

Table of Contents

Table of Contents	vii
List of Figures	xiv
List of Tables	xxix
List of Abbreviations	xxxii
List of Symbols	xxxvii
PREFACE	xxxix
Chapter 1: Introduction and Objectives	1
1.1 Introduction	1
1.2 Objectives	6
Chapter 2: Review of Literature	7
2.1 Natural compound	7
2.2 Streptomyces	8
2.3 Life cycle of <i>Streptomyces</i>	9
2.4 Therapeutic Application of <i>Streptomyces</i>	12
2.5 Industrial Application of <i>Streptomyces</i>	15
2.6 Agricultural Application of <i>Streptomyces</i>	17
2.7 <i>Streptomyces clavuligerus</i> and <i>Streptomyces fragilis</i> and their secondary metabolites production.	19
2.8 Anthracycline and its mechanism of action	20
2.9 Doxorubicin and Epirubicin	22
2.10 Microbial Production of Epirubicin	24
2.11 Submerged Fermentation and Batch Fermentation	26
2.12 Optimization Approach for Enhancement of Bioactive Compound	27
2.12.1 Optimization Using Response Surface Methodology Approach	27
2.12.2 Optimization using Artificial Neural Network (ANN)	29
2.12.3 Optimization Using Machine Learning Approach	30

2.13 Receptor Tyrosine Kinase	32
2.14 Protein Kinase Inhibitory Assay	33
Chapter 3: Preliminary screening of metabolites	35
3.1 Introduction	35
3.2 Materials and Methods	37
3.2.1 Retrieval and processing of ligand database	37
3.2.2 Retrieval and Processing of Target	37
3.2.3 ADME analysis of ligand library	38
3.2.4 Molecular Docking Simulations	39
3.2.5 Molecular Dynamics Simulations	39
3.3 Results and Discussion	40
3.3.1 ADMET Properties Predictions	40
3.3.2 Molecular Docking Study of triple mutant EGFR ^{CSTMLR}	42
3.3.3 Molecular Dynamics Studies of the mutant protein	49
3.3.4 MM/GBSA calculations and per-residue free energy decomposition	54
3.4 Conclusions	59
Chapter 4: Production of Tyrosine Kinase Inhibitors from <i>Streptomyces clavuligerus</i>	61
4.1 Introduction	61
4.2 Materials and Methods	63
4.2.1 Cancer Cell Line and <i>Streptomyces 85E</i> strain	63
4.2.2 Procurement and maintenance of <i>Streptomyces clavuligerus</i>	63
4.2.3 Colony separation and microscopic examination	63
4.2.4 Fermentation-based production of bioactive compounds	64
4.2.5 Microbial Growth and Metabolite Production	64
4.2.6 Downstream process of secondary metabolites	64
4.2.7 Protein kinase inhibitory assay (<i>Streptomyces 85E</i> assay)	65

4.2.8 Free radical scavenging activity	66
4.2.9 Cytotoxicity Assay	66
4.2.10 HR-LCMS for Identification of Secondary Metabolites	68
4.2.11 Molecular Docking of metabolites with RTKs.	68
4.2.12 Molecular Dynamics Simulation	69
4.3 Results and Discussions	70
4.3.1 Colony separation and microscopic examination	70
4.3.2 Microbial Growth and Metabolite Production	71
4.3.3 <i>Streptomyces 85E</i> assay (Kinase inhibition assay)	72
4.3.4 The free radical scavenging activity	73
4.3.5 Cytotoxicity analysis	74
4.3.6 HR-LCMS analysis	75
4.3.7 Molecular Docking	80
4.3.8 Molecular dynamics simulation	87
4.3.9 Protein-Ligand Interaction	94
4.3.10 MM/GBSA calculation	99
4.3.11 Per-residue decomposition	101
4.4 Conclusions	103
Chapter 5: Production of Tyrosine Kinase Inhibitors from	
<i>Streptomyces fragilis</i>	105
5.1 Introduction	105
5.2 Materials and methods	107
5.2.1 Isolation and maintenance of Actinomycetes	107
5.2.2 Cultural and Morphological Properties	107
5.2.3 16S rRNA Sequencing and Phylogenetic Analysis	108
5.2.4 Fermentation	108
5.2.4.1 Fermentation media	108

5.2.4.2 Fermentation condition:	109
5.2.5 Production and Downstream Process of Secondary Metabolites	109
5.2.6 Protein kinase inhibitory Assay (<i>Streptomyces 85E</i> assay)	110
5.2.7 Free radical scavenging activity	110
5.2.8 Cytotoxicity Activity by MTT Assay	111
5.2.9 Identification of secondary metabolites through HR-LCMS	112
5.2.10 Molecular docking	113
5.2.11 Molecular Dynamics Simulation	114
5.3 Results and Discussion	115
5.3.1 Strain morphology and its identification	115
5.3.2 Protein Kinase Inhibitory Assay	116
5.3.3 Antioxidant Potential	117
5.3.4 MTT Assay	118
5.3.5 Identification of Secondary Metabolites	119
5.3.6 Molecular Docking	123
5.3.7 Molecular Dynamics Simulation	128
5.3.8 Protein-Ligand Interaction	132
5.3.9 MM/GBSA calculation	136
5.3.10 Per-residue Decomposition	137
5.4 Conclusion	138
Chapter 6: Glycosidic Anthracycline Production, Optimization, Purification, and Characterization from <i>Streptomyces fragilis</i>	140
6.1. Introduction	140
6.2. Materials and methods	142
6.2.1. Microorganism and Fermentation	142
6.2.2. Microbial Growth and Metabolite Production	142
6.2.3. Optimization process for bioactive metabolites production	143

6.2.3.1. Effect of different carbon sources on metabolite production	143
6.2.3.2. Effect of different nitrogen sources on metabolite production	143
6.2.3.3. Different techniques for production process optimization	144
6.2.3.4 Production Optimization using RSM	144
6.2.3.5 Process optimization using Machine Learning methodology	146
6.2.3.6 BFGS (BFGS Quasi-Newton Algorithm)	147
6.2.3.7 GRNN (General Regression Neural Network)	148
6.2.3.8 LM (Levenberg-Marquardt Algorithm)	148
6.2.3.9 SCG (Scaled Conjugate Gradient Algorithm)	148
6.2.3.10 SVM (Support Vector Machine)	149
6.2.4 Validation of the experimentally derived model	150
6.2.5 Downstream process and detection of Anthracycline Epirubicin	150
6.2.6 Structural Elucidation and Thermal Characteristics of Epirubicin	151
6.2.6.1. FTIR Spectral Analysis	151
6.2.6.2. ¹ H and ¹³ C Nuclear Magnetic Resonance	152
6.2.6.3. High-Resolution Mass Spectroscopy	152
6.2.6.4. X-ray diffraction (XRD) analysis	152
6.2.6.5. Differential Scanning Calorimetry (DSC)	152
6.2.6.6. Thermogravimetric analysis (TGA)	153
6.3. Results and Discussions	153
6.3.1. Microbial Growth and Metabolite Production	153
6.3.2. Production of secondary metabolites using different carbon sources	154

6.3.3. Effect of Nitrogen on the production of secondary metabolites	156
6.3.4. Maximizing the production of metabolites using the RSM technique.	157
6.3.5. Prediction of metabolite production using machine learning.	160
6.3.6. Validation of the model by experiments	169
6.3.7. Anthracycline detection and separation	169
6.3.8. Structural Elucidation and Thermal Characteristics of Epirubicin	170
6.3.8.1. FTIR Spectral Analysis	170
6.3.8.2. ¹ H and ¹³ C Nuclear Magnetic Resonance	172
6.3.8.3. High-Resolution Mass Spectroscopy	173
6.3.8.4. X-ray diffraction (XRD) analysis	173
6.3.8.5. Differential Scanning Calorimetry (DSC)	174
6.3.8.6. Thermogravimetric analysis (TGA)	174
6.4. Conclusions	175
Chapter 7: Therapeutic potential of Bioactive compounds	177
7.1. Introduction	177
7.2. Materials and Methods	179
7.2.1. Free radical scavenging activity	179
7.2.2. Cell Line and Culture Conditions	179
7.2.3. Cell Viability Assay	179
7.2.4. Total RNA Extraction, Qualitative and Quantitative Analysis	180
7.2.5. Illumina 2x150 PE library preparation	180
7.2.6. Quality Control, Cluster Generation, and Sequencing	181
7.2.7. Library Validation	181
7.2.8. Functional Gene Analysis	183
7.3. Results and Discussions	183
7.3.1. Free Radical Scavenging Potential	183

7.3.2. Cytotoxicity study	184
7.3.3. Transcriptomics Analysis	184
7.3.3.1. Differentially Expressed Genes Analysis	184
7.3.3.2. Gene Ontology and KEGG Pathway Analysis	185
7.3.3.3. Gene Pathway Enrichment Analysis	185
7.3.3.4. Protein-Protein Interaction Network	189
7.3.3.5. Analysis of Core 7 Gene	192
7.3.3.6. Correlation analysis of MCM5 with core 7 gene	201
7.4. Conclusion	204
Chapter 8: Summary and Conclusions	206
Chapter 9: References	211

List of Figures

- Figure 2.1:** The life cycle of *Streptomyces* at different stages. 11
- Figure 2.2:** Mechanisms of anticancer activity of anthracycline drugs associated with inhibition of DNA topoisomerase II and overproduction of reactive oxygen species and cell apoptosis. 22
- Figure 2.3:** 2D image of Doxorubicin and Epirubicin showing the major difference at the C-4 position. 23
- Figure 2.4:** Protein Kinase Inhibitory Assay showing the growth pattern of *Streptomyces 85E* strain in the presence and absence of kinase inhibitors. 34
- Figure 3.1:** ADMET profile of 6524 compounds from the StreptomeDB 3.0 database. 41
- Figure 3.2.1:** 2D plot showing the network of the interaction of all compounds representing a: C_42, and b: C_4299 with active site residues of EGFR^(L858R/T790M/C797S) mutant during molecular docking study. 45
- Figure 3.2.2:** 2D plot showing the network of the interaction of all compounds representing a: C_4300, b: C_5702, and c: CH7, with active site residues of EGFR^(L858R/T790M/C797S) mutant during molecular docking study. 46
- Figure 3.2.3:** 2D plot showing the network of the interaction of all compounds representing a: C_5024, and b: Osi with active site residues of EGFR^(L858R/T790M/C797S) mutant during molecular docking study. 47
- Figure 3.3:** RMSD profile of triple mutant EGFR(L858R/T790M/C797S) protein complex with ligands during MD simulation of 1,000 ns... 49

Figure 3.4:	RMSF profile of triple mutant EGFR(L858R/T790M/C797S) protein's interacting residues with ligands during MD simulation of 1,000 ns.....	50
Figure 3.5:	The radius of gyration of ligands during MD simulation of triple mutant EGFR(L858R/T790M/C797S) protein-receptor complex. ..	51
Figure 3.6:	Solvent accessible surface area (SASA) of five-hit ligands, reference drug CH7233163, and co-crystallized ligand Osimertinib during MD simulation of 1,000 ns.....	51
Figure 3.7:	Hydrogen bonds formed between protein EGFR ^{L858R/T790M/C797S} and ligands during MD simulation of 1,000 ns.....	52
Figure 3.8:	Amino acid residues contacted with ligands (A: C_42, B: C_4299, C: C_4300, D: C_5024, E: C_5702, F: CH7 and G: OSI) through molecular interaction (Hydrogen-bond, Hydrophobic, ionic, and water bridges) during MD simulation of 1,000 ns.	53
Figure 3.9.1:	2D plot represents the ligands representing a: C_42, and b: C_5024 with interacting residues of EGFR ^{L858R/T790M/C797S} and their interaction percentage during MD simulation of 1,000 ns.	55
Figure 3.9.2:	2D plot represents the ligands representing a: C_4299, b: CH7, and c: C_5702 with interacting residues of EGFR ^{L858R/T790M/C797S} and their interaction percentage during MD simulation of 1,000 ns.	56
Figure 3.9.3:	2D plot represents the ligands representing a: C_4300, and b: Osi with interacting residues of EGFR ^{L858R/T790M/C797S} and their interaction percentage during MD simulation of 1,000 ns.....	57

Figure 3.10:	Per-residue-energy-decomposition analysis of all ligands against EGFR ^{L858R/T790M/C797S} structure.	59
Figure 4.1:	Plate A represents the cultured plate of <i>Streptomyces clavuligerus</i> (MTCC 1142), and plate B is the microscopic image of <i>Streptomyces clavuligerus</i> at 100X.....	71
Figure 4.2:	The growth curve (black line) and metabolite production curves (red line) for <i>S. clavuligerus</i> represent the biomass accumulation and secondary metabolite production over a 72-hour period.	72
Figure 4.3:	Plate A showed 18 mm diameter whitish hazy zones due to <i>Streptomyces clavuligerus</i> extracts, while plate B (positive control) showed a 0.66 mm diameter thin and hazy line from doxorubicin. The negative control in plate C, represented by DMSO, did not show any zone formation. Whitish hazy zones suggest the existence of Protein Kinase Inhibitors, and clear zones denote complete bacterial inhibition.....	73
Figure 4.4:	The free radical scavenging activity of various concentrations of organic and aqueous extract of <i>Streptomyces clavuligerus</i>	74
Figure 4.5:	Percentage growth of MCF-7, Hop-62, SiHa, and PC-3 cancerous cell lines (cultured at 37 °C temperature, 5:95 of CO ₂ : air composition, and 100% RH for 24hr) with varying concentrations of the organic <i>Streptomyces clavuligerus</i> extracts.....	75
Figure 4.6:	HRLCMS chromatogram of aqueous and organic extract of <i>Streptomyces clavuligerus</i> showing a great resolution peak.	76

Figure 4.7.1: 3D molecular docking interactions of top ligands and standard drug Pazopanib with FGFR Receptor Tyrosine Kinases.....	85
Figure 4.7.2: 3D molecular docking interactions of top ligands and standard drug Pazopanib with PDGFR Receptor Tyrosine Kinases.....	86
Figure 4.7.3: 3D molecular docking interactions of top ligands and standard drug Pazopanib with VEGFR Receptor Tyrosine Kinases.....	87
Figure 4.8: Root Mean Square Deviation (RMSD) profiles of the respective three targets RTKs in Apo state, in complex with ligand and Std (protein with Pazopanib) during Molecular Dynamics.	88
Figure 4.9: Root Mean Square Fluctuations (RMSF) profiles of the respective three targets RTKs in Apo state, in complex with ligand and Std (protein with Pazopanib) during Molecular Dynamics.	89
Figure 4.10: Solvent Accessible Surface Area (SASA) profiles of the respective three targets RTKs in complex with ligand and Std (protein with Pazopanib) during Molecular Dynamics.	91
Figure 4.11: Radius of Gyration (Rg) profiles of the respective three targets RTKs in complex with ligand and Std (protein with Pazopanib) during Molecular Dynamics.....	92
Figure 4.12: Number of Hydrogen Bond formation profiles of three targets RTKs in complex with ligand and Std (protein with Pazopanib) during Molecular Dynamics.....	94
Figure 4.13.1: The 2D plot represents the protein-ligands interaction analysis of ligands and standard drug Pazopanib with interaction residue of	

FGFR and their interaction percentage during an MD simulation of 200 ns	95
Figure 4.13.2: The 2D plot represents the protein-ligands interaction analysis of ligands and standard drug Pazopanib with interaction residue of VEGFR and their interaction percentage during an MD simulation of 200 ns	96
Figure 4.13.3. The 2D plot represents the protein-ligands interaction analysis of ligands and standard drug Pazopanib with interaction residue of PDGFR and their interaction percentage during an MD simulation of 200 ns	97
Figure 4.14: Energy decomposition analysis within the protein-ligand complex structure during MD simulation of 200 ns.....	102
Figure 5.1: The strain characterization. (A) Cultured plate of isolated actinomycetes and its microscopic 100x zoom image, and (B) Phylogenetic relationship tree of the actinomycetes. The isolated strain is highlighted with a red boundary.....	115
Figure 5.2: Plates labelled as aM:1, aM:2, aM:3, and aM:4 are <i>Streptomyces</i> 's aqueous extracts, whereas oM:1, oM:2, oM:3, and oM:4 are organic extracts in respective media. Dimethyl sulfoxide (DMSO) is used as a negative control, and std as doxorubicin, a positive control. Whitish hazy zones indicate the presence of Protein Kinase Inhibitors in the extract, and clear zones indicate complete inhibition of bacteria.	116

Figure 5.3:	free radical scavenging activity of epirubicin and organic (S_F_Org) and aqueous (S_F_Aq) extract of <i>Streptomyces fragilis</i>	117
Figure 5.4:	Percentage growth of MCF-7, Hop-62, SiHa, and PC-3 cancerous cell lines (cultured at 37 °C temperature, 5:95 of CO ₂ : air composition, and 100% RH for 24 hours) with varying concentrations of the organic <i>Streptomyces fragilis</i> extracts.	118
Figure 5.5:	HRLCMS chromatogram of aqueous and organic extract of <i>Streptomyces fragilis</i> showing a great resolution peak.	120
Figure 5.6:	2D Molecular docking interactions of epirubicin with various target proteins, represented as a.1: BCL2, a.2: BCR-ABL, a.3: CDK4, a.4: EGFR, a.5: JAK2, a.6: PDGFR, a.7: PI3K.	126
Figure 5.7:	Molecular docking interactions of doxorubicin with various target proteins, represented as b.1: BCL2, b.2: BCR-ABL, b.3: CDK4, b.4: EGFR, b.5: JAK2, b.6: PDGFR, b.7: PI3K.	127
Figure 5.8:	RMSD, rGyr, SASA, and Number of Hydrogen Bond formation profiles of respective ten targets in Apo state, in Epi (protein with epirubicin), and Std (protein with doxorubicin) during Molecular Dynamics.....	130
Figure 5.9:	Root Mean Square Fluctuation (RMSF) in the amino acid residue of respective target proteins with ligand during Molecular Dynamics simulation of 200 ns.....	131
Figure 5.10:	The 2D plot showcases the interaction patterns and corresponding percentages of epirubicin with protein kinase residues during a	

200 ns molecular dynamics (MD) simulation. The figures label represents specific target proteins as BCL2 (a.1), BCR-ABL (a.2), CDK4 (a.3), EGFR (a.4), JAK2 (a.5), PDGFR (a.6), and PI3K (a.7). These plots provide valuable insights into the binding interactions between epirubicin and the target proteins, offering potential avenues for targeted drug design and optimization. 134

Figure 5.11: The 2D plot represents the interaction of doxorubicin with protein kinase residues and their corresponding interaction percentages during a 200 ns molecular dynamics (MD) simulation. The figures label represents specific target proteins as BCL2 (a.1), BCR-ABL (a.2), CDK4 (a.3), EGFR (a.4), JAK2 (a.5), PDGFR (a.6), and PI3K (a.7). These 2D plots provide insights into the binding dynamics of doxorubicin with protein kinases. 135

Figure 5.12: Residues and their energy decomposition within a protein-ligand complex structure. 137

Figure 6.1: The growth curve (black line) and metabolite production curves (red line) for *S. fragilis* represent the biomass accumulation and secondary metabolite production over a 72-hour period. 155

Figure 6.2: Effect of different carbon sources on metabolite production using *S. fragilis*, along with pH 7, agitation rate 200, and temperature 30 °C. 156

Figure 6.3: Effect of different nitrogen sources on metabolite production using *S. fragilis*, along with pH 7, agitation rate 200, and temperature 30 °C. 157

Figure 6.4: Response surface plots of the quadratic model depicting the interactive effects between the process parameters. Plots showing interaction between (a) Molasses (%) and Agitation Rate (RPM) (b) Agitation Rate (RPM) and temperature (°C) (c) Molasses (%) and pH (d) Agitation Rate (RPM) and temperature (°C) (e) Agitation Rate (%) and pH (f) temperature (°C) and pH.160

Figure 6.5: This figure presents a comprehensive performance evaluation of the SVM (Support Vector Machine) model. Section A displays a scatter plot of predicted versus experimental values with a fit line, demonstrating a strong alignment between predicted and actual values. Section B shows a plot of residuals versus predicted values, indicating the distribution of errors relative to the predicted values. Section C presents a plot of residuals versus actual values, further illustrating the error distribution across the range of actual values. These visualizations confirm the SVM model's strong predictive accuracy and consistency.164

Figure 6.6: This figure presents a comprehensive performance evaluation of the GRNN (General Regression Neural Network) model. A display a scatter plot of predicted versus experimental values with a fit line, demonstrating a high degree of alignment between predicted and actual values. B shows a plot of residuals versus actual values, indicating the distribution of errors across the range of actual values. C presents a plot of residuals versus predicted values, further illustrating the error distribution relative to the

predicted values. These visualizations confirm the GRNN model's strong predictive accuracy and consistency..... 165

Figure 6.7: This figure presents a comprehensive performance evaluation of the BFGS (Broyden-Fletcher-Goldfarb-Shanno) model. A displays scatter plots of predicted versus actual target values for training ($R=0.99779$), validation ($R=0.99494$), test ($R=0.88105$), and overall datasets ($R=0.97651$), each with corresponding linear fit lines, indicating varying predictive accuracy across datasets. B shows the mean squared error (MSE) plot over 54 epochs, highlighting the best validation performance achieved at epoch 48 with a value of 0.010983. C illustrates the predicted versus experimental production values, further validating the model's predictive accuracy. 166

Figure 6.8: This figure presents a comprehensive performance evaluation of the SCG (Scaled Conjugate Gradient) model. A displays scatter plots of predicted versus actual target values for training ($R=0.97417$), validation ($R=0.99642$), test ($R=0.74653$), and overall datasets ($R=0.97108$), each with corresponding linear fit lines, indicating varying predictive accuracy across datasets. B shows the mean squared error (MSE) plot over 32 epochs, highlighting the best validation performance achieved at epoch 26 with a value of 0.005651. C illustrates the predicted versus experimental production values, further validating the model's predictive accuracy. 167

Figure 6.9:	This figure presents a performance evaluation of the LM (Levenberg-Marquardt) model. A displays scatter plots of predicted versus actual target values for training (R=0.99994), validation (R=0.9997), test (R=0.99243), and overall datasets (R=0.99593), each with corresponding linear fit lines, indicating high predictive accuracy. B shows the mean squared error (MSE) plot over epochs, highlighting the best validation performance achieved at epoch 1 with a value of 0.00031272. C illustrates the predicted versus experimental values, further validating the model's predictive accuracy.....	168
Figure 6.10:	HPLC chromatogram of the downstream part of <i>Streptomyces fragilis</i> extract (Black; Figure. 6.1.A) and pure epirubicin drug (multicolor; Figure. 6.1.B). The peak at RT of ~5 min indicates the presence of epirubicin in the extract.	169
Figure 6.11:	HPLC calibration curve of epirubicin.	170
Figure 6.12:	FT-IR spectra depicting a potential band of epirubicin (Blue) and the downstream part of <i>Streptomyces fragilis</i> extract (Red)	171
Figure 6.13:	Structural elucidation of epirubicin by ¹ H NMR (A) and ¹³ C NMR (B)	172
Figure 6.14:	HR-MS spectra of epirubicin produced by <i>S. fragilis</i>	173
Figure 6.15:	X-Ray diffraction spectra of epirubicin	173
Figure 6.16:	Differential scanning calorimetry thermogram analysis of epirubicin.....	174

Figure 6.17: Thermogravimetric analysis of epirubicin at 25°C to 600°C with a nitrogen flow rate of 100 mL/min.....	175
Figure 7.1. Free radical scavenging activity of epirubicin	183
Figure 7.2: Percentage viability of A549 cell line (Human lung cancer) cultured at 37 °C temperature, 5% CO ₂ for 24 hr., treated with varying concentrations of epirubicin.	184
Figure 7.3: Volcano plot showing differential expression of genes in A549 treated with Epirubicin. The x-axis illustrates log ₂ fold change of gene expression, while the y-axis illustrates -log ₁₀ p-value showing statistical significance of change in expression. Upregulated genes are visualized in red, downregulated genes in blue, and non-significant changes in yellow.....	187
Figure 7.4: Gene ontology analysis of epirubicin treated A549 cell representing upregulated, and downregulated genes. (Left) Bar graph of GO biological processes showing the count of proteins involved in different biological processes, including signal transduction, cell proliferation, and DNA damage response. (Right) Bar graph of GO cellular components. Majorly associated important proteins are the extracellular matrix, the external encapsulating structure, the DNA replication preinitiation complex, and the CMG complex. These suggest their possible role in structural and replication-related functions within the cell.	188

Figure 7.5: Gene Pathway Enrichment Analysis of Differentially Expressed Genes. The pathways include transcriptional misregulation in cancer, Ras signaling pathway, p53 signaling pathway, DNA replication, central carbon metabolism in cancer, cAMP signaling pathway, calcium signaling pathway, axon guidance, apoptosis, and Apelin signaling pathway. The count on the x-axis represents the number of proteins associated with each pathway, highlighting the most enriched pathways in LUSC progression.....189

Figure 7.6: Protein-protein interaction network of 597 genes, illustrating the connections between various proteins. The enlarged inset highlights a highly interconnected subnetwork consisting of seven key proteins, including MYC, MCM7, UHRF1, MCM3, MCM6, MCM5, and ECT2, suggesting their potential collaborative roles in tumor progression and survival.191

Figure 7.7: (A) Kaplan-Meier survival curve from TCGA demonstrating that patients with LUSC and high MYC expression levels (n=124) have a better prognosis compared to those patients that express MYC at low/medium levels (n=370). (B) Box plot from GEPIA comparing the expression of MYC in primary tumor samples (n=503) and normal samples (n=52). The gene presents higher expression in the primary tumor (TPM), with a certain tendency for the data to indicate the involvement of MYC in the progression of tumors.....194

Figure 7.8: Kaplan-Meier analysis of LUSC in regard to (A) overexpression and baseline-expression-based cutoff; comparing the survival

outcomes of patients who had high ECT2 expression (> 125 cases) with those who had low/medium ECT2 expression (< 369 cases). (B) Box plot indicating that the expression of ECT2 is significantly elevated within primary tumor samples (n = 503) compared to normal samples (n = 52). Expression is captured in transcripts per million, underlining the possible importance of ECT2 within tumor progression. 195

Figure 7.9: (A) Kaplan-Meier survival curve for the number of days survived by LUSC patients with high UHRF1 expression levels, n = 125, as no different from the number of days the low/medium group, n = 369. (B) Box plot illustrating UHRF1 expression in primary tumor samples, n = 503, significantly higher than in normal samples, n = 52, with expression measured in TPM, emphasizing its potential relevance to tumor biology..... 196

Figure 7.10: (A) Kaplan-Meier curve demonstrating survival between patients with LUSC who expressed MCM3 as high (n = 125) versus low/medium (n = 369). (B) Expression levels of MCM3 were significantly upregulated in primary tumor samples (n = 503) compared to normals (n = 52); expression values are in TPM, suggesting MCM3 involvement in tumor biology..... 197

Figure 7.11: (A) Kaplan-Meier survival curve of LUSC patients indicates that patients with high MCM5 expression (n=125) have significantly better survival compared with those with low/medium expressed MCM5 (n=369). (B) Box plot infers the MCM5 expression is significantly higher in primary tumor samples (n=503) compared

to normal samples (n=52), expression valued in TPMs, indicating of the possible importance of MCM5 in tumor progression.198

Figure 7.12: (A) Kaplan-Meier survival curve on LUSC patients presenting high MCM6 expression (n=124), showing slightly better survival trends compared to low/medium expressors (n=370). (B) Box plot indicative of significantly high expression of MCM6 in primary tumor samples (n=503) versus normal (n=52) with the expression measured by TPM, indicating its possible role in the biology of the tumor.199

Figure 7.13: (A) Trend of Kaplan-Meier survival curve in LUSC patients overexpressing MCM7. A total number of 123 patients have high expression of MCM7, compared to those having low/medium expression of MCM7 (n=371). (B) Box plot representing significantly higher expression levels of MCM7 in primary tumor samples from the TCGA database compared to normal samples, suggesting potential implications in tumor biology. Expression levels are in TPM (transcripts per million).200

Figure 7.14: Correlation analysis of MCM5 expression across normal respiratory tissue. Histogram showing the frequency distribution of the correlation values of MCM5 with other genes. Blue lines are the correlation values of the top genes significantly correlated with MCM5. The genes, their descriptions, r-value of correlation, and p-value have been presented in the following table and show the significant positive correlation of MCM5 with MCM7, MCM6, UHRF1, MCM3, ECT2, and MYC, thus indicating

possible interaction and coregulation among these genes in normal respiratory tissue.202

Figure 7.15: Correlation analysis of MCM5 expression across cancer respiratory tissue. Histogram showing the frequency distribution of the correlation values of MCM5 with other genes. Blue lines are the correlation values of the top genes significantly correlated with MCM5. The genes, their descriptions, r-value of correlation, and p-value have been presented in the following table and show the significant positive correlation of MCM5 with MCM7, MCM6, UHRF1, MCM3, ECT2, and MYC, thus indicating possible interaction and coregulation among these genes in cancerous respiratory tissue.....203

List of Tables

Table 2.1:	List of secondary metabolites produced by different <i>Streptomyces</i> species and their therapeutic applications with their year of FDA approval.....	13
Table 2.2:	List of compounds produced by various <i>Streptomyces</i> and their industrial application.....	16
Table 2.3:	List of secondary metabolites produced by the genus <i>Streptomyces</i> with their various agricultural uses.	18
Table 2.4:	Presents a comprehensive list of microbial strains with the potential to synthesize epirubicin, its precursors, or analogous compounds.	26
Table 3.1:	Binding energy and molecular interaction analysis (Hydrogen and Hydrophobic interaction) ligands with triple mutant EGFR ^{L858R/T790M/C797S} and their production source. The bold residues are the conserved one with control compounds. (†: name retrieved from StreptomeDB 3.0 database).....	43
Table 3.2:	Free energies (kcal/mol) of complex EGFR ^{L858R/T790M/C797S} with C_42, C_4299, C_4300, C_5024, C_5702, CH7 and Osimertinib.	58
Table 4.1:	Secondary metabolites identified by the HR-LCMS in an aqueous extract of <i>Streptomyces clavuligerus</i>	77
Table 4.2:	Secondary metabolites identified by the HR-LCMS in the organic extract of <i>Streptomyces clavuligerus</i>	78

Table 4.3:	Binding energy and interaction analysis of the top 3 secondary metabolites <i>Streptomyces clavuligerus</i> and standard ligand Pazopanib with various Receptor Tyrosine Kinases (RTKs).....	81
Table 4.4:	Free energy analysis of inhibitor complexes with target RTKs protein on the last 20 ns simulation frames.	100
Table 5.1:	Secondary metabolites identified by the HR-LCMS from the aqueous extract of <i>Streptomyces fragilis</i> (extracted from the stationary phase of fermentation, after 72 hrs of incubation period).....	120
Table 5.2:	Secondary metabolites identified by the HR-LCMS from the organic extract of <i>Streptomyces fragilis</i> (extracted from the stationary phase of fermentation, after 72 hrs of incubation period).....	122
Table 5.3:	Binding energy and Hydrogen interaction analysis of epirubicin and standard drug Doxorubicin with various types of kinase protein.....	124
Table 5.4:	Free energy analysis of epirubicin and doxorubicin complex with target kinase protein on the last 20 ns simulation frames.	136
Table 6.1:	Process parameters selected for fermentation of <i>Streptomyces fragilis</i> for the production of the metabolites.	144
Table 6.2:	A matrix for the Box-Behnken experimental design is presented, depicting four factors along with their coded and experimental values. Additionally, the response is shown in terms of both the experimental and RSM-predicted values for the production of	

metabolites. (The experimental values were calculated as the mean of triplicate values).....145

Table 6.3: ANOVA analysis of quadratic model157

List of Abbreviations

ADME - Absorption, Distribution, Metabolism, and Excretion

ALA - Alanine

ANN - Artificial Neural Networks

ASP - Aspartic Acid

ATP - Adenosine Triphosphate

BBD - Box-Behnken Design

BCL2: B-Cell Lymphoma 2

BCR-ABL: Breakpoint Cluster Region-Abelson

BFGS: Broyden-Fletcher-Goldfarb-Shanno

BGCs - Biosynthesis Gene Clusters

BLAST: Basic Local Alignment Search Tool

BLIP - Beta-Lactamase Inhibitory Protein

CCD - Central Composite Design

CDK4: Cyclin-Dependent Kinase 4

cDNA: Complementary DNA

DEGs: Differentially Expressed Genes

DGF - Di-Glycine Motif

DMEM-HG: Dulbecco's Modified Eagle Medium-High Glucose

DMSO: Dimethyl Sulfoxide

DNA - Deoxyribonucleic Acid

DPGA - Diphosphoglyceric Acid

DPPH: 2,2-Diphenyl-1-Picrylhydrazyl

EGFR - Epidermal Growth Factor Receptor

ER - Endoplasmic Reticulum

FDA - Food and Drug Administration

FGFR - Fibroblast Growth Factor Receptor

FtsZ - Filamenting Temperature-Sensitive Mutant Z

g/L: Grams per Liter

GC - Guanine-Cytosine

GLN - Glutamine

GO: Gene Ontology

GRNN: General Regression Neural Network

HERG - Human Ether-à-go-go-Related Gene

HGFR - Hepatocyte Growth Factor Receptor

HR-LCMS: High-Resolution Liquid Chromatography-Mass Spectrometry

IC50: Half Maximal Inhibitory Concentration

IGFR - Insulin Growth Factor Receptor

JAK2: Janus Kinase 2

KEGG: Kyoto Encyclopedia of Genes and Genomes

LEU - Leucine

LM: Levenberg-Marquardt

LUSC: Lung Squamous Cell Carcinoma

m/z: Mass-to-Charge Ratio

MAE - Maestro File Format

Mb - Megabase

MD - Molecular Dynamics

MET - Methionine

ML: Machine Learning

MM/GBSA - Molecular Mechanics/Generalized Born Surface Area

MRSA - Methicillin-Resistant Staphylococcus Aureus

NCBI: National Centre for Biotechnology Information

NEB: New England Biolabs (likely referring to NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina)

NPT - Number of particles, Pressure, and Temperature

ns - Nanoseconds

NSC - Non-Small Cell

NSCLC - Non-Small Cell Lung Cancer

NVT - Number of particles, Volume, and Temperature

OPLS - Optimized Potentials for Liquid Simulations

PAINS - Pan Assay Interference Compounds

PDGFR - Platelet-Derived Growth Factor Receptor

PHE - Phenylalanine

PKC - Protein Kinase C

PKI: Protein Kinase Inhibitor

PKS - Polyketide Synthases

PPI: Protein-Protein Interaction

QbD - Quality by Design

R²: Coefficient of Determination

RCSB-PDB - Research Collaboratory for Structural Bioinformatics Protein Data Bank

Rg - Radius of Gyration

rGyr: Radius of Gyration

RMSD - Root Mean Square Deviation

RMSF - Root Mean Square Fluctuation

RNA - Ribonucleic Acid

RSM - Response Surface Methodology

RTK - Receptor Tyrosine Kinase

SASA - Solvent Accessible Surface Area

SCG: Scaled Conjugate Gradient

SCLC: Small-cell lung cancer

SDF - Structure Data File

SER - Serine

SVM: Support Vector Machine

TIP3P - Transferable Intermolecular Potential with 3 Points

TSS: Trace Salt Solution

UD: Uniform Design

VAL - Valine

VEGFR - Vascular Endothelial Growth Factor Receptor

List of Symbols

- °C - Degrees Celsius
- α - Alpha
- μM - Micromolar
- \AA - Angstrom
- Δ - Delta
- ΔG_{bind} - Change in binding free energy
- ΔG_{comp} - Change in complex free energy
- $\Delta G_{\text{protein}}$ - Change in protein free energy
- ΔG_{lig} - Change in ligand free energy
- \pm - Plus-Minus
- γ - Gamma
- λ - Lambda
- Σ - Sigma
- π - Pi
- β - Beta
- μ - Mu
- σ - Sigma
- ρ - Rho
- τ - Tau
- Ω - Omega

- κ - Kappa
- θ - Theta
- φ - Phi
- ψ - Psi
- χ - Chi
- ∞ - Infinity
- \neq - Not equal to
- \approx - Approximately equal to
- \leq - Less than or equal to
- \geq - Greater than or equal to
- \rightarrow - Right arrow
- \leftrightarrow - Double-headed arrow
- $+$ - Plus
- $-$ - Minus
- \times - Multiplication
- \div - Division
- $\sqrt{\quad}$ - Square root
- \sum - Summation
- \int - Integral
- $\mu\text{g/mL}$ - Micrograms per milliliter
- 2θ - Two theta