.2	-3.3	-5.4
flav	analog	63	69378634	-7.3	-6.3	-3.0	-5.3
flav	analog	64	129892599	-7.7	-6.7	-3.4	-5.9
flav	analog	65	44592636	-7.8	-7.8	-3.8	-6.8
flav	analog	66	129671851	-6.4	-5.7	-3.1	-5.5
flav	analog	67	102137145	-7.3	-7.3	-3.1	-6.2
flav	analog	68	101653698	-6.5	-6.5	-2.9	-6.0
flav	analog	69	129670024	-7.3	-6.0	-3.3	-5.7
flav	analog	70	122372413	-5.6	-5.9	-2.8	-4.8
flav	analog	71	122372409	-6.3	-6.7	-3.0	-5.3
flav	analog	72	12098458	-7.6	-8.0	-3.6	-7.0
flav	analog	73	122372410	-5.9	-5.9	-2.8	-5.1
flav	analog	74	10362392	-6.2	-6.4	-3.2	-5.6

flav	analog	75	122372408	-7.1	-6.3	-2.9	-5.5
flav	analog	76	132596498	-6.7	-7.2	-3.1	-6.3
flav	analog	77	53342735	-8.0	-8.1	-3.7	-7.5
flav	analog	78	137199552	-7.0	-6.3	-2.8	-5.3
flav	analog	79	101683334	-7.1	-7.1	-3.3	-6.3
flav	analog	80	10535192	-6.3	-6.0	-3.0	-5.7
(+)- CATHECIN							
flav	analog	81	9064	-8.7	-8.7	-3.8	-6.2
flav	analog	82	147299	-8.0	-8.7	-4.3	-7.9
flav	analog	83	5276454	-7.9	-8.7	-3.8	**-8.2**
flav	analog	84	471393	-9.0	-9.5	-4.0	-8.0
flav	analog	85	15689618	-8.8	-8.5	-4.2	-7.9
flav	analog	86	182659	-7.0	-9.1	-3.4	-6.2
flav	analog	87	474541	-9.6	-8.7	-4.4	-7.6
flav	analog	88	10789789	-8.2	-8.4	-4.0	-7.0
flav	analog	89	21626705	-7.5	-8.7	-4.1	-7.3
flav	analog	90	6324898	-7.9	-8.2	-3.9	-7.4
flav	analog	91	100947852	-7.8	-9.2	-3.7	-7.1
(-)- EPICATHECIN							
flav	analog	92	72276	-8.0	-9.0	-3.8	-6.5
flav	analog	93	107905	-7.8	**-10.6**	-4.0	-6.9
flav	analog	94	65064	-7.8	-8.1	-4.1	-7.0
flav	analog	95	76961592	-8.6	-8.3	-4.1	-7.6
flav	analog	96	76969982	-8.1	-8.7	-4.1	-7.2
flav	analog	97	14332898	-7.4	-7.7	-3.9	-6.5
flav	analog	98	124202088	-6.9	-8.5	-4.1	-6.6
flav	analog	99	124202087	-8.4	-7.4	-4.0	-6.8
flav	analog	100	476783	**-9.5**	-8.1	**-4.5**	-7.1

Table A.5. 7UJM (ER), 1R5K (ER), 4IFI (BRCA1) and 1N0W (BRCA2) receptors molecular docking drugs and Autodock Vina docking results.

MOLECULAR DOCKING FORMAT		**ChEMBL ACTIVITIES**			**RECEPTORS Docking Results (kcal/mol)**			
pdbqt	**sdf**	**Order**	**ChEMBL Code**	**Drug Name**	**7UJM**	**1R5K**	**4IFI**	**1N0W**
druq	dru	1	CHEMBL4650316	GIREDESTRANT	-7.8	-7.5	-3.9	-6.8
druq	dru	2	CHEMBL494753	ESTRONE SULFURIC ACID	-8.1	-7.7	-4.4	-7.2
druq	dru	3	CHEMBL3545211	SR16234	-8.6	-8.5	-4.0	-7.7
druq	dru	4	CHEMBL954	ENCLOMIPHENE	-5.9	-6.2	-3.1	-6.3
druq	dru	5	CHEMBL2355051	CLOMIPHENE	**-9,6**	-7.4	-3.7	-6.7
druq	dru	6	CHEMBL3581693	BRILANESTRANT	-7.5	-7.8	-4.2	-7.6
druq	dru	7	CHEMBL3545017	GTX-758	-9.5	-7.6	-3.6	-6.5
druq	dru	8	CHEMBL4475463	AMCENESTRANT	-7.4	-8.2	-4.3	-7.1
druq	dru	9	CHEMBL4650365	CAMIZESTRANT	-7.5	-8.1	-3.4	-6.8
druq	dru	11	CHEMBL135	ESTRADIOL	-9.2	-7.5	-3.9	-6.6
druq	dru	12	CHEMBL1511	ESTRADIOL VALERATE	-7.4	-7.3	-4.0	-7.3
druq	dru	13	CHEMBL1018	DIENESTROL	-6.3	-8.6	-3.6	-6.2
druq	dru	14	CHEMBL1201151	MESTRANOL	-8.1	-7.8	-4.5	-7.2
druq	dru	15	CHEMBL691	ETHINYL ESTRADIOL	-8.0	-7.6	-4.6	-7.4
druq	dru	16	CHEMBL83	TAMOXIFEN	-6.8	**-9.4**	-3.2	-6.2
druq	dru	17	CHEMBL1200973	ESTRADIOL CYPIONATE	-8.0	-8.5	**-4.8**	**-8.5**
druq	dru	18	CHEMBL4297509	ELACESTRANT	-6.2	-7.8	-3.1	-6.3
druq	dru	19	CHEMBL411	DIETHYLSTILBESTROL	-5.4	-6.1	-3.0	-5.6
druq	dru	20	CHEMBL1405	ESTRONE	-7.3	-6.8	-4.2	-7.0
druq	dru	21	CHEMBL1200430	ESTRADIOL ACETATE	-8.0	-7.9	-4.0	-7.3

APPENDIX B

GROMACS Tutorial/ Protein-Ligand Complex

source /usr/local/gromacs/bin/GMXRC
tar -zxvf charmm36_ljpme-jul2022.ff.tgz
gmx_mpi pdb2gmx -f REC.pdb -ignh
8 (CHARMM27)
1 (TIP3P)
gmx_mpi editconf -f LIG.pdb -o LIG.gro
gedit conf.gro LIG.gro
*(Copy content from 3rd line of lig.gro to the conf.gro file up to the 2nd last line)
*(Check the column number from where the lig.gro data ends (x) in conf.gro and replace the value in 2nd line by x-3)
*(Open file in chimera to check ligand and receptor)

-----EDIT THE FOLLOWING in topol.top -----
gedit topol.top
(add

; Include ligand topology
#include "LIG.itp"

below- Include forcefield parameters
#include "amberGS.ff/forcefield.itp")

AT THE BOTTOM OF THE SAME FILE PERFORM FOLLOWING CHANGES
(add LIG 1
align exactly below-
Protein_chain_E 1)

-----EDIT THE FOLLOWING in lig.itp -----
gedit lig.itp
[moleculetype]
; Name nrexcl
lig_gmx2 3
TO

[moleculetype]

; Name nrexcl

LIG 3

(in certain cases this will already be LIG 3 so for such case no change is needed)

gmx_mpi editconf -f conf.gro -d 2.0 -bt triclinic -o box.gro

gmx_mpi solvate -cp box.gro -cs spc216.gro -p topol.top -o box_sol.gro

gmx grompp -f ions.mdp -c box_sol.gro -p topol.top -o ION.tpr

(OR)

gmx_mpi grompp -f ions.mdp -c box_sol.gro -maxwarn 2 -p topol.top -o ION.tpr

gmx_mpi genion -s ION.tpr -p topol.top -conc 0.1 -neutral -o box_sol_ion.gro

15

gmx grompp -f EM.mdp -c box_sol_ion.gro -p topol.top -o EM.tpr (OR)

gmx_mpi grompp -f EM.mdp -c box_sol_ion.gro -maxwarn 2 -p topol.top -o EM.tpr

gmx mdrun -v -deffnm EM

```
; LINES STARTING WITH ';' ARE COMMENTS
title           = Minimization  ; Title of run

; Parameters describing what to do, when to stop and what to save
integrator      = steep     ; Algorithm (steep = steepest descent minimization)
emtol           = 1000.0    ; Stop minimization when the maximum force < 10.0 kJ/mol
emstep          = 0.01      ; Energy step size
nsteps          = 50000     ; Maximum number of (minimization) steps to perform

; Parameters describing how to find the neighbors of each atom and how to calculate the interactions
nstlist         = 1             ; Frequency to update the neighbor list and long range forces
cutoff-scheme   = Verlet
ns_type         = grid          ; Method to determine neighbor list (simple, grid)
rlist           = 1.2           ; Cut-off for making neighbor list (short range forces)
coulombtype     = PME           ; Treatment of long range electrostatic interactions
rcoulomb        = 1.2           ; long range electrostatic cut-off
vdwtype         = cutoff
vdw-modifier    = force-switch
rvdw-switch     = 1.0
rvdw            = 1.2           ; long range Van der Waals cut-off
pbc             = xyz           ; Periodic Boundary Conditions
DispCorr        = no
```

Figure B.1. Protein-Ligand Complex EM.mdp file parameters.

gedit nvt.mdp (This file is already modified)

Now make index files

gmx_mpi make_ndx -f LIG.gro -o index_LIG.ndx

> 0 & ! a H*

> q

gmx_mpi genrestr -f LIG.gro -n index_LIG.ndx -o posre_LIG.itp -fc 1000 1000 1000

> select group "3"

Now, open topol.top file

at the end of the document

after

"; Include Position restraint file

#ifdef POSRES

#include "posre.itp"

#endif

"Here"

add this

; Ligand position restraints

#ifdef POSRES

#include "posre_LIG.itp"

#endif

Again, Make other Index file for System

gmx_mpi make_ndx -f EM.gro -o index.ndx

> 1 | 13

> q

-----[NVT]-----

gedit NVT.mdp (This file is already modified)

gmx_mpi grompp -f NVT.mdp -c EM.gro -r EM.gro -p topol.top -n index.ndx -maxwarn 2 -o NVT.tpr

gmx mdrun -deffnm NVT

```
title                    = Protein-ligand complex NVT equilibration
define                   = -DPOSRES  ; position restrain the protein and ligand
; Run parameters
integrator               = md        ; leap-frog integrator
nsteps                   = 50000     ; 2 * 50000 = 100 ps
dt                       = 0.002     ; 2 fs
; Output control
nstenergy                = 500   ; save energies every 1.0 ps
nstlog                   = 500   ; update log file every 1.0 ps
nstxout-compressed       = 500   ; save coordinates every 1.0 ps
; Bond parameters
continuation             = no        ; first dynamics run
constraint_algorithm     = lincs     ; holonomic constraints
constraints              = h-bonds   ; bonds to H are constrained
lincs_iter               = 1         ; accuracy of LINCS
lincs_order              = 4         ; also related to accuracy
; Neighbor searching and vdW
cutoff-scheme            = Verlet
ns_type                  = grid      ; search neighboring grid cells
nstlist                  = 20        ; largely irrelevant with Verlet
rlist                    = 1.2
vdwtype                  = cutoff
vdw-modifier             = force-switch
rvdw-switch              = 1.0
rvdw                     = 1.2       ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype              = PME       ; Particle Mesh Ewald for long-range electrostatics
rcoulomb                 = 1.2       ; short-range electrostatic cutoff (in nm)
pme_order                = 4         ; cubic interpolation
fourierspacing           = 0.16      ; grid spacing for FFT
; Temperature coupling
tcoupl                   = V-rescale                     ; modified Berendsen thermostat
tc-grps                  = Protein_LIG Water_and_ions    ; two coupling groups - more accurate
tau_t                    = 0.1   0.1                     ; time constant, in ps
ref_t                    = 300   300                     ; reference temperature, one for each group, in K
; Pressure coupling
pcoupl                   = no        ; no pressure coupling in NVT
; Periodic boundary conditions
pbc                      = xyz       ; 3-D PBC
; Dispersion correction is not used for proteins with the C36 additive FF
DispCorr                 = no
; Velocity generation
gen_vel                  = yes       ; assign velocities from Maxwell distribution
gen_temp                 = 300       ; temperature for Maxwell distribution
gen seed                 = -1        ; generate a random seed
```

Figure B.2. Protein-Ligand Complex NVT.mdp file parameters.

-----[NPT]-----

gedit NPT.mdp (This file is already modified)

gmx_mpi grompp -f NPT.mdp -c NVT.gro -r NVT.gro -p topol.top -n index.ndx -maxwarn 2 -o NPT.tpr

gmx mdrun -deffnm NPT

```
title                    = Protein-ligand complex NPT equilibration
define                   = -DPOSRES  ; position restrain the protein and ligand
; Run parameters
integrator               = md        ; leap-frog integrator
nsteps                   = 50000     ; 2 * 50000 = 100 ps
dt                       = 0.002     ; 2 fs
; Output control
nstenergy                = 500       ; save energies every 1.0 ps
nstlog                   = 500       ; update log file every 1.0 ps
nstxout-compressed       = 500       ; save coordinates every 1.0 ps
; Bond parameters
continuation             = yes       ; continuing from NVT
constraint_algorithm     = lincs     ; holonomic constraints
constraints              = h-bonds   ; bonds to H are constrained
lincs_iter               = 1         ; accuracy of LINCS
lincs_order              = 4         ; also related to accuracy
; Neighbor searching and vdW
cutoff-scheme            = Verlet
ns_type                  = grid      ; search neighboring grid cells
nstlist                  = 20        ; largely irrelevant with Verlet
rlist                    = 1.2
vdwtype                  = cutoff
vdw-modifier             = force-switch
rvdw-switch              = 1.0
rvdw                     = 1.2       ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype              = PME       ; Particle Mesh Ewald for long-range electrostatics
rcoulomb                 = 1.2
pme_order                = 4         ; cubic interpolation
fourierspacing           = 0.16      ; grid spacing for FFT
; Temperature coupling
tcoupl                   = V-rescale                     ; modified Berendsen thermostat
tc-grps                  = Protein_LIG Water_and_ions    ; two coupling groups - more accurate
tau_t                    = 0.1   0.1                     ; time constant, in ps
ref_t                    = 300   300                     ; reference temperature, one for each group, in K
; Pressure coupling
pcoupl                   = Berendsen                     ; pressure coupling is on for NPT
pcoupltype               = isotropic                     ; uniform scaling of box vectors
tau_p                    = 2.0                           ; time constant, in ps
ref_p                    = 1.0                           ; reference pressure, in bar
compressibility          = 4.5e-5                        ; isothermal compressibility of water, bar^-1
refcoord_scaling         = com
; Periodic boundary conditions
pbc                      = xyz       ; 3-D PBC
; Dispersion correction is not used for proteins with the C36 additive FF
DispCorr                 = no
; Velocity generation
gen_vel                  = no        ; velocity generation off after NVT
```

Figure B.3. Protein-Ligand Complex NPT.mdp file parameters.

-----[FINAL MD RUN/PRODUCTION]-----

gedit NPT.mdp (Change MD RUN TIME as per your need)

gmx_mpi grompp -f MD.mdp -c NPT.gro -t NPT.cpt -p topol.top -n index.ndx -maxwarn 2 -o MD.tpr

gmx mdrun -deffnm MD

```
title                    = Protein-ligand complex MD simulation
; Run parameters
integrator               = md        ; leap-frog integrator
nsteps                   = 25000000  ; 2 * 25000000 = 50000 ps (50 ns)
dt                       = 0.002     ; 2 fs
; Output control
nstenergy                = 5000      ; save energies every 10.0 ps
nstlog                   = 5000      ; update log file every 10.0 ps
nstxout-compressed       = 5000      ; save coordinates every 10.0 ps
; Bond parameters
continuation             = yes       ; continuing from NPT
constraint_algorithm     = lincs     ; holonomic constraints
constraints              = h-bonds   ; bonds to H are constrained
lincs_iter               = 1         ; accuracy of LINCS
lincs_order              = 4         ; also related to accuracy
; Neighbor searching and vdW
cutoff-scheme            = Verlet
ns_type                  = grid      ; search neighboring grid cells
nstlist                  = 20        ; largely irrelevant with Verlet
rlist                    = 1.2
vdwtype                  = cutoff
vdw-modifier             = force-switch
rvdw-switch              = 1.0
rvdw                     = 1.2       ; short-range van der Waals cutoff (in nm)
; Electrostatics
coulombtype              = PME       ; Particle Mesh Ewald for long-range electrostatics
rcoulomb                 = 1.2
pme_order                = 4         ; cubic interpolation
fourierspacing           = 0.16      ; grid spacing for FFT
; Temperature coupling
tcoupl                   = V-rescale                     ; modified Berendsen thermostat
tc-grps                  = Protein_LIG Water_and_ions    ; two coupling groups - more accurate
tau_t                    = 0.1   0.1                     ; time constant, in ps
ref_t                    = 300   300                     ; reference temperature, one for each group, in K
; Pressure coupling
pcoupl                   = Parrinello-Rahman             ; pressure coupling is on for NPT
pcoupltype               = isotropic                     ; uniform scaling of box vectors
tau_p                    = 2.0                           ; time constant, in ps
ref_p                    = 1.0                           ; reference pressure, in bar
compressibility          = 4.5e-5                        ; isothermal compressibility of water, bar^-1
; Periodic boundary conditions
pbc                      = xyz       ; 3-D PBC
; Dispersion correction is not used for proteins with the C36 additive FF
DispCorr                 = no
; Velocity generation
gen vel                  = no        ; continuing from NPT equilibration
```

Figure B.4. Protein-Ligand Complex MD.mdp file parameters.