**DISSERTATION**
Presented by

# CHAHBAOUI Narimene

To obtain the degree of
**Doctorate in Chemistry**

Option:
**Pharmaceutical Chemistry**

Entitled:

## Computational Approaches to Understanding the Structural Features of Small Molecules of Biological Interest

Publicly defended on: February 19, 2025
**In front of the jury composed of:**

| | | | |
|---|---|---|---|
| Mr. BELAIDI Salah | Professor | University of Biskra | President |
| Mrs. KHAMOULI Saida | MCA | University of Biskra | Supervisor |
| Mr. LANEZ Touhami | Professor | University of El-Oued | Examiner |
| Mr. DAOUD Ismail | Professor | University of Biskra | Examiner |
| Mrs. DJOUAMA Hanane | MCA | University of Biskra | Examiner |

*This work is dedicated to my beloved parents*

*To my dear sisters*

*And to all those who hold a special place in my heart*

# Acknowledgments

# Abstract

Pancreatic ductal adenocarcinoma (PDAC) stands out as one of the most aggressive and lethal forms of cancer, characterized by a dismal prognosis largely due to late-stage diagnosis and the inadequate effectiveness of current treatment options. Recognizing the urgent need for innovative therapies, this thesis adopts a dual approach, exploring both direct and indirect inhibition mechanisms to tackle the challenges associated with PDAC treatment. By integrating computer-aided drug design with the chemical diversity of natural products, novel small-molecule inhibitors with potential efficacy against PDAC were identified.

Divided into two main parts, the study begins with an exploration of curcumin derivatives as potential inhibitors of PI3Kα. Through virtual screening that employed pharmacophore modeling and molecular docking, two promising compounds were identified: CID154728220 and CID156189304. Both compounds exhibited favorable pharmacokinetic profiles, and subsequent MD simulations confirmed their structural stability, making them strong candidates for further preclinical evaluation.

The second part utilized a fragment-based drug discovery approach to directly target KRAS G12D. This strategy led to the discovery of two novel compounds, Hit1 and Hit2, which demonstrated higher binding affinity compared to the reference inhibitor MRTX-1133, as validated by IFD and MM-GBSA analysis. Notably, Hit2 exhibited the most favorable balance of pharmacokinetic properties, safety, and drug-likeness, positioning it as a promising candidate for further development.

This thesis represents a significant advancement in cancer research, providing a foundation for the development of novel therapeutic strategies in the treatment of pancreatic cancer.

**Keywords:** Pancreatic cancer, virtual screening, natural products, pharmacophore modeling, molecular docking, molecular dynamics simulations.

# الملخص

يُعتبر سرطان البنكرياس الغدي القنوي (PDAC) من أكثر أنواع السرطان شراسة وفتكًا، حيث يظهر آفاقًا صحية غير جيدة، ويرجع ذلك بشكل رئيسي إلى اكتشاف المرض في مراحل متأخرة وعدم فعالية العلاجات المتاحة. نظرًا للحاجة المُلِحَّة إلى علاجات مبتكرة، تتبنى هذه الأطروحة منهجًا مزدوجًا، حيث تستعرض آليات التثبيط المباشر وغير المباشر للتصدي للتحديات المرتبطة بعلاج PDAC. من خلال دمج تصميم الأدوية بمساعدة الحاسوب مع التنوع الكيميائي للمركبات الطبيعية، تم اكتشاف مثبطات جزيئية صغيرة جديدة ذات فعالية محتملة ضد سرطان البنكرياس.

تنقسم الدراسة إلى جزأين رئيسيين، حيث يركز الجزء الأول على تحديد مشتقات الكركم كمثبطات محتملة لإنزيم PI3Kα. من خلال الفحص الافتراضي الذي تضمن نمذجة الفارماكوفور والالتحام الجزيئي، تم تحديد مركبين رئيسيين هما CID154728220 و CID156189304. كلاهما أظهر خصائص حركية دوائية واعدة واستقرارًا بنيويًا في محاكاة الديناميكيات الجزيئية، مما يجعلهما مرشحين واعدين للتقييم ما قبل السريري.

اعتمد الجزء الثاني نهج اكتشاف الأدوية القائم على الأجزاء لاستهداف KRAS G12D. أدى هذا النهج إلى اكتشاف مركبين جديدين، هما Hit1 و Hit2 اللذان أظهرا قابلية ارتباط أعلى مقارنة بالمثبط المرجعي MRTX-1133، كما تم التحقق من صحة ذلك من خلال تحليل IFD و MM-GBSA. تجدر الإشارة إلى أن المركب Hit2 أظهر التوازن الأكثر ملاءمة بين الخصائص الحركية الدوائية، السلامة، و drug-likeness، مما يؤهله ليكون مرشحًا قويًا لمزيد من التطوير.

تعد هذه الأطروحة مساهمة هامة في أبحاث السرطان، حيث تفتح آفاقاً لتطوير استراتيجيات علاجية جديدة لسرطان البنكرياس.

**الكلمات المفتاحية:** سرطان البنكرياس، الفحص الافتراضي، المركبات الطبيعية، نمذجة الفارماكوفور، الالتحام الجزيئي، محاكاة الديناميكيات الجزيئية.

# Table of Contents

## Chapter I: Background and Rationale

## Chapter II: Computational Methods in Drug Discovery

## Chapter III: Application 1 – Comprehensive Computational Strategies to Identify Novel Curcumin Derivatives Against Pancreatic Cancer

## Chapter IV: Application 2 – Targeting KRAS G12D Using Fragment-Based Drug Discovery

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| **ADMET** | Absorption, Distribution, Metabolism, Excretion, and Toxicity |
| **ANNs** | Artificial Neural Networks |
| **AUC** | Area Under the Curve |
| **CADD** | Computer-Aided Drug Design |
| **CoMFA** | Comparative Molecular Field Analysis |
| **CoMSIA** | Comparative Molecular Similarity Indices Analysis |
| **DCCM** | Dynamic Cross-Correlated Motions |
| **DL** | Deep Learning |
| **EF** | Enrichment Factor |
| **EMT** | Epithelial-Mesenchymal Transition |
| **FBDD** | Fragment-Based Drug Discovery |
| **FDA** | U.S. Food and Drug Administration |
| **FOLFIRINOX** | FOLinic acid, Fluorouracil, IRINotecan, OXaliplatin |
| **GAPs** | GTPase-Activating Proteins |
| **GEFs** | Guanine nucleotide Exchange Factors |
| **GOLPE** | Generating Optimal Linear PLS Estimations |
| **HRAS** | Harvey Rat Sarcoma viral oncogene homolog |
| **HTS** | High-Throughput Screening |
| **HTVS** | High-Throughput Virtual Screening |
| **IFD** | Induced Fit Docking |
| **IUPAC** | International Union of Pure and Applied Chemistry |
| **KNIME** | Konstanz Information Miner |
| **KRAS** | Kirsten Rat Sarcoma viral oncogene homolog |
| **LBDD** | Ligand-Based Drug Design |
| **MD** | Molecular Dynamics |

| | |
|---|---|
| **ML** | Machine Learning |
| **MLR** | Multiple Linear Regression |
| **MM-GBSA** | Molecular Mechanics Generalized Born Surface Area |
| **MM-PBSA** | Molecular Mechanics Poisson-Boltzmann Surface Area |
| **MMPs** | Matrix Metalloproteinases |
| **NCEs** | New chemical entities |
| **NPT** | Isothermal-isobaric ensemble [constant number of N, P, and T] |
| **NRAS** | Neuroblastoma RAS viral oncogene homolog |
| **NVT** | Canonical ensemble [constant number of N, V, and T] |
| **PanINs** | Pancreatic Intraepithelial Neoplasia |
| **PBCs** | Periodic Boundary Conditions |
| **PCA** | Principal Component Analysis |
| **PDAC** | Pancreatic Ductal Adenocarcinoma |
| **PI3K/Akt/mTOR** | Phosphoinositide 3-kinase/protein kinase B/mammalian target of rapamycin pathway |
| **PI3Kα** | Phosphoinositide 3-kinase alpha isoform |
| **PI3Kβ** | Phosphoinositide 3-kinase beta isoform |
| **PI3Kγ** | Phosphoinositide 3-kinase gamma isoform |
| **PI3Kδ** | Phosphoinositide 3-kinase delta isoform |
| **PIK3CA** | Phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha |
| **PLS** | Partial Least Squares |
| **PNETs** | Pancreatic Neuroendocrine Tumors |
| **QSAR** | Quantitative Structure-Activity Relationship |
| **Raf/Mek/Erk** | Rapidly accelerated fibrosarcoma/ mitogen-activated protein kinase kinase/extracellular signal-regulated kinase pathway |
| **RalGDS/Ral** | Ral guanine nucleotide dissociation stimulator/Ras-like GTPases pathway |
| **RF** | Random Forests |
| **Rg** | Radius of Gyration |

| | |
|---|---|
| **RIE** | Robust Initial Enhancement |
| **RMSD** | Root Mean Square Deviation |
| **RMSF** | Root Mean Square Fluctuation |
| **ROC** | Receiver Operating Characteristic |
| **SASA** | Solvent-Accessible Surface Area |
| **SBDD** | Structure-Based Drug Design |
| **SP** | Standard Precision |
| **SVM** | Support Vector Machine |
| **TIP3P** | Transferable Intermolecular Potential with 3 Points |
| **XP** | Extra Precision |

# General Introduction

Pancreatic cancer, particularly pancreatic ductal adenocarcinoma (PDAC), is among the most aggressive and lethal cancers, with a dismal five-year survival rate of just 13% [1]. The increasing incidence of this disease is particularly alarming, positioning it to potentially become the second leading cause of cancer fatalities by 2030 [2].

Current therapeutic strategies remain extremely limited, with standard treatments such as surgery, chemotherapy, and radiation often failing to deliver long-term survival benefits. Surgical resection, the only potentially curative option, is feasible in less than 20% of cases due to late-stage diagnosis [3]. For those with non-resectable tumors, chemotherapy regimens provide only modest survival improvements, typically at the expense of significant toxicity and reduced quality of life [4].

The dire prognosis for pancreatic cancer patients underscores the pressing need for novel and more effective therapeutic strategies. Nevertheless, developing successful treatments for this malignancy is particularly challenging due to its highly complex genetic landscape and the prevalence of KRAS mutations, which occur in over 90% of cases. These mutations result in the aberrant activation of the KRAS protein, which drives the aggressive progression of PDAC, making it an attractive therapeutic target [5,6]. Despite its critical role, direct targeting of KRAS has proven difficult due to structural challenges that have rendered it "undruggable" [7,8]. Consequently, researchers have shifted their focus toward targeting downstream effectors within the KRAS signaling cascade. Among these, the PI3K pathway, particularly the PI3Kα isoform, has gained considerable attention. PI3Kα is essential for transmitting oncogenic signals from mutated KRAS, promoting tumor survival, proliferation, and metastasis. As a result, it represents a viable and promising alternative for therapeutic intervention, offering hope for improved treatment outcomes in PDAC [9].

Recent advancements have however reignited hope in directly targeting KRAS. The FDA approval of inhibitors for KRAS G12C, a specific KRAS mutation subtype prevalent in non-small cell lung cancer, marks a significant breakthrough [10,11]. This success has fueled intensified efforts to develop inhibitors for other KRAS variants, particularly KRAS G12D, the most prevalent mutation in pancreatic cancer.

Computer-aided drug design (CADD) has revolutionized the pharmaceutical industry by integrating computational tools into the drug discovery process. These tools simulate interactions between drug molecules and biological targets, enabling faster and more accurate identification of potential therapeutics. With the incorporation of artificial intelligence and machine learning, CADD continues to push the boundaries of drug design, reducing both the time and cost associated with traditional experimental methods [12]. A key component of CADD is virtual screening, which allows researchers to rapidly evaluate large libraries of compounds to identify those most likely to interact with a target of interest. This approach significantly narrows down the pool of candidate molecules for further development [13].

Natural products remain a powerful source of inspiration in drug discovery, as they often exhibit complex structures and potent biological activities that are challenging to replicate synthetically. Derived from plants, microorganisms, and marine organisms, these compounds have led to the development of many successful drugs [14].

This research aims to identify and optimize potential inhibitors for pancreatic cancer through both direct and indirect targeting strategies. By integrating natural products with cutting-edge computational techniques, the advancements discussed in this thesis not only pave the way for new treatment options but also highlight the untapped potential of naturally derived molecules in modern drug design to address unmet medical needs. To provide a structured and comprehensive examination of this topic, the thesis is organized as follows:

- **Chapter I** provides a detailed overview of pancreatic cancer, emphasizing its biology, the role of KRAS mutations, and the therapeutic challenges associated with targeting this protein. Additionally, the chapter explores the potential of natural products in cancer therapy, with a particular focus on curcumin and its mechanisms of action.

- **Chapter II** delves into the principles and applications of CADD in drug discovery, setting the stage for the computational strategies explored in subsequent chapters.

- **Chapter III** investigates the use of a comprehensive computational strategy to identify novel curcumin derivatives as inhibitors of PI3Kα for the treatment of pancreatic cancer.

- **Chapter IV** shifts focus to the direct targeting of KRAS G12D, leveraging the chemical diversity of natural products through fragment-based drug discovery.

Finally, the thesis concludes with a summary of the research findings and recommendations for future research directions.

# References

[1] R.L. Siegel, A.N. Giaquinto and A. Jemal, *Cancer statistics, 2024*, CA. Cancer J. Clin. 74 (2024), pp. 12–49.

[2] R.U. Harvitkar, H. Peri, S.N. Zallipalli, S.J. Joseph, G.B. Gattupalli and K. Ansari, *Non-Cancer Causes of Death in Patients With Pancreatic Adenocarcinoma: A Surveillance, Epidemiology, and End Results (SEER)-Based Study*, Cureus 13 (2021), pp. e20289.

[3] A. Stathis and M.J. Moore, *Advanced pancreatic carcinoma: current treatment and future challenges*, Nat. Rev. Clin. Oncol. 7 (2010), pp. 163–172.

[4] L. Buscail, B. Bournet and P. Cordelier, *Role of oncogenic KRAS in the diagnosis, prognosis and treatment of pancreatic cancer*, Nat. Rev. Gastroenterol. Hepatol. 17 (2020), pp. 153–168.

[5] S. Sankarasubramanian, U. Pfohl, C.R.A. Regenbrecht, C. Reinhard and L. Wedeken, *Context Matters—Why We Need to Change From a One Size Fits all Approach to Made-to-Measure Therapies for Individual Patients With Pancreatic Cancer*, Front. Cell Dev. Biol. 9 (2021), pp. 760705.

[6] A.M. Waters and C.J. Der, *KRAS: The Critical Driver and Therapeutic Target for Pancreatic Cancer*, Cold Spring Harb. Perspect. Med. 8 (2018), pp. a031435.

[7] C. Zhu, X. Guan, X. Zhang, X. Luan, Z. Song, X. Cheng et al., *Targeting KRAS mutant cancers: from druggable therapy to drug resistance*, Mol. Cancer 21 (2022), pp. 159.

[8] A.D. Cox, S.W. Fesik, A.C. Kimmelman, J. Luo and C.J. Der, *Drugging the undruggable Ras: mission possible?*, Nat. Rev. Drug Discov. 13 (2014), pp. 828–851.

[9] B. Thibault, F. Ramos-Delgado, E. Pons-Tostivint, N. Therville, C. Cintas, S. Arcucci et al., *Pancreatic cancer intrinsic PI3Kα activity accelerates metastasis and rewires macrophage component*, EMBO Mol. Med. 13 (2021), pp. e13502.

[10] S. Dhillon, *Adagrasib: First Approval*, Drugs 83 (2023), pp. 275–285.

[11] E.C. Nakajima, N. Drezner, X. Li, P.S. Mishra-Kalyani, Y. Liu, H. Zhao et al., *FDA Approval Summary: Sotorasib for KRAS G12C Mutated Metastatic NSCLC*, Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 28 (2022), pp. 1482–1486.

[12] X. Yang, Y. Wang, R. Byrne, G. Schneider and S. Yang, *Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery*, Chem. Rev. 119 (2019), pp. 10520–10594.

[13] C.G. Wermuth, B. Villoutreix, S. Grisoni, A. Olivier and J.-P. Rocher, *Chapter 4 - Strategies in the Search for New Lead Compounds or Original Working Hypotheses*, in *The Practice of Medicinal Chemistry (Fourth Edition)*, C.G. Wermuth, D. Aldous, P. Raboisson and D. Rognan, eds., Academic Press, San Diego, 2015, pp. 73–99.

[14] G.M. Cragg and D.J. Newman, *Natural Product Drug Discovery in the Next Millennium*, Pharm. Biol. 39 (2001), pp. 8–17.

# Chapter I

# Background and Rationale

## I.1    Introduction

Pancreatic cancer presents a unique challenge in the field of oncology. Often referred to as a "silent killer," it advances swiftly and insidiously, with symptoms that typically remain unnoticed until the disease has significantly advanced [1]. The aggressive nature of this malignancy, coupled with its resistance to standard treatments, contributes to a bleak prognosis [2]. The five-year survival rate remains dishearteningly low, hovering around a mere 13%, significantly lower than that of many other cancers [3].

Central to the aggressive nature of pancreatic cancer is the prevalence of KRAS mutations, which are found in approximately 94% of cases [4]. These mutations result in an activated KRAS protein that aberrantly triggers numerous downstream signaling pathways, thereby promoting cancer cell proliferation, immune evasion, metabolic reprogramming, and resistance to therapy [5]. Consequently, KRAS functions as a critical driver of pancreatic cancer and is linked to a poorer prognosis for affected patients [6].

Targeting KRAS directly has been challenging due to the protein's structure, which lacks classic drug-binding sites, making it historically "undruggable." Recent breakthroughs, however, have resulted in the innovation of indirect targeting strategies. These include targeting KRAS expression, processing, upstream regulators, or downstream effectors [7].

This chapter delves into the intricacies of targeting critical downstream signaling effectors of KRAS, with a particular focus on the PI3K/AKT/mTOR pathway. This pathway is a key regulator of cell survival, growth, and proliferation, and its dysregulation is often associated with the onset and advancement of numerous cancers, including pancreatic cancer. By inhibiting this pathway, researchers aim to disrupt the critical signals that cancer cells rely on for survival and growth, potentially overcoming therapeutic resistance [8]. Furthermore, the PI3K/AKT/mTOR pathway cross-talks with other signaling cascades, and its inhibition could have a broader impact on the tumor microenvironment, impacting not only the tumor cells but also the surrounding stromal and immune cells. This could lead to a more comprehensive approach to cancer therapy, enhancing the effectiveness of current treatments [9].

The unmatched chemical diversity of natural products has been instrumental in drug discovery, resulting in the development of numerous bioactive compounds. Having historically provided groundbreaking drugs, natural products remain a source of inspiration

and challenge for synthetic chemists and pharmacologists. This chapter explores these dynamics, with a particular focus on curcumin, a turmeric-derived compound with promising anticancer properties. Due to its ability to influence multiple signaling pathways, including those governing inflammation [10], apoptosis [11], and cell proliferation [12], curcumin emerges as a compelling candidate for pancreatic cancer therapy. The chapter will discuss the mechanisms by which curcumin exerts its effects and its potential to enhance the efficacy of existing treatments.

## I.2    Pancreatic cancer

### I.2.1    Pancreas anatomy and function

In humans, the pancreas typically weighs between 50 to 100 grams and measures approximately 14 to 18 centimeters in length. It is comprised of three main parts: the head, the body, and the tail. The C-shaped head of the pancreas aligns with the upper curvature of the duodenum, while its flattened body lies horizontally beneath the stomach, spanning major blood vessels such as the superior mesenteric artery and vein, the abdominal aorta, the inferior vena cava, and the portal vein. Extending from the body, the pancreatic tail reaches and connects to the hilum of the spleen [13].
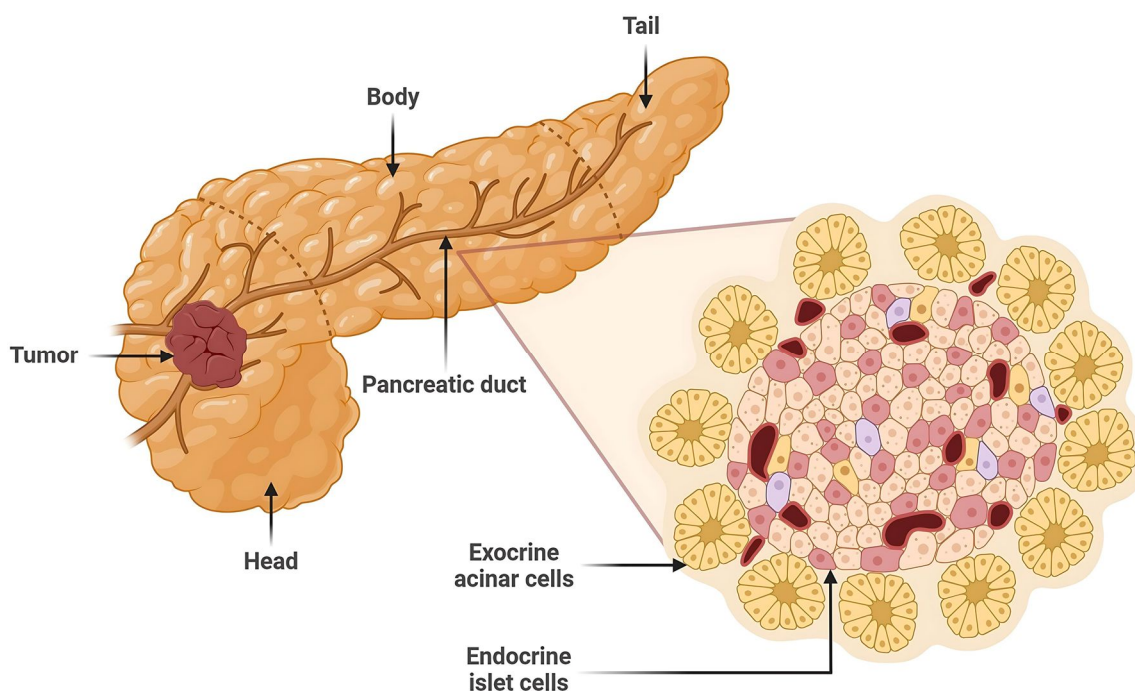


Figure I.1: Pancreatic morphology and cellular composition. Created with BioRender.com

The normal pancreas consists of both exocrine and endocrine glands, which, although functionally distinct, are anatomically intertwined. The exocrine gland releases digestive enzymes that are transported to the duodenum through a network of ducts, facilitating the digestion of nutrients such as carbohydrates, fats, and proteins. Meanwhile, the endocrine gland comprises the islets of Langerhans, housing alpha and beta cells that produce glucagon and insulin, respectively. These hormones are essential in maintaining the balance of blood glucose levels [14,15].

### I.2.2    Clinical aspects of pancreatic cancer

Pancreatic cancer was initially documented by Giovanni Battista Morgagni in his 1761 publication [16]. However, the first dependable microscopic diagnosis of this malignancy was presented by Jacob M. Da Costa in 1858 [17]. Tumors can originate from both endocrine and exocrine cells in the pancreas. Endocrine cells typically lead to the development of pancreatic neuroendocrine tumors (PNETs), constituting only 1% to 3% of all newly diagnosed pancreatic cancers [14,18]. In contrast, pancreatic ductal adenocarcinoma (PDAC), arising from exocrine cells, represents the most prevalent type of pancreatic cancer, making up approximately 90% of cases [19].

Pancreatic cancer is categorized into four stages, each representing the extent of tumor progression. In Stage I, the cancer is confined to the pancreas, with tumor sizes ranging from up to 2 cm (IA) to just under 4 cm (IB). Stage II involves local spread or borderline resectable cases, where the tumor exceeds 4 cm but remains within the pancreas or has extended to nearby lymph nodes. Stage III indicates more extensive spread, with potential involvement of nearby blood vessels or nerves but no distant metastasis. Finally, Stage IV represents metastatic cancer, with confirmed spread to distant organs [20].

During its initial stages, pancreatic cancer often manifests without discernible symptoms [1], earning it the ominous moniker of the "silent killer." As the tumor progresses, indications become evident through a gradual onset of non-specific symptoms, including fatigue, weight loss, jaundice, light-colored stools, and abdominal pain [21]. The existing diagnostic tests lack specificity and may overlook individuals in the early stages of the disease [1].

### I.2.3 Global trends and risk factors in pancreatic cancer

Pancreatic cancer is the seventh leading cause of cancer-related deaths globally and the third most prevalent in the United States [20]. It is characterized by a dishearteningly low 5-year survival rate, standing at a mere 13% [3]. The global scale of its devastation becomes apparent in the staggering statistics from 2022, revealing 511,000 new cases and 467,000 deaths, with a minimal disparity between incidence and mortality rates [22].

This malignancy is a rare occurrence before the age of 40 [23], reaching its highest incidence in individuals aged over 70, and exhibiting a slightly elevated prevalence in men compared to women [24,25]. The mortality risk associated with pancreatic cancer shows a significant age-dependent increase, rising from less than 2 deaths per 100,000 person-years for individuals aged 35−39 to exceeding 90 deaths per 100,000 person-years for those over 80 in the United States [26].
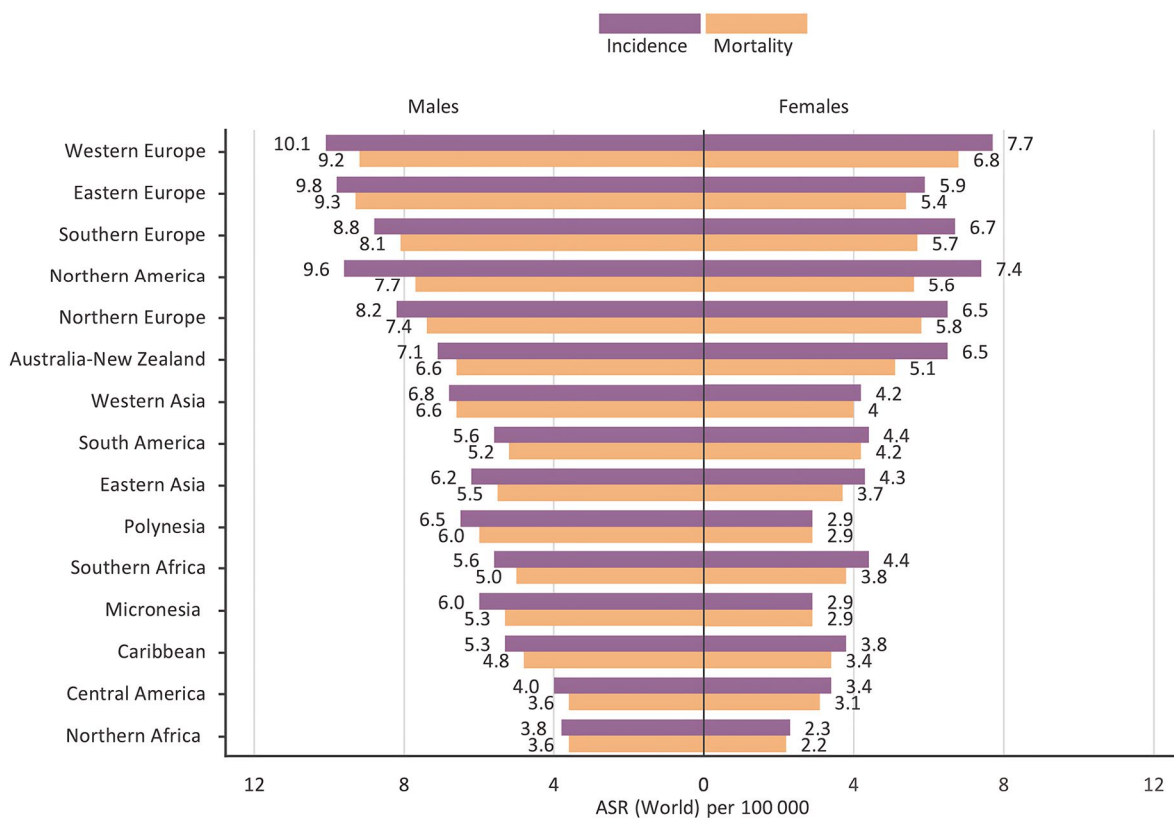


Figure I.2: Global age-standardized pancreatic cancer incidence and mortality rates by gender in 2022.

As global health advances and life expectancy increases, the incidence of this deadly disease is expected to rise. Alarming projections suggest that pancreatic cancer could

become the second leading cause of cancer fatalities by 2030 [27]. This dire forecast underscores the urgent need for increased awareness, research, and interventions to mitigate the growing impact of pancreatic cancer on public health.

Aside from aging, several other risk factors contribute to the development of this malignancy, including:

- **Smoking:** With over a billion smokers globally, tobacco use is the primary environmental risk factor for pancreatic cancer. The causal relationship between smoking and pancreatic cancer has been unequivocally established by the International Agency for Research on Cancer [28]. The risk of developing pancreatic cancer is directly associated with both the duration of smoking and the daily quantity of cigarettes consumed. Smokers face an almost twofold higher risk compared to non-smokers [29,30]. Moreover, a comprehensive meta-analysis of 82 studies has provided compelling evidence, revealing a relative risk of 1.74 for current smokers and 1.2 for former smokers, with the risk persisting for at least a decade after cessation [31–33].

- **Obesity:** Obesity is closely associated with a heightened risk of pancreatic cancer [34], contributing to both increased incidence and mortality rates [35,36]. Research by Li and colleagues [37] indicates that being overweight or obese in early adulthood raises the risk of pancreatic cancer. Additionally, obesity in older age (30−79 years) has been correlated with poorer overall survival. The American Cancer Society found that obese individuals have a significantly higher risk of pancreatic cancer compared to those with a healthy BMI, with a relative risk of 2.08 [36].

- **Diabetes mellitus:** Diabetes mellitus acts both as a risk factor for and a consequence of pancreatic cancer. Many individuals newly diagnosed with pancreatic cancer report the onset of diabetes, while those with pre-existing diabetes often experience a worsening of their condition. Notably, having long-standing diabetes (lasting more than 3 years) is linked to a 1.5−2.4-fold increased risk of developing pancreatic cancer [38–40].

- **Pancreatitis:** Marked by inflammation of the pancreas, this condition can manifest in acute or chronic forms, potentially leading to organ damage. Emerging evidence has increasingly associated chronic pancreatitis with a heightened risk of pancreatic cancer [41–43]. Approximately 1.8% of those with chronic pancreatitis develop pancreatic cancer within a decade of diagnosis, with the risk rising to 4% after 20 years [20].

### I.2.4 Current management of pancreatic cancer

Due to the inherent challenges in early-stage PDAC diagnosis, merely 10−15% of patients with stage I or II disease are eligible for surgical resection [44]. Unfortunately, post-tumor resection outcomes remain unsatisfactory, with large-scale randomized clinical trials for postoperative chemotherapy showing a modest median disease-free survival of 20−22 months [45–49]. This indicates that, in the majority of cases, despite early detection, surgical intervention, and chemotherapy, pancreatic ductal adenocarcinoma inevitably progresses to advanced and metastatic stages, leading to mortality.

Various surgical techniques are employed for pancreatic cancer based on its location within the pancreas. The pancreaticoduodenectomy, commonly known as the Whipple procedure, is the most frequently performed surgery to remove the head of the pancreas when cancer is confined in this part. Distal pancreatectomy involves the removal of the pancreatic tail, often accompanied by the spleen. Total pancreatectomy, on the other hand, entails the complete removal of the pancreas and adjacent tissues or organs if cancer has extensively spread throughout the pancreas [50].

In the case of 35% of patients diagnosed with locally advanced disease (stage III) and 50% presenting with metastatic disease (stage IV), the exclusive therapeutic recourse is palliative chemotherapy [44]. Gemcitabine and 5-fluorouracil have been the cornerstone of therapeutic regimens, albeit with limited survival benefits [51]. The recent introduction of the FOLFIRINOX combination therapy has demonstrated a notable yet incremental enhancement in overall survival for individuals with locally advanced and metastatic disease compared to gemcitabine alone. Patients with locally advanced PDAC who undergo treatment with FOLFIRINOX exhibit a median overall survival of 24.2 months, a significant improvement compared to the 6−13 months typically observed with gemcitabine. Likewise, in the case of metastatic patients, the duration of overall survival is extended to 11.1 months with FOLFIRINOX, whereas it stands at 6.8 months with gemcitabine [52,53].

### I.2.5 Genetic mutations in PDAC

In pancreatic cancer, genetic aberrations are prevalent in about 97% of cases, involving point mutations, amplifications, deletions, translocations, and inversions. KRAS mutations, predominant in 94% of cases, significantly contribute to the initiation of pancreatic ductal

adenocarcinoma. Other frequently altered genes include TP53 (64%), SMAD4 (21%), and CDKN2A (17%) [4,54].

PDAC progression unfolds over approximately 12 years [56], marked by a stepwise transformation of pancreatic duct epithelium into noninvasive microscopic ductal lesions known as pancreatic intraepithelial neoplasias (PanINs). Early-stage PanINs exhibit distinct changes, including the transformation of flat pancreatic epithelial cells into a cuboidal appearance, heightened mucin production, and atypical cytological and morphological features [57,58]. High-grade PanINs are characterized by a papillary morphology [58].
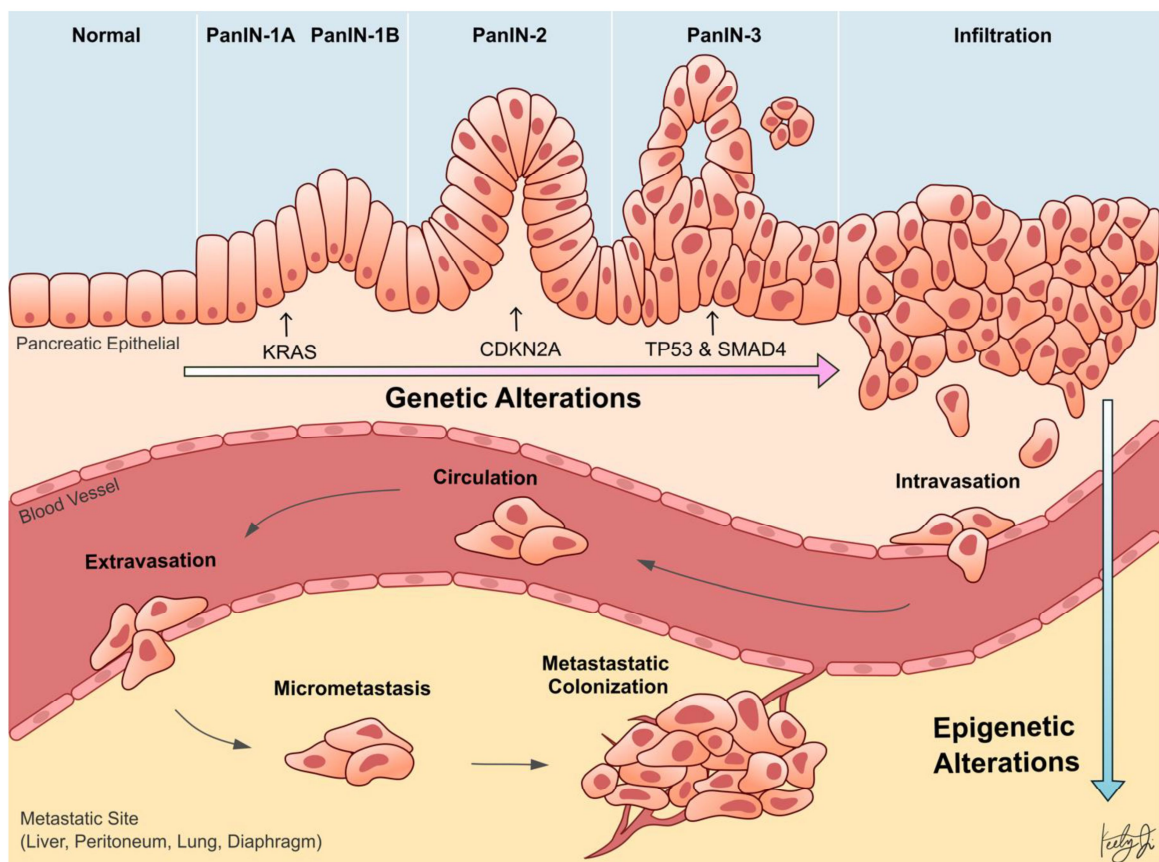


Figure I.3: Illustration depicting the progression of PDAC from normal pancreatic epithelial, through pancreatic intraepithelial neoplasia, to metastasis. Adapted from [55].

Developing effective treatments for pancreatic cancer is particularly challenging due to the complex and heterogeneous genetic landscape of the disease. This genetic variability makes it difficult to develop one-size-fits-all therapies, as different mutations may lead to varying responses to treatment, reducing the effectiveness of standard therapies and necessitating the development of personalized approaches [59]. Amid this complexity,

KRAS mutations stand out as the most prevalent genetic drivers of pancreatic ductal adenocarcinoma, playing a crucial role in tumor initiation, progression, and maintenance [60]. This central role makes KRAS an especially compelling target, as its effective inhibition could potentially reduce tumor growth and promote regression.

### I.2.5.1 Oncogenic KRAS mutations

The RAS family consists of three genes: KRAS (Kristen-Ras), HRAS (Harvey-RAS), and NRAS (neuroblastoma-RAS). Among these, KRAS emerges as the primary mutated RAS gene in cancers, accounting for approximately 84% of all RAS missense mutations. NRAS follows, representing around 12%, while HRAS mutations are infrequent, occurring in only about 4% of cases [4].
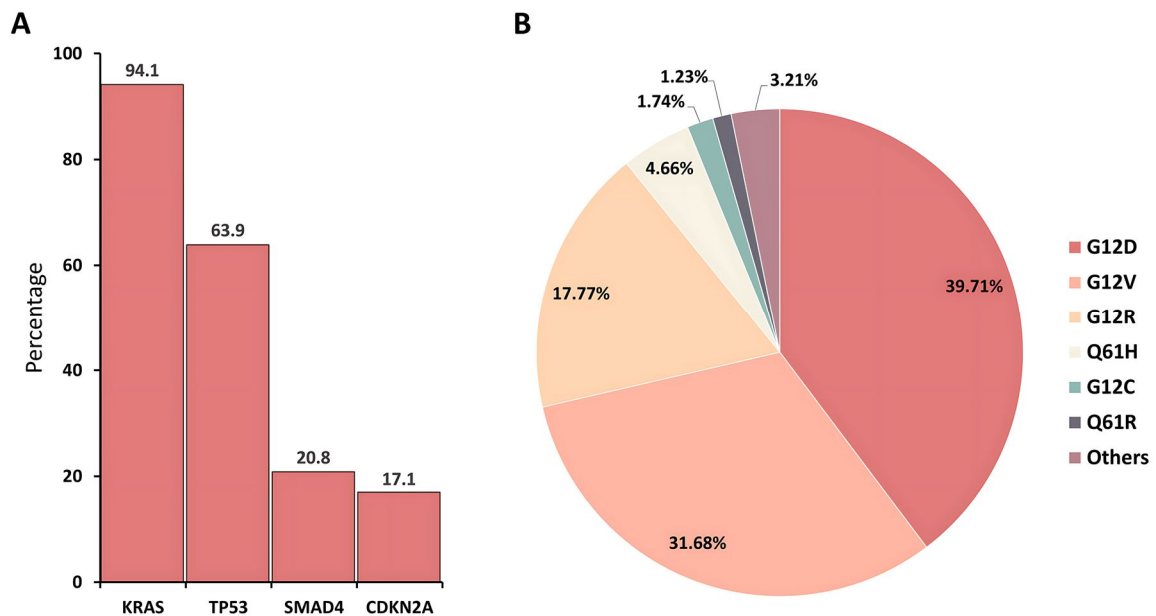


Figure I.4: KRAS mutations in pancreatic cancer. (A) Frequency of mutations in the four major genes in pancreatic cancer. (B) KRAS amino acid substitutions in pancreatic cancer. Adapted from [4,61].

Oncogenic KRAS genes are characterized by missense mutations that encode single amino acid substitutions, predominantly occurring at one of three mutational hotspots: glycine-12 (G12), glycine-13 (G13), or glutamine-61 (Q61) [62]. KRAS mutations are observed in numerous cancers, exhibiting varying mutation frequencies and a diverse range of mutation subtypes. In pancreatic ductal adenocarcinoma, for instance, the predominant

mutant subtype is G12D, characterized by a glycine-to-aspartic acid substitution at codon 12. This is followed by G12V and G12R mutations [61].

### I.2.5.2   RAS proteins

RAS genes encode small GTP-binding proteins that function as molecular switches to regulate a variety of cellular processes, including but not limited to cell proliferation, invasion, autophagy, differentiation, apoptosis, migration, gene expression, and tumor microenvironment [63]. RAS protein activity is regulated through a cycle of guanine nucleotide binding, in which the protein alternates between two states: an inactive form bound to GDP and an active form bound to GTP.

Activation is driven by guanine nucleotide exchange factors (GEFs), which catalyze the release of GDP. Due to the high intracellular GTP concentration, this release facilitates the rapid exchange of GDP for GTP. Binding of GTP induces a conformational change in RAS, which enhances its interaction with effector proteins and activates downstream signaling pathways. The inactivation of RAS signaling is achieved through GTPase-activating proteins (GAPs), which expedite the conversion of GTP to GDP, thus returning RAS to its inactive state [64].
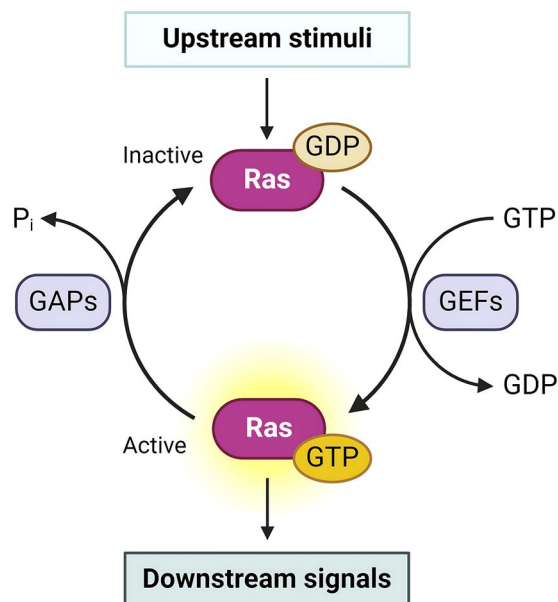


Figure I.5: Regulation of the Ras molecular switch by RasGEFs and RasGAPs. Created with BioRender.com

Mutations in RAS genes, such as those occurring at codon 12, can render the mutant proteins GAP-resistant, causing RAS to remain persistently bound to GTP and thereby constitutively active, independent of extracellular stimuli. This leads to the overstimulation of downstream signaling pathways, triggering several hallmarks of cancer, including cell migration and metastasis, sustained proliferation, remodeling of the tumor microenvironment, metabolic reprogramming, evasion of the immune response as well as resistance to apoptosis [65].

### I.2.6    Targeting KRAS
### I.2.6.1    Directly

Efforts to discover GTP-competitive inhibitors for RAS proteins, similar to the successful ATP-competitive inhibitors for protein kinases, have encountered significant challenges. Unlike ATP, which binds to protein kinases with low micromolar affinity, GTP binds to RAS proteins with a much higher picomolar affinity, making it extremely difficult to displace [66]. Additionally, the structural features of RAS proteins, being nearly spherical and lacking distinct drug-binding sites, further complicate the development of effective inhibitors by providing fewer interaction points for potential drugs [7]. As a result, KRAS has been deemed "undruggable" for over 40 years. Nonetheless, recent breakthroughs have led to the development of targeted therapies for specific KRAS mutations.

FDA-approved drugs such as Sotorasib and Adagrasib have shown efficacy in treating cancers with the KRAS G12C mutation, particularly non-small cell lung cancer [67,68]. Despite this progress, these drugs exhibit limited effectiveness in the case of PDAC and other cancers characterized by the KRAS G12D mutation [69]. A promising development on this front is the ongoing clinical trials for MRTX-1133, a drug specifically tailored to target KRAS G12D. It has shown great promise in preclinical studies and is currently in clinical development for the treatment of solid tumors, including pancreatic and colorectal cancer [70,71].

### I.2.6.2    Indirectly

In light of historical challenges in directly inhibiting mutant KRAS, current and past strategies for developing therapeutics to block its function have predominantly centered around indirect approaches. These involve targeting downstream effectors and pathways influenced by KRAS activation which are more easily targeted with small molecule

inhibitors. In pancreatic ductal adenocarcinoma, oncogenic KRAS signaling is widely recognized to be orchestrated through three key pathways: Raf/Mek/Erk, PI3K/Akt/mTOR, and the RalGDS/Ral pathway [5,72,73].
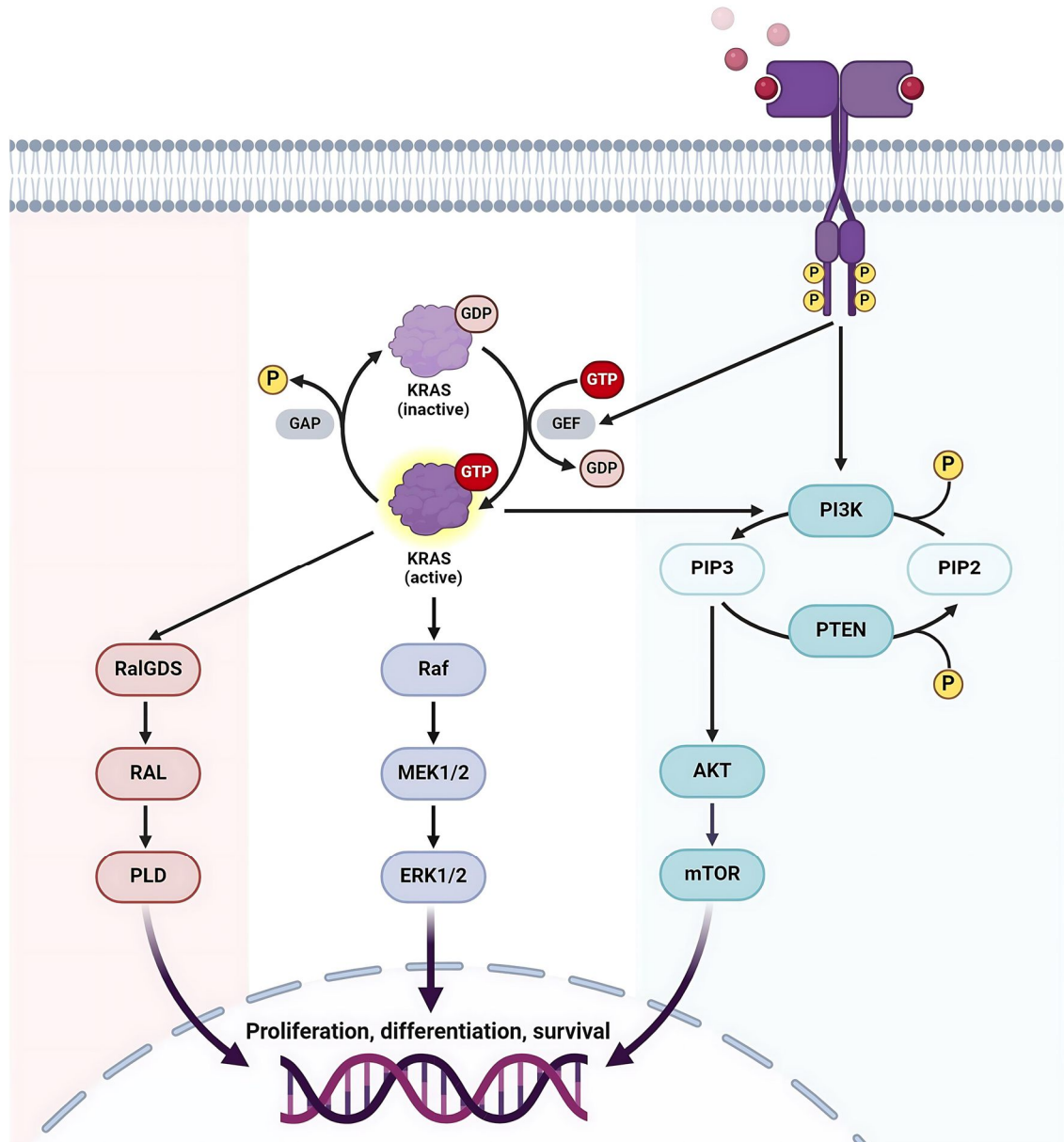


Figure I.6: KRAS-mediated signaling pathways in pancreatic cancer. Created with BioRender.com

The PI3K/AKT/mTOR pathway is a major intracellular signaling cascade that regulates several essential cellular processes for tumorigenesis and metastasis, including growth, survival, proliferation, and metabolism [74]. The hyperactivation of this pathway in pancreatic cancer is associated with poor overall survival in patients and is a significant contributor to the aggressiveness and therapeutic resistance of the disease. Targeting the

PI3K/Akt/mTOR pathway shows promise in treating pancreatic ductal adenocarcinoma, with studies indicating reduced tumor burden and improved patient survival [9,75].

### I.2.7    PI3K/AKT/mTOR pathway

The phosphoinositide 3-kinase (PI3K) family is divided into three classes, with Class I being the most commonly implicated in cancer. Class I PI3Ks are further categorized into Class IA and Class IB. Class IA PI3Ks are heterodimers composed of a regulatory subunit (p85α, p55α, p50α, p85β, or p55γ) and a catalytic subunit (p110α, p110β, or p110δ). In contrast, Class IB PI3Ks consist of a regulatory subunit (p101) and a catalytic subunit (p110γ). Together, these combinations form the four isoforms of Class I PI3Ks: PI3Kα, PI3Kβ, PI3Kδ, and PI3Kγ [76].
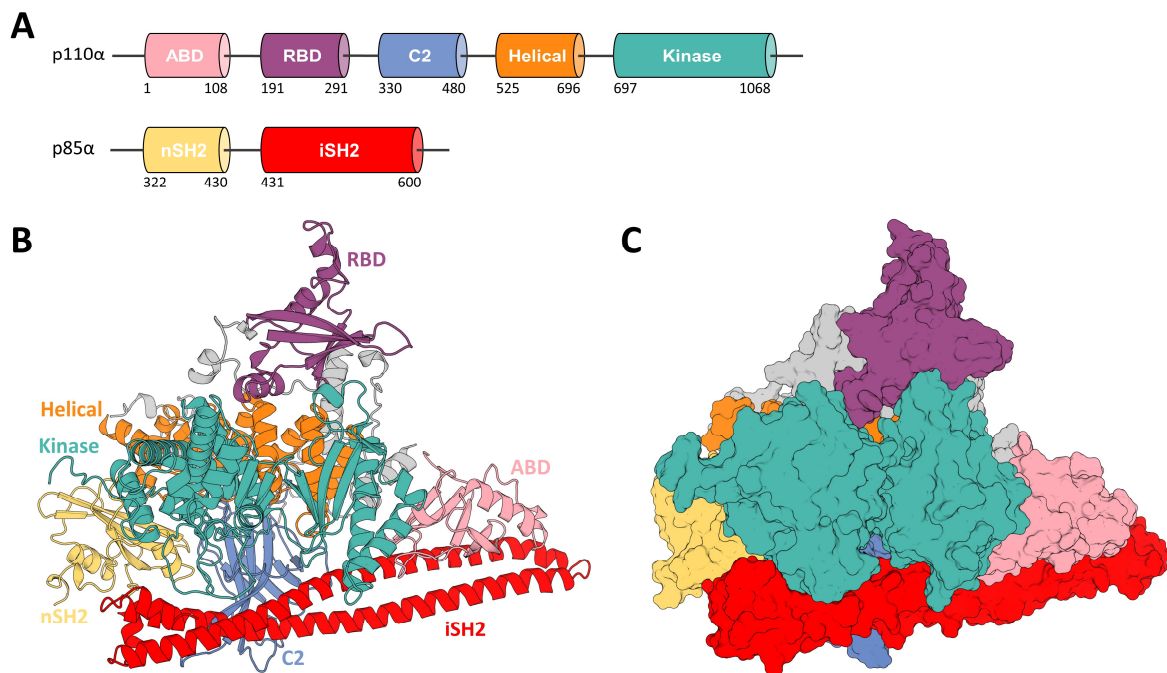


Figure I.7: Overview of the p110α/niSH2 heterodimer. (A) Schematic representation of domain organization. Gray regions are linkers between domains. (B) Diagram illustrating the structure of the p110α/niSH2 heterodimer. (C) Surface representation of the p110α/niSH2 heterodimer, shown from an alternative perspective. Generated using PyMOL (PDB ID: 4L1B [77]).

The p110α catalytic subunit comprises five domains: an adaptor-binding domain (ABD), a Ras-binding domain1(RBD), a C2 domain, a helical domain, and a kinase domain. The p85α regulatory subunit contains multiple domains, including the N-terminal SH2(nSH2), and inter-SH2 (iSH2) domains, which interact with p110α and inhibit its catalytic activity. Binding of RAS proteins to the RBD domain leads to conformational changes that relieve the inhibitory effect of p85α on the p110α subunit, leading to the activation of PI3Kα and facilitating its recruitment to the plasma membrane where it can catalyze the conversion of PIP2 to PIP3 [78].
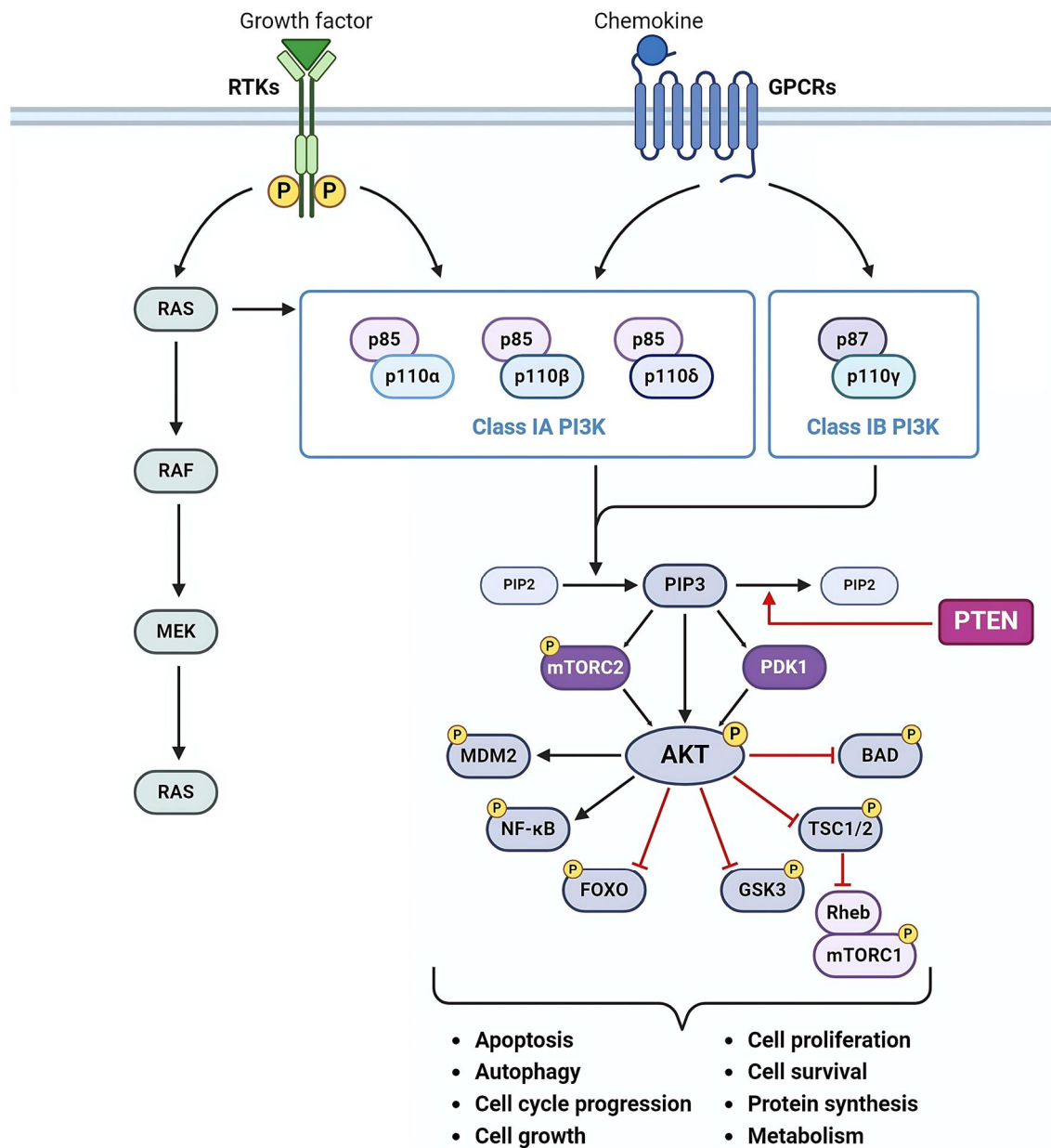
Figure I.8: Schematic representation of the PI3K/AKT/mTOR signaling cascade. Created with BioRender.com

Class I PI3Ks are activated in response to extracellular signals such as growth factors through receptor tyrosine kinases, G protein-coupled receptors, or GTPases. Following activation, these enzymes translocate to the cell membrane, where they catalyze the conversion of phosphatidylinositol 4,5-bisphosphate (PIP2) to phosphatidylinositol (3,4,5)-trisphosphate (PIP3) [79].

PIP3 serves as a second messenger, attracting proteins with pleckstrin homology domains, such as Akt. Upon membrane recruitment, Akt is phosphorylated at Threonine 308

by phosphoinositide-dependent kinase 1 (PDK1) and at Serine 473 by the mammalian target of rapamycin complex 2 (mTORC2). This phosphorylation activates Akt, which then phosphorylates various downstream targets involved in regulating metabolism, growth, proliferation, survival, and apoptosis. The activity of this signaling cascade is negatively regulated by the lipid phosphatase PTEN (phosphatase and tensin homolog), which dephosphorylates PIP3 back to PIP2, thereby attenuating Akt activation [76,80].

In PDAC, reduced PI3K expression has been linked to improved survival rates, whereas increased PI3K expression is associated with poor prognosis, tumor recurrence, and chemotherapy resistance [81]. Consequently, targeting PI3Ks with small-molecule inhibitors presents a promising therapeutic strategy for treating pancreatic cancer and overcoming treatment resistance.

### I.2.7.1 PI3Kα as a potential target for PDAC treatment

Activated RAS proteins can interact directly with the catalytic subunits p110α, p110γ, and p110δ of Class I PI3Ks, whereas the p110β subunit is primarily regulated by Rho family GTPases, specifically RAC1 and CDC42. Consequently, KRAS activation initiates signal transduction through the PI3K/AKT/mTOR pathway via the PI3Kα, PI3Kγ, and PI3Kδ isoforms. Among these isoforms, PI3Kα stands out as the most promising therapeutic target for pancreatic cancer treatment. This is largely due to the differential tissue distribution of these isoforms. PI3Kγ and PI3Kδ are predominantly expressed in hematopoietic cells, which limits their role to hematologic malignancies such as certain types of leukemias and lymphomas. In contrast, PI3Kα exhibits ubiquitous expression across a broad range of tissues. This widespread distribution makes PI3Kα particularly relevant in the context of solid tumors driven by KRAS mutations [82].

In pancreatic cancer, PI3Kα activation is essential for the migratory ability of pancreatic cancer cells, regardless of the genetic landscape of the tumor. This kinase promotes the reorganization of the actin cytoskeleton, facilitating cell motility and invasive capabilities, which are crucial for metastasis. Selective inhibition of PI3Kα using inhibitors like A66 and BYL-719 (Alpelisib) has demonstrated significant results, effectively reducing the migratory and motility capacities of pancreatic cancer cells in a concentration-dependent manner. Targeting PI3Kα also disrupts the maintenance of early pancreatic cancer lesions induced by oncogenic KRAS, especially under inflammatory conditions. Thus, therapeutic strategies

focusing on PI3Kα can potentially hinder pancreatic cancer metastasis and progression, offering a targeted approach to managing this aggressive malignancy [83].

## I.3    Natural products in drug design

### I.3.1    From historical milestones to modern rediscovery

The renewed interest in natural products as sources for new drugs is deeply rooted in the historical success of several medications derived from nature that have profoundly impacted modern medicine. For instance, Salicin, derived from willow bark (*Salix alba*), led to the synthesis of aspirin in 1899 by Felix Hoffmann at Bayer, becoming one of the most widely used medications for pain, fever, and inflammation. Similarly, Penicillin, discovered by Alexander Fleming in 1928 from the mold *Penicillium notatum*, initiated the antibiotic era. Its development by Howard Florey and Ernst Boris Chain in the early 1940s enabled mass production, drastically reducing bacterial infection mortality. In 1967, Monroe Wall and Mansukh Wani isolated paclitaxel from the Pacific yew tree (*Taxus brevifolia*), introducing a novel cancer treatment by stabilizing microtubules. The FDA approved paclitaxel for ovarian cancer in 1992 and for breast cancer in 1994 [84].

Cyclosporin A, discovered in 1971 from the soil fungus *Tolypocladium inflatum* by Sandoz Laboratories, revolutionized organ transplantation with its immunosuppressive properties, significantly reducing rejection rates since its clinical introduction in 1983. The isolation of lovastatin from *Aspergillus terreus* by Merck scientists in the late 1970s led to its FDA approval in 1987. Statins have since been critical in managing cholesterol levels and preventing cardiovascular diseases [85].

Newman and Cragg's review of drug sources from 1981 to 2019 reveals that 50.6% of the 1,881 new chemical entities (NCEs) approved by the FDA during that period are synthetic [86]. However, a significant portion is derived from natural products: 14.5% are synthetic drugs based on natural product pharmacophores, and 11.5% are inspired by natural products, designed to mimic the pharmacological effects of natural substances and competitively inhibit the endogenous substrate of the active site, such as ATP. Thus, only 24.6% of the 1,881 NCEs can be classified as truly synthetic in origin.
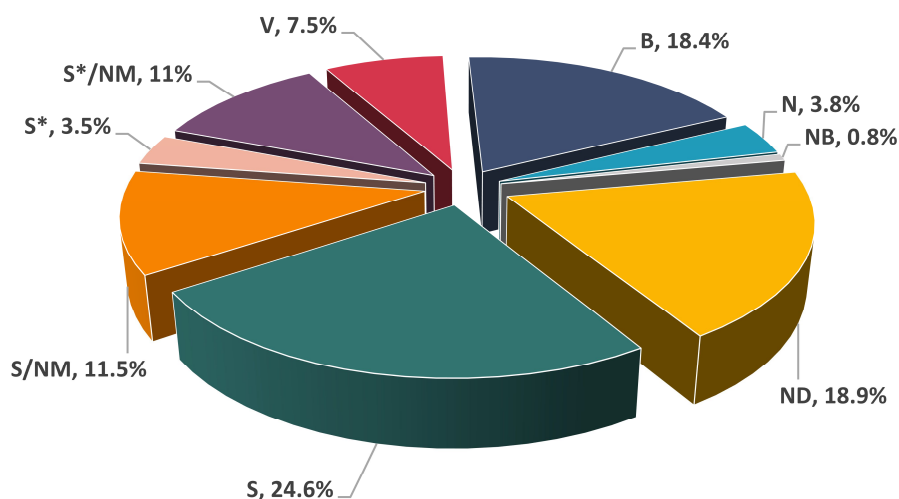
Figure I.9: Drugs approved for all diseases by the FDA or their equivalents in other countries from January 1, 1981, to September 30, 2019 (n = 1,881). Adapted from [86]. B: biological macromolecule; N: unaltered natural product; NB: botanical drug (defined mixture); ND: natural product derivative; S: synthetic drug; S*: synthetic drug modeled after the pharmacophore of a natural product; V: vaccine; NM: natural product mimic.

During the 1980s and 1990s, numerous major pharmaceutical firms aggressively screened plant samples for potential drug compounds. Nonetheless, after about a decade or so, many of these companies abandoned this strategy. They transitioned towards exploring libraries of synthetic chemicals, combinatorial chemistry, and alternative methods [87]. This transition coincided with a period in which the pharmaceutical industry experienced a downturn in research and development productivity, marked by a significant decline in the number of new chemical entities. By 2001, the industry reached a 20-year low, with only 37 new NCEs identified. This decline was accompanied by a reduction in new drug applications to the FDA, which fell from 24 in the previous year to 16 in 2001.

This downturn in productivity was attributed to several factors, including the decreasing interest in natural products by major pharmaceutical companies. As the industry faced the stark realization that the number of new chemical entities in their development pipelines was dwindling, a renewed interest emerged in rediscovering natural products as potential sources for new drugs [88].

Multiple reasons underlie the enduring relevance of natural products in modern medicine despite the rise of synthetic drug production. One reason, discussed by Larsson and colleagues [89], is that natural products have undergone pre-validation and fine-tuning by evolutionary forces over millions of years, rendering them inherently well-suited for

biological systems. Another perspective posits that natural products synthesized through intricate enzymatic pathways offer access to chemical space beyond the reach of conventional organic synthesis methods [90].

This underscores the invaluable role that natural products have played and continue to play in drug discovery, providing unique structural diversity and serving as sources of innovative pharmaceutical therapeutic agents. Their contribution to the development of new drugs remains significant, especially in the treatment of cancer, where more than 60% of the approved anticancer drugs originate from natural sources [91].

### I.3.2    Curcumin : The Indian solid gold

Curcumin, the primary active component of the dietary spice turmeric, is extracted from the rhizomes of *Curcuma longa L.*, a member of the ginger family Zingiberaceae. The identification of curcumin dates back approximately two centuries ago, when Vogel and Pelletier first isolated a "yellow coloring-matter" from the rhizomes of Curcuma longa, naming it curcumin. In 1842, Vogel Jr. managed to obtain a pure preparation of curcumin. Decades later, in 1910, Melabedzka and Lampe reported the structure of curcumin as diferuloylmethane, or 1,6-heptadiene-3,5-dione-1,7-bis (4-hydroxy-3-methoxyphenyl)-(1E,6E). Further work by the same group in 1913 led to the successful synthesis of the compound [92].

### I.3.2.1   Occurrence and physicochemical properties

Turmeric extracts generally contain 1−6% curcuminoids and 3−7% volatile essential oils. Among the curcuminoids, curcumin is the most abundant and extensively studied bioactive component, making up 60−70% of this group. This is followed by demethoxycurcumin (DMC) at 20−27% and bisdemethoxycurcumin (BDMC) at 10−15%. Curcumin appears as an orange-yellow crystalline powder with the chemical formula $C_{21}H_{20}O_6$, a molecular weight of 368.38 g/mol, and a melting point of 183 °C. Due to its hydrophobic nature, curcumin has an octanol/water partition coefficient (logP) of 3.2, which makes it insoluble in water but soluble in several organic solvents, including chloroform, ethanol, methanol, acetonitrile, ethyl acetate, and dimethyl sulfoxide.
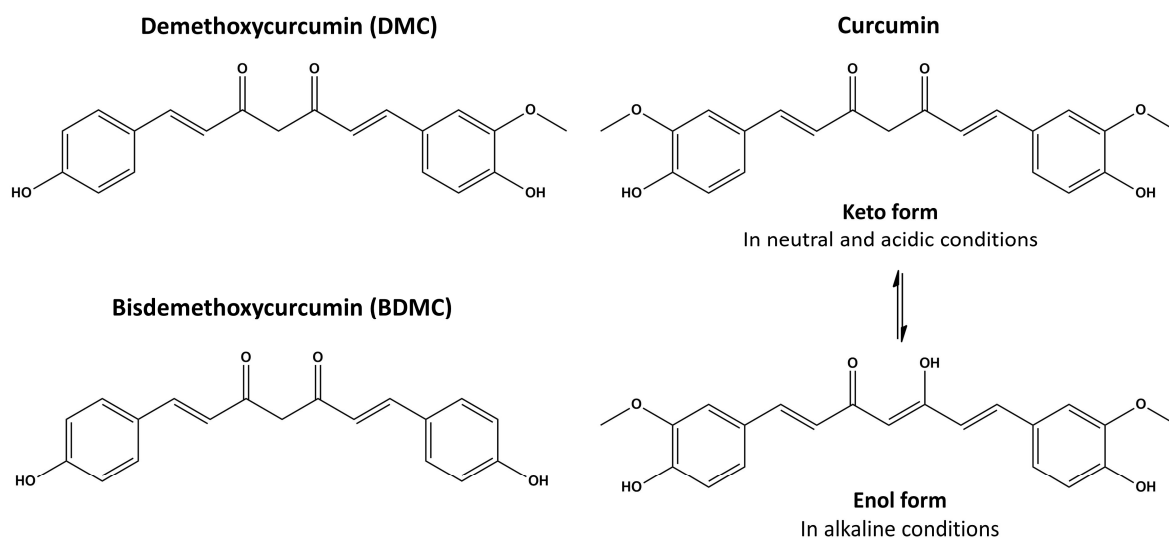
Figure I.10: Chemical structures of curcumin, demethoxycurcumin and bisdemethoxycurcumin.

The chemical structure of curcumin is characterized by two aromatic rings, each bearing a hydroxy and a methoxy group. These rings are bridged by a seven-carbon chain that features an α,β-unsaturated β-diketone moiety. This particular moiety exhibits keto-enol tautomerization, where the keto form predominates under neutral and acidic conditions, while the enol form is more stable in alkaline solutions [93,94].

### I.3.2.2   Curcumin in cancer treatment

Curcumin has shown considerable potential in cancer therapy through its modulation of various critical pathways and molecular targets. Its effects extend to cytokines, protein kinases, growth factors, transcription factors, and microRNAs [95], resulting in multiple mechanisms that impede cancer progression and enhance therapeutic outcomes.

- **Antioxidant and anti-inflammatory properties:** Curcumin exhibits antioxidant effects by effectively neutralizing free radicals and boosting the activity of antioxidant enzymes. This action helps mitigate oxidative stress and prevent DNA damage, which can contribute to the onset of cancer [96]. Additionally, curcumin's anti-inflammatory properties are largely attributed to its inhibition of Nuclear Factor kappa B (NF-κB). NF-κB is a key transcription factor involved in regulating inflammatory processes. By inhibiting NF-κB, curcumin reduces the levels of pro-inflammatory cytokines and disrupts the inflammatory microenvironment that typically supports tumor growth and progression [10].

- **Cell cycle arrest and apoptosis:** Dysregulation of the cell cycle is a hallmark of cancer, resulting in uncontrolled cellular proliferation and tumor growth. Curcumin exhibits anti-cancer properties by interfering with cell cycle progression, particularly at the G2/M phase transition. This interference is mediated through the modulation of cyclins and cyclin-dependent kinases, key regulators of cell cycle checkpoints. The consequent cell cycle arrest attenuates cancer cell proliferation, thereby inhibiting tumor growth [12].

  Evasion of apoptosis is another characteristic of cancer cells that contributes to their survival and proliferation. Curcumin induces apoptosis through both the intrinsic (mitochondrial) and extrinsic (death receptor) pathways. This is accomplished by upregulating pro-apoptotic proteins such as Bax while simultaneously downregulating anti-apoptotic proteins like Bcl-2. Furthermore, curcumin activates caspases, which are critical executioners of the apoptotic process. Through these mechanisms, curcumin leads to cellular degradation and death [11].

- **Inhibition of cancer cell migration and invasion:** Curcumin demonstrates anti-metastatic properties by modulating key signaling pathways, including the PI3K/Akt cascade. This modulation impedes critical processes involved in metastasis, such as epithelial-mesenchymal transition (EMT) and the activity of matrix metalloproteinases (MMPs) [97]. EMT facilitates the acquisition of migratory and invasive capabilities by cancer cells through the loss of epithelial characteristics and the gain of mesenchymal traits. By inhibiting EMT, curcumin prevents the detachment of cancer cells from the primary tumor and their subsequent invasion of distant tissues [98]. Additionally, curcumin reduces the expression and activity of MMPs, enzymes that degrade the extracellular matrix, further limiting the invasive potential of cancer cells [99].

- **Anti-angiogenic properties:** Curcumin demonstrates anti-angiogenic effects by inhibiting the vascular endothelial growth factor (VEGF) signaling pathway. This inhibition impedes the development of new blood vessels that supply tumors with essential nutrients and oxygen, ultimately leading to tumor starvation and restricting its growth [96].
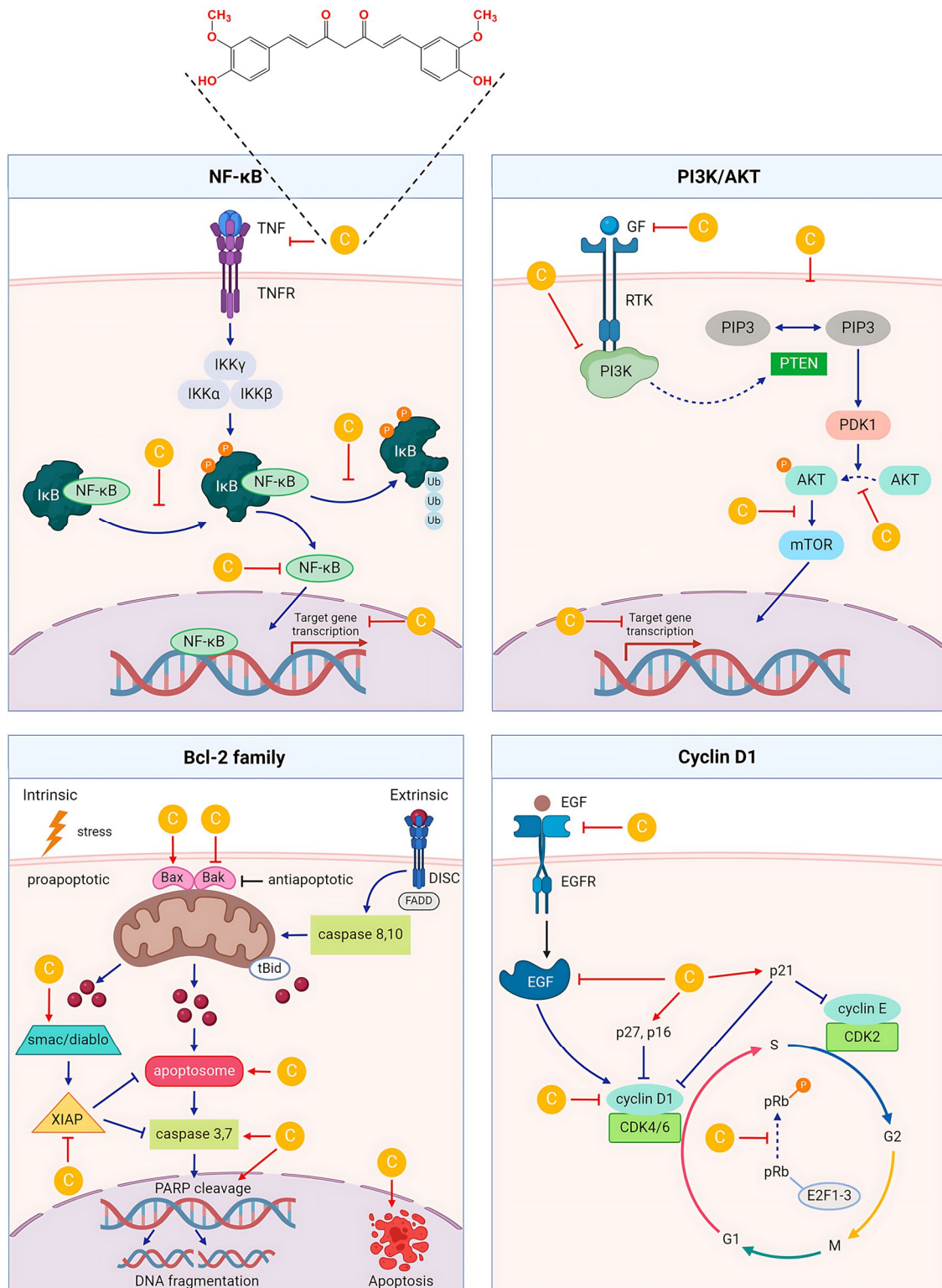
Figure I.11: Molecular mechanisms of curcumin-induced anti-cancer effects on key signaling pathways. Created with BioRender.com

- **Synergistic effects:** Research has demonstrated that curcumin can enhance the efficacy of various chemotherapeutic agents through multiple mechanisms, such as increasing

apoptosis, inhibiting proliferation, modulating critical signaling pathways, and reducing drug resistance. These synergistic effects have been observed in various cancer types, highlighting curcumin's potential as an adjunctive therapy. For instance, combining curcumin with mitomycin C significantly enhanced apoptosis in MCF-7 breast cancer stem-like cells. Similarly, the combination of curcumin with bortezomib effectively inhibited proliferation and increased apoptosis in human multiple myeloma MM1R cells. Additionally, when used in conjunction with gemcitabine, curcumin demonstrated enhanced anti-proliferative and pro-apoptotic effects, as well as reduced cell migration and invasion, particularly in pancreatic ductal adenocarcinoma cells [100].

### I.3.2.3　Curcumin bioavailability

Despite the well-documented benefits of curcumin, its approval as a therapeutic agent remains elusive, primarily due to its poor bioavailability. This issue arises from several factors, including poor water solubility, limited absorption, low tissue distribution, rapid metabolism, and swift elimination from the body [101]. To tackle these issues, a range of approaches has been implemented to optimize the pharmacokinetic properties and cellular uptake of curcumin. These strategies encompass the synthesis of curcumin derivatives and the design of advanced drug delivery systems to enhance its solubility and prolong its systemic circulation time.

➤ **Curcumin derivatives**

Chemical modifications of curcumin are extensively employed to enhance its pharmacological properties. Structural alterations of the phenolic rings, including glycosylation, acylation, aminoacylation, and alkylation of hydroxyl groups, are effective methods to improve the solubility and stability of curcumin. Glycosylation, in particular, has been demonstrated to significantly increase water solubility and kinetic stability, resulting in an enhanced therapeutic response [101]. Notably, monofunctionalized derivatives often exhibit better bioactivity compared to their difunctionalized counterparts [102].

The C7 linker has also been a target for modification, with hydrogenation of double bonds and carbonyl groups yielding analogues such as dihydrocurcumin, tetrahydrocurcumin, hexahydrocurcumin, and octahydrocurcumin. While these derivatives generally demonstrate enhanced antioxidant activity, their impact on apoptosis induction may be compromised [11].

The central methylene group can also undergo alkylation, acylation, or arylidene substitution. Additionally, the β-diketone moiety has been exploited for its reactivity with compounds like hydroxylamine and hydrazine to form bis(styryl)pyrazoles and isoxazoles. These modifications effectively mask and rigidify the central 1,3-diketone system, potentially enhancing the therapeutic properties of curcumin by increasing its stability and altering its reactivity profile [103].
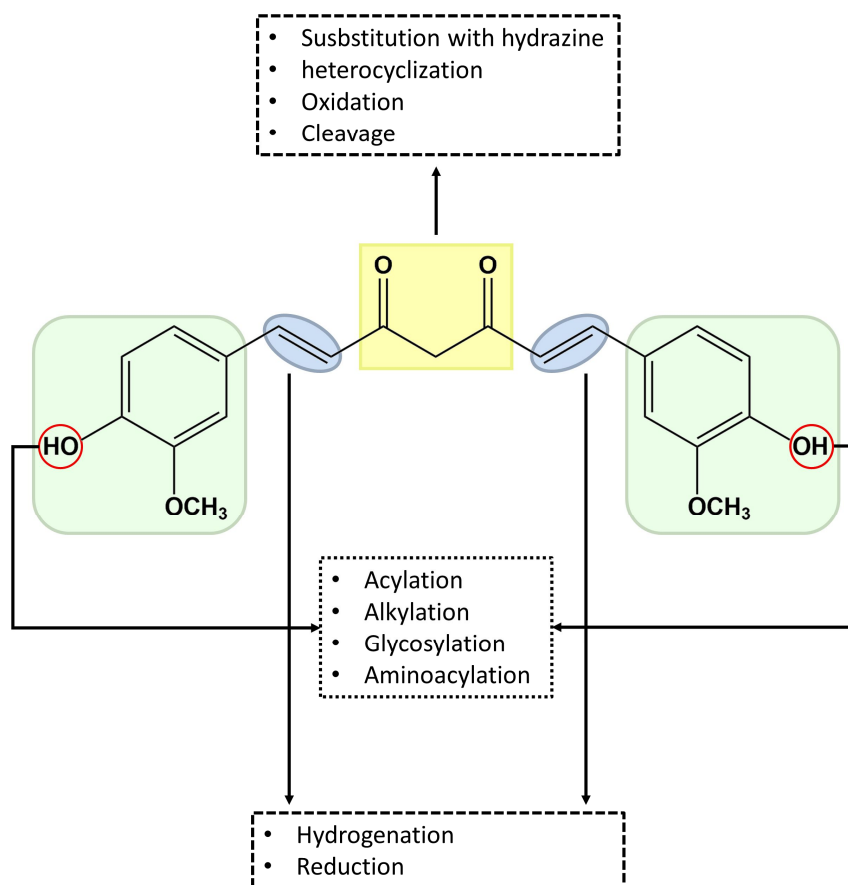


Figure I.12: Structural modifications that can be performed to improve the physicochemical and pharmacological properties of curcumin. Adapted from [101].

An innovative approach involves metal complexation. The α,β-unsaturated β-diketo moiety of curcumin is recognized for its strong chelating properties, enabling it to interact with a variety of metal ions, including platinum, copper, zinc, gold, vanadium, boron, silver, lanthanum, cobalt, ruthenium, gallium, iron, manganese, gadolinium, nickel, and palladium. Research indicates that curcumin-metal complexes exhibit enhanced solubility, stability, bioavailability, and biochemical activities compared to free curcumin [104]. Additionally, altering the central carbon chain length has also yielded promising results [105]. For

instance, diarylpentanoid curcumin analogues, featuring two aromatic rings joined by five carbon atoms, have shown more potent antitumor activity than curcumin by modulating multiple signaling pathways such as NF-κB, AKT-PTEN, STAT, and MAPK [106].

➤ **Curcumin delivery systems**

Key strategies to improve curcumin's pharmacokinetics include:

a. **Delaying metabolism:** Curcumin can be protected from rapid metabolism by encapsulation within hydrophobic matrices, which shields it from interaction with aqueous environments and metabolic enzymes.

b. **Enhancing bioaccessibility:** Increasing the amount of curcumin dissolved in mixed micelles within the small intestine by adding surfactants, fatty acids, monoglycerides, or phospholipids to curcumin-loaded carrier particles.

c. **Facilitating absorption:** By loading curcumin into delivery systems that contain agents to improve epithelial cell membrane permeability or inhibit efflux transporters [101].

Various formulations have been employed to enhance the pharmacokinetic properties of curcumin, including solid lipid nanoparticles, micelles, liposomes, magnetic nanoparticles, solid dispersions, cyclodextrin complexes, and phytosomes. Additionally, conjugating curcumin with bioenhancers, such as piperine, which inhibits its metabolism, has been shown to significantly increase its serum concentration [107,108].

# References

[1] M.S.D. De La Cruz, A.P. Young and M.T. Ruffin, *Diagnosis and management of pancreatic cancer*, Am. Fam. Physician 89 (2014), pp. 626–632.

[2] O.T. Olaoba, T.I. Adelusi, M. Yang, T. Maidens, E.T. Kimchi, K.F. Staveley-O'Carroll et al., *Driver Mutations in Pancreatic Cancer and Opportunities for Targeted Therapy*, Cancers 16 (2024), pp. 1808.

[3] R.L. Siegel, A.N. Giaquinto and A. Jemal, *Cancer statistics, 2024*, CA. Cancer J. Clin. 74 (2024), pp. 12–49.

[4] A.M. Waters and C.J. Der, *KRAS: The Critical Driver and Therapeutic Target for Pancreatic Cancer*, Cold Spring Harb. Perspect. Med. 8 (2018), pp. a031435.

[5] Z. Zhang, H. Zhang, X. Liao and H. Tsai, *KRAS mutation: The booster of pancreatic ductal adenocarcinoma transformation and progression*, Front. Cell Dev. Biol. 11 (2023).

[6] L. Buscail, B. Bournet and P. Cordelier, *Role of oncogenic KRAS in the diagnosis, prognosis and treatment of pancreatic cancer*, Nat. Rev. Gastroenterol. Hepatol. 17 (2020), pp. 153–168.

[7] C. Zhu, X. Guan, X. Zhang, X. Luan, Z. Song, X. Cheng et al., *Targeting KRAS mutant cancers: from druggable therapy to drug resistance*, Mol. Cancer 21 (2022), pp. 159.

[8] A. Glaviano, A.S.C. Foo, H.Y. Lam, K.C.H. Yap, W. Jacot, R.H. Jones et al., *PI3K/AKT/mTOR signaling transduction pathway and targeted therapies in cancer*, Mol. Cancer 22 (2023), pp. 138.

[9] S. Stanciu, F. Ionita-Radu, C. Stefani, D. Miricescu, I.-I. Stanescu-Spinu, M. Greabu et al., *Targeting PI3K/AKT/mTOR Signaling Pathway in Pancreatic Cancer: From Molecular to Clinical Aspects*, Int. J. Mol. Sci. 23 (2022), pp. 10132.

[10] D.-O. Moon, *Curcumin in Cancer and Inflammation: An In-Depth Exploration of Molecular Interactions, Therapeutic Potentials, and the Role in Disease Management*, Int. J. Mol. Sci. 25 (2024), pp. 2911.

[11] M.A. Tomeh, R. Hadianamrei and X. Zhao, *A Review of Curcumin and Its Derivatives as Anticancer Agents*, Int. J. Mol. Sci. 20 (2019), pp. 1033.

[12] S. Wang, X. Gao, J. Li, S. Wei, Y. Shao, Y. Yin et al., *The anticancer effects of curcumin and clinical research progress on its effects on esophageal cancer*, Front. Pharmacol. 13 (2022).

[13] J. Dolenšek, M.S. Rupnik and A. Stožer, *Structural similarities and differences between the human and the mouse pancreas*, Islets 7 (2015), pp. e1024405.

[14] L.E. Oldfield, A.A. Connor and S. Gallinger, *Molecular Events in the Natural History of Pancreatic Cancer*, Trends Cancer 3 (2017), pp. 336–346.

[15] Q. Zhou and D.A. Melton, *Pancreas regeneration*, Nature 557 (2018), pp. 351–358.

[16] D. Ansari, B. Tingstedt, B. Andersson, F. Holmquist, C. Sturesson, C. Williamsson et al., *Pancreatic cancer: yesterday, today and tomorrow*, Future Oncol. Lond. Engl. 12 (2016), pp. 1929–1946.

[17] J.F. Griffin, K.E. Poruk and C.L. Wolfgang, *Pancreatic cancer surgery: past, present, and future*, Chin. J. Cancer Res. 27 (2015), pp. 332–348.

[18] P.D. Chatani, S.K. Agarwal and S.M. Sadowski, *Molecular Signatures and Their Clinical Utility in Pancreatic Neuroendocrine Tumors*, Front. Endocrinol. 11 (2020), pp. 575620.

[19] M.J. Pishvaian and J.R. Brody, *Therapeutic Implications of Molecular Subtyping for Pancreatic Cancer*, Oncol. Williston Park N 31 (2017), pp. 159–166, 168.

[20] P. Rawla, T. Sunkara and V. Gaduputi, *Epidemiology of Pancreatic Cancer: Global Trends, Etiology and Risk Factors*, World J. Oncol. 10 (2019), pp. 10–27.

[21] B.J. Kenner, *Early Detection of Pancreatic Cancer*, Pancreas 47 (2018), pp. 363–367.

[22] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R.L. Siegel, I. Soerjomataram et al., *Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries*, CA. Cancer J. Clin. 74 (2024), pp. 229–263.

[23] Z. Li and I. Salik, *A Dangerous Mimic: Chronic Pancreatitis Masquerading As Pancreatic Adenocarcinoma*, Cureus 13 (2021), pp. e19795.

[24] M. Malvezzi, G. Carioli, P. Bertuccio, T. Rosso, P. Boffetta, F. Levi et al., *European cancer mortality predictions for the year 2016 with focus on leukaemias*, Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. 27 (2016), pp. 725–731.

[25] C. Bosetti, P. Bertuccio, E. Negri, C. La Vecchia, M.P. Zeegers and P. Boffetta, *Pancreatic cancer: overview of descriptive epidemiology*, Mol. Carcinog. 51 (2012), pp. 3–13.

[26] A.P. Klein, *Pancreatic cancer epidemiology: understanding the role of lifestyle and inherited risk factors*, Nat. Rev. Gastroenterol. Hepatol. 18 (2021), pp. 493–502.

[27] R.U. Harvitkar, H. Peri, S.N. Zallipalli, S.J. Joseph, G.B. Gattupalli and K. Ansari, *Non-Cancer Causes of Death in Patients With Pancreatic Adenocarcinoma: A Surveillance, Epidemiology, and End Results (SEER)-Based Study*, Cureus 13 (2021), pp. e20289.

[28] M. Ezzati, S.J. Henley, A.D. Lopez and M.J. Thun, *Role of smoking in global and regional cancer epidemiology: Current patterns and data needs*, Int. J. Cancer 116 (2005), pp. 963–971.

[29] I. Kuzmickiene, R. Everatt, D. Virviciute, A. Tamosiunas, R. Radisauskas, R. Reklaitiene et al., *Smoking and other risk factors for pancreatic cancer: a cohort study in men in Lithuania*, Cancer Epidemiol. 37 (2013), pp. 133–139.

[30] C. Pelucchi, C. Galeone, J. Polesel, M. Manzari, A. Zucchetto, R. Talamini et al., *Smoking and body mass index and survival in pancreatic cancer patients*, Pancreas 43 (2014), pp. 47–52.

[31] T. Rastogi, S. Devesa, P. Mangtani, A. Mathew, N. Cooper, R. Kao et al., *Cancer incidence rates among South Asians in four geographic regions: India, Singapore, UK and US*, Int. J. Epidemiol. 37 (2008), pp. 147–160.

[32]  S. Iodice, S. Gandini, P. Maisonneuve and A.B. Lowenfels, *Tobacco and the risk of pancreatic cancer: a review and meta-analysis*, Langenbecks Arch. Surg. 393 (2008), pp. 535–545.

[33]  A. Vrieling, H.B. Bueno-de-Mesquita, H.C. Boshuizen, D.S. Michaud, M.T. Severinsen, K. Overvad et al., *Cigarette smoking, environmental tobacco smoke exposure and pancreatic cancer risk in the European Prospective Investigation into Cancer and Nutrition*, Int. J. Cancer 126 (2010), pp. 2394–2403.

[34]  S.H. Davoodi, T. Malek-Shahabi, A. Malekshahi-Moghadam, R. Shahbazi and S. Esmaeili, *Obesity as an Important Risk Factor for Certain Types of Cancer*, Iran. J. Cancer Prev. 6 (2013), pp. 186–194.

[35]  A. Berrington de Gonzalez, S. Sweetland and E. Spencer, *A meta-analysis of obesity and the risk of pancreatic cancer*, Br. J. Cancer 89 (2003), pp. 519–523.

[36]  E.E. Calle, C. Rodriguez, K. Walker-Thurmond and M.J. Thun, *Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults*, N. Engl. J. Med. 348 (2003), pp. 1625–1638.

[37]  D. Li, J.S. Morris, J. Liu, M.M. Hassan, R.S. Day, M.L. Bondy et al., *Body mass index and risk, age of onset, and survival in patients with pancreatic cancer*, JAMA 301 (2009), pp. 2553–2562.

[38]  R. Huxley, A. Ansary-Moghaddam, A. Berrington de González, F. Barzi and M. Woodward, *Type-II diabetes and pancreatic cancer: a meta-analysis of 36 studies*, Br. J. Cancer 92 (2005), pp. 2076–2083.

[39]  C. Bosetti, V. Rosato, D. Li, D. Silverman, G.M. Petersen, P.M. Bracci et al., *Diabetes, antidiabetic medications, and pancreatic cancer risk: an analysis from the International Pancreatic Cancer Case-Control Consortium*, Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. 25 (2014), pp. 2065–2072.

[40]  J.W. Elena, E. Steplowski, K. Yu, P. Hartge, G.S. Tobias, M.J. Brotzman et al., *Diabetes and risk of pancreatic cancer: a pooled analysis from the pancreatic cancer cohort consortium*, Cancer Causes Control CCC 24 (2013), pp. 13–25.

[41]  A. Ekbom, J.K. McLaughlin, B.M. Karlsson, O. Nyrén, G. Gridley, H.O. Adami et al., *Pancreatitis and pancreatic cancer: a population-based study*, J. Natl. Cancer Inst. 86 (1994), pp. 625–627.

[42]  S. Raimondi, A.B. Lowenfels, A.M. Morselli-Labate, P. Maisonneuve and R. Pezzilli, *Pancreatic cancer in chronic pancreatitis; aetiology, incidence, and early detection*, Best Pract. Res. Clin. Gastroenterol. 24 (2010), pp. 349–358.

[43]  E.J. Duell, E. Lucenteforte, S.H. Olson, P.M. Bracci, D. Li, H.A. Risch et al., *Pancreatitis and pancreatic cancer risk: a pooled analysis in the International Pancreatic Cancer Case-Control Consortium (PanC4)*, Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. 23 (2012), pp. 2964–2970.

[44]  A. Stathis and M.J. Moore, *Advanced pancreatic carcinoma: current treatment and future challenges*, Nat. Rev. Clin. Oncol. 7 (2010), pp. 163–172.

[45]  J.P. Neoptolemos, D.D. Stocken, H. Friess, C. Bassi, J.A. Dunn, H. Hickey et al., *A randomized trial of chemoradiotherapy and chemotherapy after resection of pancreatic cancer*, N. Engl. J. Med. 350 (2004), pp. 1200–1210.

[46] H. Oettle, S. Post, P. Neuhaus, K. Gellert, J. Langrehr, K. Ridwelski et al., *Adjuvant chemotherapy with gemcitabine vs observation in patients undergoing curative-intent resection of pancreatic cancer: a randomized controlled trial*, JAMA 297 (2007), pp. 267–277.

[47] W.F. Regine, K.A. Winter, R.A. Abrams, H. Safran, J.P. Hoffman, A. Konski et al., *Fluorouracil vs gemcitabine chemotherapy before and after fluorouracil-based chemoradiation following resection of pancreatic adenocarcinoma: a randomized controlled trial*, JAMA 299 (2008), pp. 1019–1026.

[48] M. Hidalgo, *Pancreatic cancer*, N. Engl. J. Med. 362 (2010), pp. 1605–1617.

[49] T. Conroy, P. Hammel, M. Hebbar, M. Ben Abdelghani, A.C. Wei, J.-L. Raoul et al., *FOLFIRINOX or Gemcitabine as Adjuvant Therapy for Pancreatic Cancer*, N. Engl. J. Med. 379 (2018), pp. 2395–2406.

[50] T. Kamisawa, L.D. Wood, T. Itoi and K. Takaori, *Pancreatic cancer*, Lancet Lond. Engl. 388 (2016), pp. 73–85.

[51] H.A. Burris, M.J. Moore, J. Andersen, M.R. Green, M.L. Rothenberg, M.R. Modiano et al., *Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial*, J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 15 (1997), pp. 2403–2413.

[52] M. Suker, B.R. Beumer, E. Sadot, L. Marthey, J.E. Faris, E.A. Mellon et al., *FOLFIRINOX for locally advanced pancreatic cancer: a systematic review and patient-level meta-analysis*, Lancet Oncol. 17 (2016), pp. 801–810.

[53] T. Conroy, F. Desseigne, M. Ychou, O. Bouché, R. Guimbaud, Y. Bécouarn et al., *FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer*, N. Engl. J. Med. 364 (2011), pp. 1817–1825.

[54] J. Cicenas, K. Kvederaviciute, I. Meskinyte, E. Meskinyte-Kausiliene, A. Skeberdyte and J. Cicenas, *KRAS, TP53, CDKN2A, SMAD4, BRCA1, and BRCA2 Mutations in Pancreatic Cancer*, Cancers 9 (2017), pp. 42.

[55] S.S. Wang, J. Xu, K.Y. Ji and C.-I. Hwang, *Epigenetic Alterations in Pancreatic Cancer Metastasis*, Biomolecules 11 (2021), pp. 1082.

[56] C.A. Iacobuzio-Donahue, V.E. Velculescu, C.L. Wolfgang and R.H. Hruban, *Genetic basis of pancreas cancer development and progression: insights from whole-exome and whole-genome sequencing*, Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 18 (2012), pp. 4257–4265.

[57] R.H. Hruban, K. Takaori, D.S. Klimstra, N.V. Adsay, J. Albores-Saavedra, A.V. Biankin et al., *An illustrated consensus on the classification of pancreatic intraepithelial neoplasia and intraductal papillary mucinous neoplasms*, Am. J. Surg. Pathol. 28 (2004), pp. 977–987.

[58] M. Distler, D. Aust, J. Weitz, C. Pilarsky and R. Grützmann, *Precursor Lesions for Sporadic Pancreatic Cancer: PanIN, IPMN, and MCN*, BioMed Res. Int. 2014 (2014), pp. 474905.

[59] S. Sankarasubramanian, U. Pfohl, C.R.A. Regenbrecht, C. Reinhard and L. Wedeken, *Context Matters—Why We Need to Change From a One Size Fits all Approach to Made-to-Measure Therapies for Individual Patients With Pancreatic Cancer*, Front. Cell Dev. Biol. 9 (2021), pp. 760705.

[60] A. Argentiero, A. Andriano, I.C. Caradonna, G. de Martino and V. Desantis, *Decoding the Intricate Landscape of Pancreatic Cancer: Insights into Tumor Biology, Microenvironment, and Therapeutic Interventions*, Cancers 16 (2024), pp. 2438.

[61] L. Huang, Z. Guo, F. Wang and L. Fu, *KRAS mutation: from undruggable to druggable in cancer*, Signal Transduct. Target. Ther. 6 (2021), pp. 386.

[62] C.A. Stalnecker and C.J. Der, *RAS, wanted dead or alive: Advances in targeting RAS-mutant cancers*, Sci. Signal. 13 (2020), pp. eaay6013.

[63] A. Ferreira, F. Pereira, C. Reis, M.J. Oliveira, M.J. Sousa and A. Preto, *Crucial Role of Oncogenic KRAS Mutations in Apoptosis and Autophagy Regulation: Therapeutic Implications*, Cells 11 (2022), pp. 2183.

[64] A.D. Cox and M.J. DeCristo, *Ras*, in *Molecular Oncology: Causes of Cancer and Targets for Treatment*, C.L. Sawyers, E.P. Gelmann and I. Rauscher Frank J., eds., Cambridge University Press, Cambridge, 2013, pp. 258–271.

[65] S. Eser, A. Schnieke, G. Schneider and D. Saur, *Oncogenic KRAS signalling in pancreatic cancer*, Br. J. Cancer 111 (2014), pp. 817–822.

[66] A.D. Cox, S.W. Fesik, A.C. Kimmelman, J. Luo and C.J. Der, *Drugging the undruggable Ras: mission possible?*, Nat. Rev. Drug Discov. 13 (2014), pp. 828–851.

[67] S. Dhillon, *Adagrasib: First Approval*, Drugs 83 (2023), pp. 275–285.

[68] E.C. Nakajima, N. Drezner, X. Li, P.S. Mishra-Kalyani, Y. Liu, H. Zhao et al., *FDA Approval Summary: Sotorasib for KRAS G12C Mutated Metastatic NSCLC*, Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 28 (2022), pp. 1482–1486.

[69] L. Li, J. Liu, Z. Yang, H. Zhao, B. Deng, Y. Ren et al., *Discovery of Thieno[2,3-d]pyrimidine-based KRAS G12D inhibitors as potential anticancer agents via combinatorial virtual screening*, Eur. J. Med. Chem. 233 (2022), pp. 114243.

[70] R.B. Kargbo, *Targeting KRASG12D Mutations: Discovery of Small Molecule Inhibitors for the Potential Treatment of Intractable Cancers*, ACS Med. Chem. Lett. 14 (2023), pp. 1041–1042.

[71] X. Wang, S. Allen, J.F. Blake, V. Bowcut, D.M. Briere, A. Calinisan et al., *Identification of MRTX1133, a Noncovalent, Potent, and Selective KRASG12D Inhibitor*, J. Med. Chem. 65 (2022), pp. 3123–3133.

[72] E.A. Collisson, C.L. Trejo, J.M. Silva, S. Gu, J.E. Korkola, L.M. Heiser et al., *A central role for RAF→MEK→ERK signaling in the genesis of pancreatic ductal adenocarcinoma*, Cancer Discov. 2 (2012), pp. 685–693.

[73] S. Eser, N. Reiff, M. Messer, B. Seidler, K. Gottschalk, M. Dobler et al., *Selective requirement of PI3K/PDK1 signaling for Kras oncogene-driven pancreatic cell plasticity and cancer*, Cancer Cell 23 (2013), pp. 406–420.

[74] M. Li Lung, W. Dai and J.M.-Y. Ko, *Chapter 2 - Nasopharyngeal Carcinoma: Genetics and Genomics*, in *Nasopharyngeal Carcinoma*, A.W.M. Lee, M.L. Lung and W.T. Ng, eds., Academic Press, 2019, pp. 17–44.

[75] S. Mehra, N. Deshpande and N. Nagathihalli, *Targeting PI3K Pathway in Pancreatic Ductal Adenocarcinoma: Rationale and Progress*, Cancers 13 (2021), pp. 4434.

[76]  S.S. Chopra and L.C. Cantley, *PI3K-Akt-mTOR Signaling in Cancer and Cancer Therapeutics*, in *PI3K-mTOR in Cancer and Cancer Therapy*, N. Dey, P. De and B. Leyland-Jones, eds., Springer International Publishing, Cham, 2016, pp. 1–25.

[77]  Y. Zhao, X. Zhang, Y. Chen, S. Lu, Y. Peng, X. Wang et al., *Crystal Structures of PI3Kα Complexed with PI103 and Its Derivatives: New Directions for Inhibitors Design*, ACS Med. Chem. Lett. 5 (2014), pp. 138–142.

[78]  C. Cuesta, C. Arévalo-Alameda and E. Castellano, *The Importance of Being PI3K in the RAS Signaling Network*, Genes 12 (2021), pp. 1094.

[79]  P. Singh, M.S. Dar and M.J. Dar, *p110α and p110β isoforms of PI3K signaling: are they two sides of the same coin?*, FEBS Lett. 590 (2016), pp. 3071–3082.

[80]  J.W. Harris, T. Gao and B. Mark Evers, *The role of PI3K signaling pathway in intestinal tumorigenesis*, in *Intestinal Tumorigenesis: Mechanisms of Development & Progression*, 2015, pp. 101–135.

[81]  J.R. Conway, D. Herrmann, T.J. Evans, J.P. Morton and P. Timpson, *Combating pancreatic cancer with PI3K pathway inhibitors in the era of personalised medicine*, Gut 68 (2019), pp. 742–758.

[82]  R. Fritsch, I. de Krijger, K. Fritsch, R. George, B. Reason, M.S. Kumar et al., *RAS and RHO Families of GTPases Directly Regulate Distinct Phosphoinositide 3-Kinase Isoforms*, Cell 153 (2013), pp. 1050–1063.

[83]  B. Thibault, F. Ramos-Delgado, E. Pons-Tostivint, N. Therville, C. Cintas, S. Arcucci et al., *Pancreatic cancer intrinsic PI3Kα activity accelerates metastasis and rewires macrophage component*, EMBO Mol. Med. 13 (2021), pp. e13502.

[84]  D.A. Dias, S. Urban and U. Roessner, *A Historical Overview of Natural Products in Drug Discovery*, Metabolites 2 (2012), pp. 303–336.

[85]  T.A.K. Prescott, R. Hill, E. Mas-Claret, E. Gaya and E. Burns, *Fungal Drug Discovery for Chronic Disease: History, New Discoveries and New Approaches*, Biomolecules 13 (2023), pp. 986.

[86]  D.J. Newman and G.M. Cragg, *Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019*, J. Nat. Prod. 83 (2020), pp. 770–803.

[87]  J.S. Miller, *The Discovery of Medicines from Plants: A Current Biological Perspective1*, Econ. Bot. 65 (2011), pp. 396–407.

[88]  G.M. Cragg and D.J. Newman, *1.08 - Natural Product Sources of Drugs: Plants, Microbes, Marine Organisms, and Animals*, in *Comprehensive Medicinal Chemistry II*, J.B. Taylor and D.J. Triggle, eds., Elsevier, Oxford, 2007, pp. 355–403.

[89]  J. Larsson, J. Gottfries, S. Muresan and A. Backlund, *ChemGPS-NP: Tuned for Navigation in Biologically Relevant Chemical Space*, J. Nat. Prod. 70 (2007), pp. 789–794.

[90]  C. Tringali, *Bioactive Compounds from Natural Sources, Second Edition: Natural Products as Lead Compounds in Drug Discovery*, CRC Press, 2011.

[91]  A. Rayan, J. Raiyn and M. Falah, *Nature is the best source of anticancer drugs: Indexing natural products for their anticancer bioactivity*, PLoS ONE 12 (2017), pp. e0187925.

[92] S.C. Gupta, S. Patchva, W. Koh and B.B. Aggarwal, *Discovery of Curcumin, a Component of the Golden Spice, and Its Miraculous Biological Activities*, Clin. Exp. Pharmacol. Physiol. 39 (2012), pp. 283–299.

[93] M.T. Moetlediwa, R. Ramashia, C. Pheiffer, S.J.J. Titinchi, S.E. Mazibuko-Mbeje and B.U. Jack, *Therapeutic Effects of Curcumin Derivatives against Obesity and Associated Metabolic Complications: A Review of In Vitro and In Vivo Studies*, Int. J. Mol. Sci. 24 (2023), pp. 14366.

[94] H.A. Zhang and D.D. Kitts, *Turmeric and its bioactive constituents trigger cell signaling mechanisms that protect against diabetes and cardiovascular diseases*, Mol. Cell. Biochem. 476 (2021), pp. 3785–3814.

[95] S. Sudhesh Dev, S.A. Zainal Abidin, R. Farghadani, I. Othman and R. Naidu, *Receptor Tyrosine Kinases and Their Signaling Pathways as Therapeutic Targets of Curcumin in Cancer*, Front. Pharmacol. 12 (2021), pp. 772510.

[96] K. Mansouri, S. Rasoulpoor, A. Daneshkhah, S. Abolfathi, N. Salari, M. Mohammadi et al., *Clinical effects of curcumin in enhancing cancer therapy: A systematic review*, BMC Cancer 20 (2020), pp. 791.

[97] A. Bahrami, M. Majeed and A. Sahebkar, *Curcumin: a potent agent to reverse epithelial-to-mesenchymal transition*, Cell. Oncol. 42 (2019), pp. 405–421.

[98] W. Xu, Z. Yang and N. Lu, *A new role for the PI3K/Akt signaling pathway in the epithelial-mesenchymal transition*, Cell Adhes. Migr. 9 (2015), pp. 317–324.

[99] J. Cao, Z. Han, L. Tian, K. Chen, Y. Fan, B. Ye et al., *Curcumin inhibits EMMPRIN and MMP-9 expression through AMPK-MAPK and PKC signaling in PMA induced macrophages*, J. Transl. Med. 12 (2014), pp. 266.

[100] M. Kedhari Sundaram, S. Silas and A. Hussain, *Combinational Therapy Using Chemotherapeutic Agents and Dietary Bioactive Compounds: A Pragmatic Approach to Cancer Treatment*, 2021, pp. 188–214.

[101] M.C. Nocito, A. De Luca, F. Prestia, P. Avena, D. La Padula, L. Zavaglia et al., *Antitumoral Activities of Curcumin and Recent Advances to ImProve Its Oral Bioavailability*, Biomedicines 9 (2021), pp. 1476.

[102] Y. González, R. Mojica-Flores, D. Moreno-Labrador, L. Cubilla-Rios, K.S.J. Rao, P.L. Fernández et al., *Polyphenols with Anti-Inflammatory Properties: Synthesis and Biological Activity of Novel Curcumin Derivatives*, Int. J. Mol. Sci. 24 (2023), pp. 3691.

[103] A.P. Gupta, S. Khan, M.M. Manzoor, A.K. Yadav, G. Sharma, R. Anand et al., *Chapter 10 - Anticancer Curcumin: Natural Analogues and Structure-Activity Relationship*, in *Studies in Natural Products Chemistry*, Atta-ur-Rahman, ed., Elsevier, 2017, pp. 355–401.

[104] S. Prasad, D. DuBourdieu, A. Srivastava, P. Kumar and R. Lall, *Metal–Curcumin Complexes in Therapeutics: An Approach to Enhance Pharmacological Effects of Curcumin*, Int. J. Mol. Sci. 22 (2021), pp. 7094.

[105] K. Yerdelen, H. Gul, H. Sakagami, N. Umemura and M. Sukuroglu, *Synthesis and Cytotoxic Activities of a Curcumin Analogue and Its bis- Mannich Derivatives*, Lett. Drug Des. Discov. 12 (2015), pp. 643–649.

[106] F. Paulraj, F. Abas, N. H. Lajis, I. Othman and R. Naidu, *Molecular Pathways Modulated by Curcumin Analogue, Diarylpentanoids in Cancer*, Biomolecules 9 (2019), pp. 270.

[107] R. Tabanelli, S. Brogi and V. Calderone, *Improving Curcumin Bioavailability: Current Strategies and Future Perspectives*, Pharmaceutics 13 (2021), pp. 1715.

[108] J. Yakubu and A.V. Pandey, *Innovative Delivery Systems for Curcumin: Exploring Nanosized and Conventional Formulations*, Pharmaceutics 16 (2024), pp. 637.

# Chapter II

# Computational Methods in Drug Discovery

## II.1   Introduction

The development of new therapeutic agents is a challenging and complex endeavor. This process has traditionally been characterized by its time-consuming and resource-intensive nature, often requiring decades of research and substantial financial investment without guaranteed success.

Historically, drug discovery relied heavily on empirical approaches and serendipitous findings. However, the late 20th century witnessed a revolutionary shift with the advent of computer-aided drug design (CADD). This innovative methodology integrates computational algorithms with extensive chemical and biological databases to streamline and rationalize the drug discovery process. The fundamental principle of CADD involves the application of in silico techniques to model and predict molecular interactions between potential drug candidates and their biological targets, allowing researchers to prioritize promising compounds before experimental testing [1].

The implementation of CADD methodologies was facilitated by advancements in structural biology, which provided high-resolution three-dimensional structures of biomolecules, and by the exponential increase in computational power along with the development of sophisticated algorithms, which enabled the execution of complex molecular simulations within practical timeframes.

CADD encompasses a range of computational techniques, broadly categorized into structure-based and ligand-based approaches. Structure-based drug design (SBDD) leverages the 3D structures of biological targets, typically obtained through techniques like X-ray crystallography or NMR spectroscopy, to guide the design of molecules that can effectively interact with these targets. SBDD methods include molecular docking, fragment-based drug discovery, and molecular dynamics simulations. Ligand-based drug design (LBDD), on the other hand, relies on the knowledge of known active compounds to predict and design new potential drug candidates. This approach is particularly useful when the 3D structure of the target protein is unknown. LBDD techniques include quantitative structure-activity relationship (QSAR) studies, pharmacophore modeling, and similarity searching.

Both LBDD and SBDD have demonstrated their value in the drug discovery process, often being used in complementary ways to maximize the chances of identifying promising

drug candidates. This chapter will delve into the various methods and principles underlying CADD, illustrating how each approach contributes to the drug discovery process.
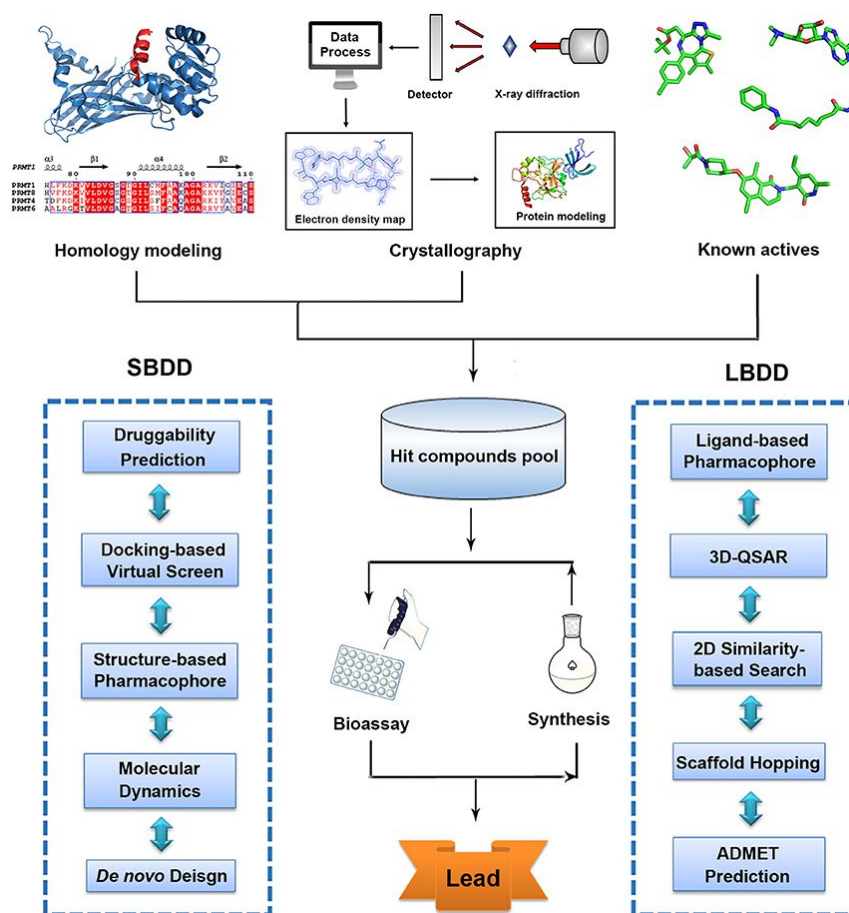


Figure II.1: Integrated approaches in modern drug discovery. Adapted from [2].

## II.2 Virtual screening

Virtual screening has emerged as a pivotal strategy in the drug discovery process, facilitating the identification of potential therapeutic agents by computationally evaluating large compound libraries. This approach can be categorized into ligand-based and structure-based methods.

### II.2.1 Ligand-based virtual screening

### II.2.1.1 Ligand-based pharmacophore modeling

Pharmacophore theory, pioneered by Paul Ehrlich in the late 19th century, initially proposed that specific structural elements or functional groups within a molecule were key to its biological effects. This early concept suggested that compounds with similar biological

activities likely shared common structural or functional characteristics. The term "pharmacophore" itself was later coined by Schueler in his 1960 publication *Chemobiodynamics and Drug Design*, where he defined it as [3]:

"A molecular framework that carries (phoros) the features that are essential for the biological activity of a drug (pharmacon)."

Since then, the understanding of pharmacophores has evolved, shifting from a focus on individual "chemical groups" to recognizing "patterns of abstract features." In 1998, the IUPAC officially defined a pharmacophore as [4]:

"The ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response."

Pharmacophore models have become invaluable tools in virtual screening, enabling the efficient analysis of extensive compound libraries that would otherwise be too large to handle with more resource-intensive structure-based techniques, such as docking. Furthermore, these models can be used to generate feature-based alignments to develop 3D quantitative structure-activity relationships, offering deeper insights into molecular interactions and guiding the design of new drugs.

### II.2.1.1.1 Model construction

Ligand-based pharmacophore modeling is a technique that relies on identifying common features among a set of molecules known to be active against a particular biological target. This method requires a set of compounds with a broad range of activity, ideally spanning a 1000-fold difference, although a minimum of a 50-fold difference can also be sufficient. The approach operates under the premise that all active molecules interact similarly with the target's binding site and share a common mechanism of action. While this assumption may not always hold true, it provides a straightforward starting point for analysis [5].

To construct a robust and reliable ligand-based pharmacophore model, several critical steps must be followed. These include selecting and preparing the dataset, conducting conformational analysis, enumerating possible pharmacophore hypotheses, and ranking and selecting the most representative models. The quality of the pharmacophore model is heavily dependent on the dataset used for its construction, making careful selection and accurate

representation of the dataset essential. The molecules used in this process, known as the "training set," should ideally be tested using the same bioassay procedure, preferably within the same laboratory. During data preparation, it is important to ensure that the chemical structures are accurately represented, paying attention to atom types, bond orders, stereochemistry, tautomers, and charge states. The next step involves exploring the conformational space of each molecule, acknowledging that the bioactive conformation may not always correspond to the lowest energy state. Thus, a thorough conformational search is necessary to identify potential bioactive conformations [6].

Once the conformational search is complete, the next step is to identify pharmacophore features that are common across all active compounds. However, this requirement can be relaxed to allow a common pharmacophore to match only a subset of the active compounds. These identified features are then used to construct various pharmacophore models, each representing potential binding interactions with the target. The generated models are then ranked based on their fitness and the extent to which the active set molecules map onto them [7].

Pharmacophore modeling tools employ various ranking methods to evaluate the quality and relevance of generated models. For instance, Schrödinger's Phase software uses a "survival score" to rank pharmacophore hypotheses. This score reflects the quality and overall desirability of the model based on multiple factors, including the alignment of active molecules with the pharmacophore, volume overlap, and selectivity [8]. The higher the survival score, the more favorable the pharmacophore model is considered, indicating its strong potential in identifying active compounds and effectively guiding drug discovery efforts.

### II.2.1.1.2   Pharmacophore validation

Following the construction of a pharmacophore model, rigorous validation is crucial to ensure its accuracy and reliability for virtual screening applications. This process typically involves assessing the model with a dataset of known active compounds alongside a larger set of presumed inactive decoys. The performance of the pharmacophore model is then assessed by how well it can distinguish between these active compounds and decoys. Common metrics used in this evaluation include the receiver operating characteristic (ROC) curve, the area under the curve (AUC), the enrichment factor (EF), and the robust initial enhancement (RIE).

The ROC curve visually illustrates the true positive rate (TPR) against the false positive rate (FPR) to indicate the frequency of inactive compounds (decoys) being misclassified as active [9]. Figure II.2 displays a theoretical ROC curve.
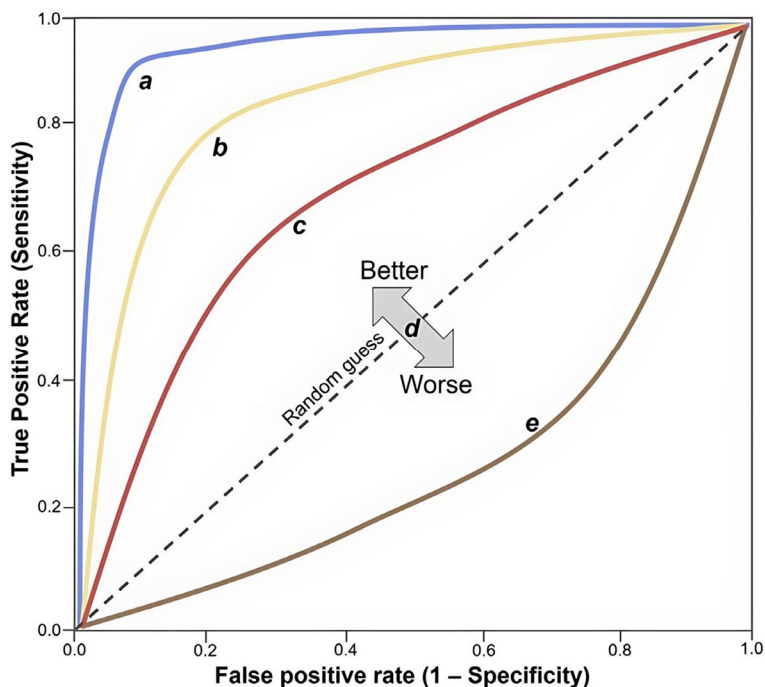


Figure II.2: ROC curves for comparative model performance evaluation. Adapted from [10].

This visualization distinguishes between high (a), good (b), and acceptable (c) performance levels for identifying actives and decoys, thereby facilitating a straightforward interpretation of enrichment. Ideally, the ROC curve should shift towards the upper-left corner, indicating that all actives are identified before any decoys, which corresponds to a TPR of 1 and an FPR of 0. Conversely, random (d) and poor (e) performance curves typically gravitate towards a TPR below 0.5 and an FPR above 0.5, especially as the number of actives and decoys increases [9].

The AUC quantifies the overall performance of a predictive model by measuring the area under the entire ROC curve, assessing how well a predicted active compound ranks against a randomly selected decoy [9,11]. The AUC is computed using Eq. (2.1):

$$AUC = \frac{1}{n}\sum_{i=1}^{n}(1 - f_1) \tag{2.1}$$

Where $f_1$ is the fraction of decoys ranked higher than the $i$th active.

The AUC classification guidelines are as follows:

$AUC < 0.5$ indicates failure, $0.5 \leq AUC < 0.7$ is poor, $0.7 \leq AUC < 0.8$ is acceptable, $0.8 \leq AUC < 0.9$ is good, and $0.9 \leq AUC \leq 1$ is excellent [12].

While AUC is an effective and rapid measure for global virtual screening performance, it is not suitable for evaluating early recognition of actives in the top x% of the ranked list. Many researchers in the field place greater emphasis on early performance, specifically the ability to identify active compounds among the top-ranked results [11]. To address this specific need, the enrichment factor (EF) was developed as a more targeted metric. The EF quantifies the effectiveness of a screening method by comparing its performance to that of random selection. It achieves this by calculating the proportion of active compounds found within a specified top percentile of the ranked dataset, and then contrasting this with the overall ratio of active compounds to decoys in the total dataset under evaluation [11,12]. This metric is typically reported for specific percentages of the filtered dataset. For instance, EF1% represents the EF value derived from the top 1% of ranked compounds. The EF is defined by Eq. (2.2):

$$EF_{x\%} = \frac{n_{x\%}/N_{x\%}}{n/N} \qquad (2.2)$$

Where $n$ represents the total number of actives, $N$ represents the total number of compounds (actives and decoys) in the database, $n_{x\%}$ represents the number of actives, and $N_{x\%}$ represents the total number of compounds in the x% ordered list, respectively.

Despite its usefulness, the EF has notable limitations. Primarily, it exhibits a strong dependence on the ratio of active compounds in the evaluation dataset. Furthermore, it lacks the capability to elucidate the distribution of active compounds within the top-ranked portion of the list [12]. While both the AUC and EF provide valuable insights, they fall short in discerning the precise ordering of active compounds and decoys within the highest-ranked subset. To address these shortcomings, particularly the "early recognition" challenge, researchers developed the robust initial enhancement (RIE) metric [13].

The RIE shares a conceptual similarity with EF, as it quantifies the frequency at which active compounds appear in the top-ranked subset compared to a random ranking, however, unlike EF, the RIE metric has the advantage of including the contributions of actives into the final score [13,14]. Another advantage of RIE is its ability to estimate the distribution of actives at the top of the ordered list. The RIE is defined by Eq. (2.3):

$$RIE = \frac{\frac{1}{n}\sum_{i=n}^{n} e^{-\alpha x_i}}{\frac{1}{n}\left(\frac{1 - e^{-\alpha}}{e^{\frac{\alpha}{n}} - 1}\right)} \qquad (2.3)$$

Where $x_i$ is the relative rank of the $i$th active and $\alpha$ is a tuning parameter.

Understanding and applying these metrics are essential for optimizing pharmacophore models, ultimately aiding in the discovery of new therapeutic agents. This detailed validation ensures that only the most promising candidates are prioritized for further development, streamlining the drug discovery process.

### II.2.1.2   3D-QSAR

Unlike traditional QSAR methods, which rely on molecular descriptors that often fail to capture the spatial characteristics of molecules, 3D-QSAR incorporates the three-dimensional arrangement of atoms. This approach allows for a more precise prediction of how variations in molecular geometry influence biological properties.

Currently, several 3D-QSAR methods are in use, including Comparative Molecular Field Analysis (CoMFA), Comparative Molecular Similarity Indices Analysis (CoMSIA), GRID/GOLPE, and Phase.

- **Comparative Molecular Field Analysis (CoMFA):** CoMFA [15] was one of the pioneering methods in 3D-QSAR modeling. The central concept behind CoMFA is that differences in biological activity among molecules can be explained by variations in the shape and strength of their surrounding non-covalent interaction fields. In essence, steric and electronic fields are sufficient to predict the biological properties of a set of compounds [16].

In a CoMFA study, molecules with known and similar biological activities are virtually aligned and positioned within a cubic grid, typically with a resolution of 2 Å. Interaction energies between the molecules and "probe" atoms are calculated at each grid point using Lennard-Jones and Coulomb potentials. These probes, modeled as $sp^3$ carbon atoms with a charge of +1.0 [17], facilitate the assessment of interaction energy values, which are subsequently compiled into a table. In this table, each row corresponds to a different molecule within the study set, and each column represents the interaction energies at various grid points. Partial Least Squares (PLS) regression analysis is used to correlate

these interaction energies with the biological activity of the molecules, resulting in a regression model characterized by numerous coefficients. The results are often visualized as contour maps, which illustrate regions around the molecules where steric or electrostatic interactions either enhance or hinder biological activity.

- **Comparative Molecular Similarity Indices Analysis (CoMSIA):** CoMFA studies face challenges due to the steep gradients in the calculated electrostatic and steric fields, often resulting in contour maps that are fragmented and difficult to interpret. To overcome this limitation, Klebe and colleagues [18] developed CoMSIA, a method that uses similarity indices rather than interaction fields. This approach, based on a modified SEAL (Steric and Electrostatic Alignment) algorithm [19], utilizes Gaussian functions to describe molecular properties, allowing for smoother and more interpretable contour maps [16].

  In CoMSIA, molecules are placed in a cubic grid, and probes are positioned at each grid point. The method calculates similarity indices focusing on steric, electrostatic, hydrophobic, and hydrogen bond properties, which are then correlated with biological activity. Like CoMFA, CoMSIA results are represented as contour maps, but with improved interpretability due to the smoother distribution of molecular properties.

- **GRID/GOLPE:** The GRID program [20], used in conjunction with GOLPE (Generating Optimal Linear PLS Estimations) [21], offers another approach to 3D-QSAR. GRID calculates interaction fields using more interpretable functions than those in CoMFA, and it allows for a variety of probe types (e.g., methyl, carbonyl, hydroxyl, amines). GOLPE handles the statistical analysis, offering tools to select significant variables for predicting biological activity. As in other 3D-QSAR approaches, results are visualized as contour maps around the ligands, providing insights into regions influencing activity.

- **Phase:** Phase software [8] extends its functionality beyond pharmacophore generation to include 3D-QSAR modeling. In this approach, a pharmacophore model is first generated for the training molecules, which are then aligned according to this pharmacophore and placed in a cubic grid with a resolution of 1 Å. The grid is populated with van der Waals spheres whose radii vary depending on the atomic type.

  Each molecule is represented by a series of "bit" values indicating which grid cubes are occupied by specific atomic classes (e.g., hydrogen bond donors, hydrophobic groups). PLS analysis is used to generate the QSAR model, which is then visualized as blue and

red cubes. Blue cubes indicate positive coefficients that enhance biological activity, while red cubes represent negative coefficients associated with decreased activity.

### II.2.1.2.1   Model validation

The validation of QSAR models is imperative, irrespective of their construction method. Whether developed through conventional or sophisticated techniques, all QSAR models must undergo rigorous evaluation to assess their robustness and predictive capabilities.

➤ **Internal validation**

- *Goodness of fit ($R^2$)*

Goodness of fit, popularly known as the $R^2$ value, is one of the most commonly employed parameters for the internal assessment of the QSAR model. The $R^2$ tells us about the reproducibility of the QSAR model but provides little or no information regarding its robustness and predictability. In QSAR modeling, an R² value above 0.7 is generally deemed acceptable. The R² value is calculated using the formula:

$$R^2 = 1 - \frac{RSS}{TSS} \tag{2.4}$$

Where $RSS$ represents the residual sum of squares and $TSS$ is the total sum of squares. Both of the parameters are calculated using the following equations:

$$RSS = \sum (Y - Y_{pred})^2 \tag{2.5}$$

$$TSS = \sum (Y - Y_{mean})^2 \tag{2.6}$$

Where $Y$ represents the experimentally obtained biological activity of the molecule, $Y_{pred}$ is the biological activity predicted by the QSAR model, and $Y_{mean}$ is the mean of the experimental biological activity values for all molecules.

While a high $R^2$ value indicates a good fit, it can be misleading, especially if many descriptors are used in the model. To address this, the adjusted $R^2_{adj}$ is introduced to account for the number of descriptors:

$$R^2_{adj} = \frac{R^2 - pn - 1}{n - p + 1} \tag{2.7}$$

Where $p$ represents the number of descriptors and $n$ is the number of molecules employed for the development of the QSAR model. Generally, the difference between $R^2$ and $R^2_{adj}$ should not be greater than 0.3 [22].

- ***Cross validation***

Cross-validation methods are divided into two types: leave-one-out (LOO) and leave-many-out (LMO). These methods, collectively referred to as "internal cross-validation," assess the model's predictive performance. In these techniques, one or more portions of the data are systematically removed, and the model is iteratively refitted to the remaining data points. The biological activities of the excluded compounds are then predicted, and the observed versus predicted values are compared. This process results in the calculation of the cross-validated R², often referred to as $Q^2$ or $q^2$, which is determined using the following equation:

$$q^2 = 1 - \frac{\sum(y_i - \tilde{y}_i)^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{PRESS}{SSD} \tag{2.8}$$

Here, $SSD$ is the sum of squared deviations for each actual activity value $y_i$ from the average activity $\bar{y}$, over the entire dataset. $PRESS$, the predictive residual sum of squares, is the sum of squared differences between the actual activity $y_i$ and the predicted activity $\tilde{y}_i$. An acceptable QSAR model should have a $q^2$ value greater than 0.5 [23].

➢ **External validation**

Golbraikh and Tropsha emphasized that external validation is crucial for QSAR model reliability. They argued that while internal validation is necessary, it is not sufficient to guarantee the predictive power of a model. In their work, they proposed several criteria for assessing the performance of QSAR models during external validation. Some of the key criteria they suggested include [23]:

$$R^2_{pred} = 1 - \frac{\sum(Y - Y_{pred})^2}{\sum(Y - Y_{mean})^2} > 0.6 \tag{2.9}$$

Where $Y$ is the experimentally observed biological activity of the molecule used in the test dataset, $Y_{pred}$ is its value obtained by the QSAR model, and $Y_{mean}$ is the mean value of the biological activity of the training set molecules.

$$\frac{(R^2 - R_0^2)}{R^2} < 0.1 \qquad \text{or} \qquad \frac{(R^2 - R_0'^2)}{R^2} < 0.1 \qquad (2.10)$$

$$|R_0^2 - R_0'^2| < 0.3 \qquad (2.11)$$

$$0.85 \leq K \leq 1.15 \qquad \text{or} \qquad 0.85 \leq K' \leq 1.15 \qquad (2.12)$$

Where $R^2$ is the correlation coefficient between the predicted and observed activities, $R_0^2$ is the coefficient of determination of predicted versus observed activities, $R_0'^2$ is the coefficient of determination of observed versus predicted activities. Additionally, $K$ and $K'$ denote the slopes of the regression lines through the origin for predicted versus observed activities and observed versus predicted activities, respectively.

➤ **Y-randomization test**

The Y-randomization test, also known as Y-scrambling, is another important validation approach that is widely used to establish model robustness. In this process, the dependent variable, typically representing biological activity, is randomly permuted across the training set. Subsequently, new QSAR models are constructed using the shuffled data. This process is generally repeated multiple times, with a minimum of five iterations being recommended, though increasing the number of repetitions can provide more reliable results. A robust original model is expected to exhibit significantly higher predictive performance compared to the models generated from randomized data. If the Y-randomized models demonstrate substantially lower predictive power, it indicates that the correlation observed in the original model is not a consequence of overfitting or random chance but is instead reflective of a genuine underlying structure-activity relationship [24,25].

The $cR^2p$ metric is an additional parameter used in this validation process, where:

$$cR^2p = R\sqrt{R^2 - R_r^2} \qquad (2.13)$$

Here, $R$ is the correlation coefficient of the original model, and $R_r$ is the average $R$ values obtained from Y-randomization runs.

For a model to be considered robust, $cR^2p$ should typically be greater than 0.5. This threshold indicates that the original model has a substantial predictive power beyond what would be expected from random noise [26].

### II.2.1.3   Similarity searching

Similarity searching is a foundational ligand-based approach in virtual screening that operates on the principle that structurally similar molecules often exhibit comparable biological activities. In practice, the method uses a known active compound as a reference to screen large chemical databases, aiming to identify structurally similar molecules with potentially similar biological properties. Molecules are represented using various descriptors, which can be 2D (e.g., molecular fingerprints) or 3D (e.g., molecular conformations). Similarity metrics, such as the Tanimoto coefficient and shape-based measures, quantify the degree of resemblance between the reference compound and database molecules. While similarity searching is computationally efficient and does not require protein structure information, its effectiveness depends on the choice of reference compound and molecular representations [27,28].

### II.2.1.4   ADMET

In-silico ADMET predictions play a crucial role in assisting medicinal chemists by enabling the early exclusion of undesirable new chemical entities (NCEs) during the initial stages of drug design. These predictions offer essential feedback on ADMET profiles throughout the lead optimization process and help reduce the reliance on animal testing in laboratory settings. A wide range of in-silico ADMET tools is available, including ADMETlab, SwissADME, pkCSM, Pred-hERG, and QikProp. These tools predict a comprehensive array of properties based on the chemical structure of NCEs, such as drug-likeness, acute toxicity, human intestinal absorption, metabolism, hERG toxicity, and overall pharmacological profile [29].

### II.2.2   Structure-based virtual screening
### II.2.2.1   Molecular docking

Molecular docking is a computational technique widely used in drug discovery to predict the interaction between a small molecule (ligand) and a target protein. The core idea behind molecular docking is to simulate the binding process, which can provide insights into the affinity and orientation of the ligand within the active site of the protein. This technique helps identify potential drug candidates by predicting how well a molecule fits and binds to a particular target, thus accelerating the drug development process [30].

Docking involves two main steps: predicting the conformation, positioning, and orientation of the ligand within the protein binding site (referred to as the pose), and assessing the quality of this pose using a scoring function. Ideally, the sampling algorithm should replicate the experimentally observed binding mode, with the scoring function correctly ranking this pose as the most favorable among the generated poses.

### II.2.2.1.1   Search algorithms

The sampling process in molecular docking is inherently complex. The conformational space encompasses numerous degrees of freedom, including the rotation and translation of one molecule relative to another, as well as the additional conformational flexibility of both the ligand and the protein. In some cases, there are also degrees of freedom associated with the solvent [31].

Given the current limitations in computational resources, it is not feasible to explore every possible conformation and orientation of a molecule within the protein binding site within the realistic time constraints required for virtual screening. Thus, achieving efficient and thorough sampling of the conformational space remains a significant challenge. Early docking methods addressed this issue by treating both the ligand and receptor as rigid bodies, thereby reducing the degrees of freedom to six. These approaches relied primarily on shape complementarity between the ligand and the protein binding site. For instance, programs like DOCK [32] utilize pharmacophore spheres of varying radii to model the ligand and protein. The search algorithm aligns these spheres by matching their internal distances, positioning the ligand within the binding site through a least-squares fitting of atoms to the sphere centers [33,34]. If steric clashes are detected, the algorithm iteratively reorients the ligand until a satisfactory fit is achieved. The final configuration is typically evaluated based on the degree of overlap between the ligand and protein pharmacophore spheres.

While rigid docking methods are computationally efficient, they are constrained by a significant limitation: the inability to account for molecular flexibility. Protein-ligand interactions typically involve conformational changes in both the ligand and the receptor, which rigid docking methods fail to capture, leading to less accurate predictions. To overcome this challenge, modern docking algorithms have been developed to incorporate ligand flexibility. These approaches explore the conformational space of the ligand using a mix of systematic and stochastic techniques, while often limiting protein flexibility to specific regions, such as the binding site, or treating the protein as entirely rigid.

- **Systematic methods**

Systematic search methods aim to explore all possible conformations of a ligand by systematically varying its structural parameters. There are three types of systematic search methods: exhaustive search, fragmentation and conformational ensemble.

Exhaustive search methods involve systematically rotating every possible rotatable bond in the ligand at specified intervals. While this approach is comprehensive, the exponential increase in possible combinations with the number of rotatable bonds makes it more practical for small, relatively flexible ligands. To enhance practicality, geometric or chemical constraints are frequently applied during the initial ligand pose screening. Glide is a prominent example of a docking program that utilizes exhaustive sampling [35,36].

The fragmentation approach adopts an incremental strategy by deconstructing ligands into fragments. One fragment is initially anchored to the protein binding site, with additional fragments added sequentially to build the ligand conformation. FlexX is a notable docking program that employs this fragmentation method [37].

Conformational ensemble methods address ligand flexibility by docking a pre-generated set of ligand conformations. This approach reduces computational demands but relies on additional tools for ensemble generation. A key limitation of this method is that the pre-generated ensemble may not capture the ligand's bioactive conformation [30].

- **Stochastic methods**

Stochastic algorithms in molecular docking sample ligand binding orientations and conformations by introducing random modifications, which are then evaluated and either accepted or rejected based on a predefined criterion. These methods offer the advantage of generating extensive conformational ensembles and exploring a broad segment of the energy landscape, thus enhancing the probability of locating the global energy minimum. However, this comprehensive exploration often results in substantial computational costs. Notable examples of stochastic algorithms include genetic algorithms, Monte Carlo methods, Ant Colony Optimization, and tabu search [30].

Genetic algorithms draw inspiration from the principles of natural selection and evolution. In this approach, ligand conformations are encoded as "chromosomes" that undergo evolutionary processes such as mutation, crossover, and selection. These algorithms are particularly effective in exploring a diverse range of conformations by randomly

modifying ligand structures to optimize their fit within the protein binding site. This technique has been successfully implemented in several docking programs, such as GOLD [38].

Monte Carlo methods, such as those implemented in DockVision, rely on random sampling to explore ligand conformations. They use random rotations and translations of the ligand to generate diverse poses, which are then evaluated against a scoring function. If a generated conformation meets certain energy criteria, it is selected and further refined. This method allows the ligand to overcome energy barriers, making it suitable for discovering novel binding poses that might be missed by other deterministic methods [39].

Inspired by the foraging behavior of ants, Ant Colony Optimization algorithms involve a population of "agents" that mimic ants searching for food. These agents leave behind pheromones that influence the search patterns of subsequent agents. In docking, this translates to identifying optimal binding conformations through iterative optimization, where the best solutions are reinforced over time [40].

The tabu search algorithm enhances the search process by using a memory-based approach to avoid revisiting previously explored conformations, thus reducing redundancy. By keeping a "tabu list" of recently visited states, it ensures a more efficient exploration of the search space. Tabu search is advantageous for refining docking poses, especially in scenarios where local optimization is necessary but traditional methods might get trapped in suboptimal solutions [41].

### II.2.2.1.2   Scoring functions

Scoring functions fall into four main categories: physics-based, empirical, knowledge-based, and machine learning-based [42].

Traditional scoring functions—including physics-based, empirical, and knowledge-based models—operate under the assumption that the binding free energy can be linearly decomposed into distinct components, such as van der Waals forces, electrostatics, and hydrogen bonding. This linear decomposition, while computationally efficient, often oversimplifies the inherently complex and dynamic nature of protein-ligand interactions. In contrast, machine learning-based scoring functions leverage advanced techniques like RF, SVM, and DL to capture the non-linear and multi-faceted relationships between molecular features and binding affinity. These models are capable of incorporating complex

dependencies and subtle interaction patterns that traditional approaches often miss, thereby offering a more sophisticated and adaptable framework for predicting binding affinities [43].

Physics-based, or force-field-based, scoring functions in molecular docking compute binding energy by considering both bonded interactions (like bond stretching, angle bending, and torsional angles) and non-bonded interactions (such as van der Waals forces and electrostatics). These functions emphasize enthalpic contributions to binding energy, and hydrogen bonds are usually treated as either an additional energy term or included within the electrostatic component. One significant challenge for physics-based scoring functions is accurately modeling solvent effects during ligand binding. To address this, implicit solvent models like Poisson-Boltzmann and Generalized Born are commonly used due to their efficiency [44]. For greater accuracy, however, more computationally demanding methods, such as FEP and TI, explicitly account for water molecules, providing a more detailed solvent environment [45]. Furthermore, capturing entropic effects in physics-based scoring functions is challenging due to the lack of a robust physical model. While individual energy terms are crucial, separating them into independent contributions is only feasible when the system can be divided into mutually exclusive subsystems [46]. In practice, many of these terms interact and affect binding affinity in complex ways [47]. The feasibility of decomposing the free energy of ligand binding into a linear combination of interaction terms, without calculating the partition function, remains uncertain. While simplified, these physics-based scoring functions are valued for their speed and adaptability in high-throughput settings, benefiting from continuous improvements in force-field accuracy and computational methods. Widely used examples of these scoring functions include GoldScore [48] and AutoDock [49].

Empirical scoring functions estimate binding affinity by summing various energetic contributions, including hydrogen bonding, hydrophobic effects, and steric clashes between the protein and ligand. The weights for these contributions are determined through multiple linear regression analysis of a dataset containing known binding affinities. While empirical methods are generally faster than physics-based approaches, their accuracy heavily relies on the quality of the training set. Notable examples of empirical scoring functions include ChemScore [51], GlideScore [36], and ChemPLP [40,52].

Knowledge-based scoring functions derive their potentials from statistical analyses of extensive databases containing protein-ligand crystal structures. These methods convert

interatomic contact frequencies into distance-dependent pairwise potentials using the Boltzmann law, operating under the assumption that more favorable interactions are encountered more frequently [30]. This approach effectively balances accuracy and computational speed without the need for ab initio calculations. However, establishing an appropriate reference state for these potentials presents challenges, and corrections for phenomena like excluded volume effects are essential [53]. DrugScore [54,55] and GOLD/ASP [56] are prominent examples of knowledge-based scoring functions that have been successfully applied in docking studies.

Over time, various studies have compared different scoring functions, each with its own strengths and weaknesses. No single scoring function consistently outperforms others in all scenarios, although some excel in specific applications [57,58]. To address individual scoring functions' limitations, consensus scoring—where multiple functions or components are combined—has become a common approach to improve hit rates and reduce errors [59]. Although consensus scoring has shown promise, there are instances where ranking based on a single scoring function proves more effective [60].

Machine learning (ML) techniques have been increasingly integrated into scoring function development, driven by the growing availability of experimental data. These ML-based approaches incorporate QSAR analyses and generate statistical models that compute binding scores based on ligand and protein properties, interaction patterns, and geometrical descriptors [61]. Unlike traditional scoring functions, ML-based models can capture non-linear relationships and automatically extract relevant features from training data. Trained on datasets of active and inactive ligands, these models can accurately rank ligands by potency, making them a promising alternative to classical methods [62,63]. Recent advancements in this area include the development of scoring functions using random forests, support vector machines, artificial neural networks, gradient boosting decision trees, and convolutional neural networks [42]. However, despite their superior performance, ML-based scoring functions are not yet widely integrated into docking software and are primarily used for rescoring [62,64]. Examples of these scoring functions include RF-Score [65] and SVM-Score [66].

## II.2.2.2  Fragment-based drug discovery

Fragment-based drug discovery (FBDD) has become increasingly prominent in the lead identification and optimization stages of drug development [67]. This approach typically

begins with the screening of small, low-molecular-weight compounds, referred to as fragments, to identify those that can bind to a target protein. Unlike more complex drug-like molecules, fragments generally have simpler chemical structures and adhere to the rule of three (RO3), which stipulates that they should have a molecular weight of $\leq 300$ Da, no more than three hydrogen-bond donors, no more than three hydrogen-bond acceptors, and a logP value of $\leq 3$ [68]. These criteria ensure that fragments are large enough to probe interactions within the target protein but small enough to leave ample room for optimization. Once fragments that bind to the target are identified, they are typically optimized by growing, linking, or merging to improve their binding affinity and specificity [69]. This stepwise build-up allows for the systematic enhancement of binding interactions while maintaining a balance between molecular complexity and drug-likeness.

The rationale for adopting FBDD becomes evident when considering the substantial resources invested in combinatorial chemistry and high-throughput screening (HTS) technologies within the pharmaceutical industry. The challenge of discovering novel chemical entities for new protein targets is significant. It is estimated that the chemical universe comprises approximately $10^{60}$ compounds [70], and screening a set of $10^6$ compounds—a substantial corporate repository—only scratches the surface of this vast chemical space. Furthermore, many existing compound libraries are tailored to historical targets, which limits chemical diversity and reduces the likelihood of discovering new leads. To overcome this limitation, pharmaceutical companies have devoted significant resources to expanding the size and diversity of their chemical libraries [71,72].

In contrast, the fragment universe is much smaller, with estimates suggesting around 14 million compounds in the chemical space below 160 Da [73]. Screening a fragment library consisting of 10,000 compounds therefore provides a broader range of chemical diversity compared to conventional high-throughput screens. Additionally, fragment-based screening is advantageous due to the higher hit rates observed with less complex molecules targeting protein structures, as proposed theoretically by Hann and colleagues [74]. Thus, although a typical fragment screen may cover less than 1% of the low-molecular-weight chemical universe, the probability of identifying viable leads is significantly enhanced, thereby improving the efficiency of the screening process. This theoretical framework is supported by recent findings from the Novartis group [75], which demonstrated that hit rates for fragment screens are 10 to 1,000 times higher than those observed in conventional HTS. It is crucial to note that the higher hit rates in fragment-based screens are not merely a

consequence of using higher compound concentrations but are due to the broader chemical diversity facilitated by the fragments' simpler nature [76].

FBDD has notably contributed to the development of six marketed drugs, including Asciminib, Venetoclax, Pexidartinib, Sotorasib, Erdafitinib, and Vemurafenib, and has also generated numerous clinical candidates through fragment screening and optimization approaches [77,78].

## II.3    Molecular dynamics simulations

Molecular dynamics (MD) simulations serve as powerful computational tools that enable detailed representation of molecular system conformations over time. MD holds immense value for the scientific community by providing a fully quantifiable dataset, which is often difficult to achieve in traditional experimental settings. Beyond enabling the analysis of spontaneous processes like protein folding, MD is also extensively applied in predicting and quantifying the physicochemical properties of molecular and macromolecular systems [79,80].

### II.3.1    Foundations of molecular dynamics

The core components required for molecular dynamics simulations are: (i) the interaction potential, which represents the potential energy between particles and from which the forces can be derived, and (ii) the equations of motion that dictate the behavior and dynamics of the particles. These simulations follow the laws of classical mechanics, mainly Newton's law:

$$F_i = m_i a_i \tag{2.14}$$

Here, $m_i$ is the atom mass, $a_i$ its acceleration and $F_i$ the force acting upon it due to the interactions with the other atoms.

Alternatively, the force $F$ on body $i$ can be expressed as the negative gradient of the overall potential energy with respect to the change in the body's position:

$$F_i = m_i a_i = -\nabla U_i = -\frac{dU}{dr_i} \tag{2.15}$$

By using the atomic forces and masses, one can determine the positions of atoms over a sequence of very small-time intervals, typically in the range of femtoseconds. The velocities of the atoms are then obtained from their accelerations:

$$a_i = \frac{dv_i}{dt} \qquad (2.16)$$

Finally, the positions are calculated from the velocities:

$$v_i = \frac{dr_i}{dt} \qquad (2.17)$$

In essence, at each timestep, the forces acting on the atoms are determined and combined with their current positions and velocities to predict their new states in the subsequent interval. It is assumed that the forces remain constant within each time increment. The atoms are then moved to their new positions, the forces are recalculated, and the process is repeated for the next cycle [81].

Solving the equations of motion provides a trajectory that describes how the particles evolve over time. For simple systems involving only two particles, these differential equations can be solved analytically. However, in the context of MD simulations, where systems often involve thousands or even millions of particles, analytical solutions become infeasible. As a result, numerical integration methods are employed to approximate the particles' velocities and positions at each time step. Several numerical integration schemes are commonly used in MD simulations, including the Verlet, Leapfrog, and Velocity-Verlet algorithms [81].

In MD simulations, the timestep is the increment of time over which the equations of motion are integrated. The choice of timestep is crucial because it directly affects the accuracy and efficiency of the simulation. If the timestep is too small, the simulation will require an enormous number of calculations to simulate even a brief period, making it computationally expensive and inefficient. Conversely, if the timestep is too large, the simulation may run faster, but it risks missing critical details or introducing significant errors. This can result in the simulation failing to accurately capture the system's dynamics, potentially leading to incorrect results. In biomolecular systems, the fastest vibrations typically involve hydrogen atoms, with a frequency corresponding to a period of about 10 femtoseconds (fs). To balance accuracy with computational efficiency, a timestep of 1 or 2

fs is commonly used in MD simulations, ensuring that the system's dynamics are captured accurately without unnecessary computational expense [82].

## II.3.2    Periodic boundary conditions

Finite and infinite systems have distinct characteristics, and there is no definitive answer for determining the minimum size a relatively small system must achieve to accurately mimic the behavior of an infinite system. In simulations, the system is enclosed within a container, which might initially be perceived as having rigid walls that atoms collide with when trying to exit the simulation area. In macroscopic systems, only a tiny fraction of atoms is close enough to the walls to experience deviations from the conditions within the system's interior. For example, in a three-dimensional system with $N_m = 10^{21}$ atoms at liquid density, the number of atoms near the walls is on the order of $N_m^{2/3}$, which equals approximately $10^{14}$ atoms—just one in $10^7$. However, with a more typical molecular dynamics (MD) value of $N_m = 1000$, around 500 atoms are near the walls, leaving very few atoms in the interior. Excluding the first two layers, only 216 atoms remain. As a result, the simulation may not accurately capture the typical behavior of interior atoms, leading to measurements that reflect this limitation. Therefore, unless the objective is to study behavior near actual walls, it is preferable to eliminate these boundaries [83,84].
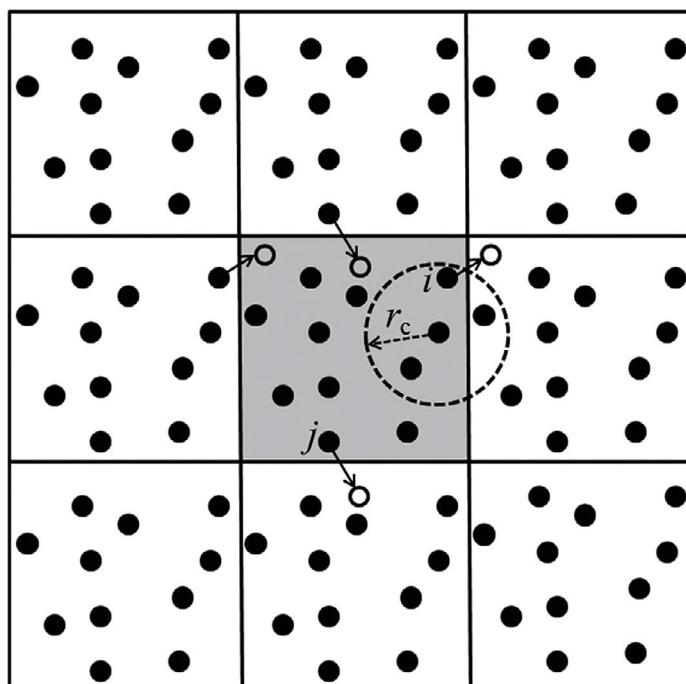


Figure II.3: The schematic of PBC in a two-dimensional system. Adapted from [85].

This can be accomplished by applying periodic boundary conditions (PBCs), as illustrated in Figure II.3. With these boundary conditions, the system behaves as if it is replicated infinitely, forming a continuous lattice of identical simulation cells. This periodic arrangement leads to two primary effects: first, when an atom exits the simulation box through one boundary, it immediately re-enters through the opposite boundary; second, atoms within a cutoff distance $r_c$ near the boundary can interact with atoms in adjacent cells or with those near the opposite boundary, creating a seamless "wraparound" effect. This approach effectively removes the limitations imposed by finite simulation boundaries, allowing the system to be treated as if it were spatially infinite, while still maintaining homogeneity at the boundaries [83,84].

## II.3.3  Statistical ensemble

Statistical mechanics serves as a fundamental framework that bridges the microscopic behavior of individual particles with the macroscopic physical properties we observe in experiments. This framework is deeply rooted in Gibbs ensemble theory, which posits that a wide array of microscopic states can result in the same macroscopic characteristics for a large system in equilibrium. This concept is crucial because it implies that, rather than tracking the detailed motion of every particle in a system, we can instead use statistical methods to predict its overall behavior. The key insight provided by ensemble theory is that macroscopic properties can be determined by averaging over a large number of hypothetical copies of the system, each existing in a different microstate consistent with certain macroscopic constraints. These copies form what is known as a statistical ensemble, where each microstate reflects a possible configuration of the system under the given conditions. By computing ensemble averages, we obtain properties that correspond to the time-averaged measurements observed in experiments. This ensemble-based approach provides a powerful method to derive thermodynamic properties without the need for exhaustive time-dependent simulations [81].

Statistical ensembles are classified based on the specific thermodynamic parameters that remain fixed, such as energy (E), temperature (T), pressure (P), volume (V), particle number (N), or chemical potential (μ). Depending on which of these variables are controlled, different ensembles—such as the canonical (NVT) and isothermal-isobaric (NPT) ensembles—are used to model systems under varying physical conditions [81].

Thermostats and barostats are algorithms used during MD simulations to regulate temperature and pressure, respectively. Popular thermostats include the Simple Velocity Rescaling thermostat, Nosé-Hoover thermostat, and Langevin thermostat. Frequently used barostats are the Parrinello-Rahman barostat, Martyna-Tuckerman-Tobias-Klein barostat, and Berendsen barostat [79].

### II.3.4   Main steps of a molecular dynamics simulation

The general workflow of a molecular dynamics (MD) simulation typically involves the following key steps:

- **System preparation:** This phase ensures that the molecular system is in a suitable starting configuration for MD simulations. It includes selecting the molecular structure (e.g., a protein-ligand complex), validating its geometry (correcting steric clashes or missing atoms), and assigning appropriate force field parameters [88].

- **Solvation:** After the input structures are prepared, the system is solvated to replicate a realistic aqueous environment. The solvation process involves placing the biomolecular system in a box of explicit water molecules using standard water models like TIP3P. Periodic boundary conditions are applied to simulate an infinite system by allowing atoms to move across boundaries and re-enter from the opposite side, which helps to eliminate edge effects. Additionally, counterions are added to neutralize the system's overall charge and mimic physiological salt concentrations, ensuring that the simulated environment closely resembles biological conditions [86].

- **Energy minimization:** Once solvated, the system undergoes energy minimization to eliminate unfavorable atomic contacts or steric clashes that may have been introduced during system preparation. This step involves optimizing atomic positions to find a local minimum on the potential energy surface, thereby stabilizing the system. Common algorithms used for this process include steepest descent and conjugate gradient methods [87].

- **Equilibration:** The equilibration phase aims to bring the system to a stable thermodynamic state by adjusting variables such as temperature and pressure to match the desired ensemble conditions. This process typically consists of two stages: first, the system is equilibrated using the NVT ensemble (constant number of particles, volume, and temperature) to stabilize the temperature, followed by the NPT ensemble (constant

number of particles, pressure, and temperature) to ensure proper density and pressure equilibration [87].

- **Production:** In the final step, the actual MD simulation is performed under the desired ensemble conditions (often NPT) for a sufficiently long time to study the system's dynamics. During this phase, no restraints are applied, and the system evolves naturally according to Newton's equations of motion. The simulation generates trajectories, which are a series of snapshots of the system's atomic coordinates over time. These trajectories are then analyzed to extract properties like root-mean-square deviation, radius of gyration, and hydrogen bonding patterns, providing insights into the system's structural and dynamic behavior [79].

## II.4 Binding energy

Calculating the binding free energy between a ligand and its target is crucial for understanding the stability and affinity of the interaction. The two widely used methods for free energy estimation are MM-PBSA (Molecular Mechanics Poisson-Boltzmann Surface Area) and MM-GBSA (Molecular Mechanics Generalized Born Surface Area). The binding energy using MM-GBSA or MM-PBSA method can be calculated as a difference in energy of the bound and unbound states as shown in Eq. 2.18 [88]:

$$\Delta G_{bind} = \Delta G_{complex} - \left(\Delta G_{receptor} + \Delta G_{ligand}\right) \tag{2.18}$$

Where $\Delta G_{complex}$, $\Delta G_{receptor}$, and $\Delta G_{ligand}$ represent the average free energies of the complex, receptor, and ligand, respectively. $\Delta G$ can be expressed as:

$$\Delta G = \Delta G_{gas} + \Delta G_{sol} \tag{2.19}$$

Where $\Delta G_{gas}$ is the gas-phase energy and $\Delta G_{sol}$ is the solvation energy. $\Delta G_{gas}$ is calculated as the sum of internal (bond, angle, dihedral), electrostatic, and vdW energies:

$$\Delta G_{gas} = \Delta E_{bond} + \Delta E_{angle} + \Delta E_{dihedral} + \Delta E_{ele} + \Delta E_{vdW} \tag{2.20}$$

The solvation free energy ($\Delta G_{sol}$) is divided into two components:

$$\Delta G_{sol} = \Delta E_{pol} + \Delta E_{nonpol} \tag{2.21}$$

Where $\Delta E_{pol}$ and $\Delta E_{nonpol}$ are the polar and non-polar contributions, respectively. The polar component of binding free energy can be calculated using the Poissone-Boltzmann or the

Generalized Born approximation models. The nonpolar contribution of the solvation free energy is calculated as a function of the solvent-accessible surface area (SASA), as follows:

$$\Delta E_{nonpol} = \gamma SASA + \beta \tag{2.22}$$

Where $\gamma$ is a proportionality constant related to the surface tension, and $\beta$ is a constant offset.

# References

[1] S.K. Niazi and Z. Mariam, *Computer-Aided Drug Design and Drug Discovery: A Prospective Analysis*, Pharmaceuticals 17 (2024), pp. 22.

[2] W. Lu, R. Zhang, H. Jiang, H. Zhang and C. Luo, *Computer-Aided Drug Design in Epigenetics*, Front. Chem. 6 (2018), .

[3] X. Qing, X.Y. Lee, J.D. Raeymaecker, J.R. Tame, K.Y. Zhang, M.D. Maeyer et al., *Pharmacophore modeling: advances, limitations, and current utility in drug discovery*, J. Recept. Ligand Channel Res. 7 (2014), pp. 81–92.

[4] C.G. Wermuth, C.R. Ganellin, P. Lindberg and L.A. Mitscher, *Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998)*, Pure Appl. Chem. 70 (1998), pp. 1129–1143.

[5] R. Lewis and R.G. Karki, *Structure-based Pharmacophores and Screening*, in *Computational Structural Biology*, WORLD SCIENTIFIC, 2008, pp. 501–512.

[6] S.L. Dixon, *Pharmacophore methods*, in *Drug Design: Structure- and Ligand-Based Approaches*, C.H. Reynolds, D. Ringe and J. Merz Kenneth M., eds., Cambridge University Press, Cambridge, 2010, pp. 137–150.

[7] D. Giordano, C. Biancaniello, M.A. Argenio and A. Facchiano, *Drug Design by Pharmacophore and Virtual Screening Approach*, Pharmaceuticals 15 (2022), pp. 646.

[8] S.L. Dixon, A.M. Smondyrev, E.H. Knoll, S.N. Rao, D.E. Shaw and R.A. Friesner, *PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results*, J. Comput. Aided Mol. Des. 20 (2006), pp. 647–671.

[9] T. Fawcett, *An introduction to ROC analysis*, Pattern Recognit. Lett. 27 (2006), pp. 861–874.

[10] B.J. Neves, M. Mottin, J.T. Moreira-Filho, B.K. de P. Sousa, S.S. Mendonca and C.H. Andrade, *Chapter 4 - Best Practices for Docking-Based Virtual Screening*, in *Molecular Docking for Computer-Aided Drug Design*, M.S. Coumar, ed., Academic Press, 2021, pp. 75–98.

[11] M.D. Mackey and J.L. Melville, *Better than Random? The Chemotype Enrichment Problem*, J. Chem. Inf. Model. 49 (2009), pp. 1154–1162.

[12] R.C. Braga and C.H. Andrade, *Assessing the performance of 3D pharmacophore models in virtual screening: how good are they?*, Curr. Top. Med. Chem. 13 (2013), pp. 1127–1138.

[13] R.P. Sheridan, S.B. Singh, E.M. Fluder and S.K. Kearsley, *Protocols for Bridging the Peptide to Nonpeptide Gap in Topological Similarity Searches*, J. Chem. Inf. Comput. Sci. 41 (2001), pp. 1395–1406.

[14] J.-F. Truchon and C.I. Bayly, *Evaluating Virtual Screening Methods: Good and Bad Metrics for the "Early Recognition" Problem*, J. Chem. Inf. Model. 47 (2007), pp. 488–508.

[15] R.D. Cramer, D.E. Patterson and J.D. Bunce, *Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins*, J. Am. Chem. Soc. 110 (1988), pp. 5959–5967.

[16] W. Sippl, *3D-QSAR – Applications, Recent Advances, and Limitations*, 2010, pp. 103–125.

[17] R.D. Cramer, D.E. Patterson and J.D. Bunce, *Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins*, J. Am. Chem. Soc. 110 (1988), pp. 5959–5967.

[18] G. Klebe, U. Abraham and T. Mietzner, *Molecular Similarity Indices in a Comparative Analysis (CoMSIA) of Drug Molecules to Correlate and Predict Their Biological Activity*, J. Med. Chem. 37 (1994), pp. 4130–4146.

[19] S.K. Kearsley and G.M. Smith, *An alternative method for the alignment of molecular structures: Maximizing electrostatic and steric overlap*, Tetrahedron Comput. Methodol. 3 (1990), pp. 615–633.

[20] C.A. Reynolds, R.C. Wade and P.J. Goodford, *Identifying targets for bioreductive agents: Using GRID to predict selective binding regions of proteins*, J. Mol. Graph. 7 (1989), pp. 103–108.

[21] J. Nilsson, H. Wikström, A. Smilde, S. Glase, T. Pugsley, G. Cruciani et al., *GRID/GOLPE 3D quantitative structure-activity relationship study on a set of benzamides and naphthamides, with affinity for the dopamine D3 receptor subtype*, J. Med. Chem. 40 (1997), pp. 833–840.

[22] A. Thakur, V. Mehta, P. Nagu and K. Goutam, *Computer-Aided Drug Design: QSAR, Molecular Docking, Virtual Screening, Homology and Pharmacophore Modeling*, 2024.

[23] A. Golbraikh and A. Tropsha, *Beware of q2!*, J. Mol. Graph. Model. 20 (2002), pp. 269–276.

[24] A. Tropsha, *Predictive Quantitative Structure–Activity Relationship Modeling*, in *Comprehensive Medicinal Chemistry II*, Elsevier, 2007, pp. 149–165.

[25] A. Golbraikh, X.S. Wang, H. Zhu and A. Tropsha, *Predictive QSAR Modeling: Methods and Applications in Drug Discovery and Chemical Risk Assessment*, in *Handbook of Computational Chemistry*, J. Leszczynski, ed., Springer Netherlands, Dordrecht, 2016, pp. 1–38.

[26] S. Khamouli, S. Belaidi, M. Bakhouch, S. Chtita, M.A. Hashmi and F.A. Qais, *QSAR modeling, molecular docking, ADMET prediction and molecular dynamics simulations of some 6-arylquinazolin-4-amine derivatives as DYRK1A inhibitors*, J. Mol. Struct. 1258 (2022), pp. 132659.

[27] P. Willett, *Similarity-based virtual screening using 2D fingerprints*, Drug Discov. Today 11 (2006), pp. 1046–1053.

[28] A. Kumar and K.Y.J. Zhang, *Advances in the Development of Shape Similarity Methods and Their Application in Drug Discovery*, Front. Chem. 6 (2018), .

[29] P.K. Parikh, J.K. Savjani, A.K. Gajjar and M.T. Chhabria, *Bioinformatics and Cheminformatics Tools in Early Drug Discovery*, in *Bioinformatics Tools for Pharmaceutical Drug Product Development*, John Wiley & Sons, Ltd, 2023, pp. 147–181.

[30] F. Stanzione, I. Giangreco and J.C. Cole, *Chapter Four - Use of molecular docking computational tools in drug discovery*, in *Progress in Medicinal Chemistry*, D.R. Witty and B. Cox, eds., Elsevier, 2021, pp. 273–343.

[31] L.G. Ferreira, R.N. Dos Santos, G. Oliva and A.D. Andricopulo, *Molecular Docking and Structure-Based Drug Design Strategies*, Molecules 20 (2015), pp. 13384–13421.

[32] I.D. Kuntz, J.M. Blaney, S.J. Oatley, R. Langridge and T.E. Ferrin, *A geometric approach to macromolecule-ligand interactions*, J. Mol. Biol. 161 (1982), pp. 269–288.

[33] W. Kabsch, *A solution for the best rotation to relate two sets of vectors*, Acta Crystallogr. Sect. A 32 (1976), pp. 922–923.

[34] W. Kabsch, *A discussion of the solution for the best rotation to relate two sets of vectors*, Acta Crystallogr. Sect. A 34 (1978), pp. 827–828.

[35] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz et al., *Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy*, J. Med. Chem. 47 (2004), pp. 1739–1749.

[36] T.A. Halgren, R.B. Murphy, R.A. Friesner, H.S. Beard, L.L. Frye, W.T. Pollard et al., *Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening*, J. Med. Chem. 47 (2004), pp. 1750–1759.

[37] M. Rarey, B. Kramer, T. Lengauer and G. Klebe, *A Fast Flexible Docking Method using an Incremental Construction Algorithm*, J. Mol. Biol. 261 (1996), pp. 470–489.

[38] G. Jones, P. Willett, R.C. Glen, A.R. Leach and R. Taylor, *Development and validation of a genetic algorithm for flexible docking1*, J. Mol. Biol. 267 (1997), pp. 727–748.

[39] T.N. Hart and R.J. Read, *A multiple-start Monte Carlo docking method*, Proteins Struct. Funct. Bioinforma. 13 (1992), pp. 206–222.

[40] O. Korb, T. Stützle and T.E. Exner, PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design, in Ant Colony Optimization and Swarm Intelligence, 2006, pp. 247–258.

[41] J. Pei, Q. Wang, Z. Liu, Q. Li, K. Yang and L. Lai, *PSI-DOCK: Towards highly efficient and accurate flexible ligand docking*, Proteins Struct. Funct. Bioinforma. 62 (2006), pp. 934–946.

[42] J. Li, A. Fu and L. Zhang, *An Overview of Scoring Functions Used for Protein–Ligand Interactions in Molecular Docking*, Interdiscip. Sci. Comput. Life Sci. 11 (2019), pp. 320–328.

[43] Q.U. Ain, A. Aleksandrova, F.D. Roessler and P.J. Ballester, *Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening*, WIREs Comput. Mol. Sci. 5 (2015), pp. 405–424.

[44] X. Zou, Yaxiong and I.D. Kuntz, *Inclusion of Solvation in Ligand Binding Free Energy Calculations Using the Generalized-Born Model*, J. Am. Chem. Soc. 121 (1999), pp. 8033–8043.

[45] Z. Cournia, B. Allen and W. Sherman, *Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations*, J. Chem. Inf. Model. 57 (2017), pp. 2911–2937.

[46] A.E. Mark and W.F. van Gunsteren, *Decomposition of the Free Energy of a System in Terms of Specific Interactions: Implications for Theoretical and Experimental Studies*, J. Mol. Biol. 240 (1994), pp. 167–176.

[47] D.H. Williams, A.J. Maguire, W. Tsuzuki and M.S. Westwell, *An Analysis of the Origins of a Cooperative Binding Energy of Dimerization*, Science 280 (1998), pp. 711–714.

[48] M.L. Verdonk, J.C. Cole, M.J. Hartshorn, C.W. Murray and R.D. Taylor, *Improved protein–ligand docking using GOLD*, Proteins Struct. Funct. Bioinforma. 52 (2003), pp. 609–623.

[49] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew et al., *Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function*, J. Comput. Chem. 19 (1998), pp. 1639–1662.

[50] C.R. Corbeil, C.I. Williams and P. Labute, *Variability in docking success rates due to dataset preparation*, J. Comput. Aided Mol. Des. 26 (2012), pp. 775–786.

[51] M.D. Eldridge, C.W. Murray, T.R. Auton, G.V. Paolini and R.P. Mee, *Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes*, J. Comput. Aided Mol. Des. 11 (1997), pp. 425–445.

[52] O. Korb, T. Stützle and T.E. Exner, *Empirical Scoring Functions for Advanced Protein−Ligand Docking with PLANTS*, J. Chem. Inf. Model. 49 (2009), pp. 84–96.

[53] S.-Y. Huang and X. Zou, *An iterative knowledge-based scoring function to predict protein–ligand interactions: II. Validation of the scoring function*, J. Comput. Chem. 27 (2006), pp. 1876–1882.

[54] H.F.G. Velec, H. Gohlke and G. Klebe, *DrugScoreCSDKnowledge-Based Scoring Function Derived from Small Molecule Crystal Data with Superior Recognition Rate of Near-Native Ligand Poses and Better Affinity Prediction*, J. Med. Chem. 48 (2005), pp. 6296–6303.

[55] H. Gohlke, M. Hendlich and G. Klebe, *Knowledge-based scoring function to predict protein-ligand interactions1*, J. Mol. Biol. 295 (2000), pp. 337–356.

[56] W.T.M. Mooij and M.L. Verdonk, *General and targeted statistical potentials for protein–ligand interactions*, Proteins Struct. Funct. Bioinforma. 61 (2005), pp. 272–287.

[57] Y. Li, L. Han, Z. Liu and R. Wang, *Comparative Assessment of Scoring Functions on an Updated Benchmark: 2. Evaluation Methods and General Results*, J. Chem. Inf. Model. 54 (2014), pp. 1717–1736.

[58] T. Cheng, X. Li, Y. Li, Z. Liu and R. Wang, *Comparative Assessment of Scoring Functions on a Diverse Test Set*, J. Chem. Inf. Model. 49 (2009), pp. 1079–1093.

[59] P.S. Charifson, J.J. Corkery, M.A. Murcko and W.P. Walters, *Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins*, J. Med. Chem. 42 (1999), pp. 5100–5109.

[60] M.L. Verdonk, V. Berdini, M.J. Hartshorn, W.T.M. Mooij, C.W. Murray, R.D. Taylor et al., *Virtual Screening Using Protein−Ligand Docking: Avoiding Artificial Enrichment*, J. Chem. Inf. Comput. Sci. 44 (2004), pp. 793–806.

[61] S.-Y. Huang, S.Z. Grinter and X. Zou, *Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions*, Phys. Chem. Chem. Phys. 12 (2010), pp. 12899–12908.

[62] M. Wójcikowski, P.J. Ballester and P. Siedlecki, *Performance of machine-learning scoring functions in structure-based virtual screening*, Sci. Rep. 7 (2017), pp. 46710.

[63] C. Shen, Y. Hu, Z. Wang, X. Zhang, H. Zhong, G. Wang et al., *Can machine learning consistently improve the scoring power of classical scoring functions? Insights into the role of machine learning in scoring functions*, Brief. Bioinform. 22 (2021), pp. 497–514.

[64] M.A. Khamis, W. Gomaa and W.F. Ahmed, *Machine learning in computational docking*, Artif. Intell. Med. 63 (2015), pp. 135–152.

[65] P.J. Ballester and J.B.O. Mitchell, *A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking*, Bioinformatics 26 (2010), pp. 1169–1175.

[66] L. Li, B. Wang and S.O. Meroueh, *Support Vector Regression Scoring of Receptor–Ligand Complexes for Rank-Ordering and Virtual Screening of Chemical Libraries*, J. Chem. Inf. Model. 51 (2011), pp. 2132–2138.

[67] C. Fernández and W. Jahnke, *New approaches for NMR screening in drug discovery*, Drug Discov. Today Technol. 1 (2004), pp. 277–283.

[68] P. Kirsch, A.M. Hartman, A.K.H. Hirsch and M. Empting, *Concepts and Core Principles of Fragment-Based Drug Design*, Molecules 24 (2019), pp. 4309.

[69] E.V. Bedwell, W.J. McCarthy, A.G. Coyne and C. Abell, *Development of potent inhibitors by fragment-linking strategies*, Chem. Biol. Drug Des. 100 (2022), pp. 469–486.

[70] Y.C. Martin, *Challenges and prospects for computational aids to molecular diversity*, Perspect. Drug Discov. Des. 7 (1996), pp. 159–172.

[71] E. Jacoby, A. Schuffenhauer, M. Popov, K. Azzaoui, B. Havill, U. Schopfer et al., *Key aspects of the Novartis compound collection enhancement project for the compilation of a comprehensive chemogenomics drug discovery screening collection*, Curr. Top. Med. Chem. 5 (2005), pp. 397–411.

[72] A.M. Davis, D.J. Keeling, J. Steele, N.P. Tomkinson and A.C. Tinker, *Components of successful lead generation*, Curr. Top. Med. Chem. 5 (2005), pp. 421–439.

[73] T. Fink, H. Bruggesser and J.-L. Reymond, *Virtual Exploration of the Small-Molecule Chemical Universe below 160 Daltons*, Angew. Chem. Int. Ed. 44 (2005), pp. 1504–1508.

[74] M.M. Hann, A.R. Leach and G. Harper, *Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery*, J. Chem. Inf. Comput. Sci. 41 (2001), pp. 856–864.

[75] A. Schuffenhauer, S. Ruedisser, A.L. Marzinzik, W. Jahnke, M. Blommers, P. Selzer et al., *Library design for fragment based screening*, Curr. Top. Med. Chem. 5 (2005), pp. 751–762.

[76] P.J. Hajduk and J. Greer, *A decade of fragment-based drug design: strategic advances and lessons learned*, Nat. Rev. Drug Discov. 6 (2007), pp. 211–219.

[77] M. Bon, A. Bilsland, J. Bower and K. McAulay, *Fragment-based drug discovery—the importance of high-quality molecule libraries*, Mol. Oncol. 16 (2022), pp. 3761.

[78] Q. Li, *Application of Fragment-Based Drug Discovery to Versatile Targets*, Front. Mol. Biosci. 7 (2020), .

[79] E. Braun, J. Gilmer, H.B. Mayes, D.L. Mobley, J.I. Monroe, S. Prasad et al., *Best Practices for Foundations in Molecular Simulations [Article v1.0]*, Living J. Comput. Mol. Sci. 1 (2019), pp. 5957.

[80] S. Bottaro and K. Lindorff-Larsen, *Biophysical experiments and biomolecular simulations: A perfect match?*, Science 361 (2018), pp. 355–360.

[81] R. Schneider, A.R. Sharma and A. Rai, *Introduction to Molecular Dynamics*, in *Computational Many-Particle Physics*, H. Fehske, R. Schneider and A. Weiße, eds., Springer, Berlin, Heidelberg, 2008, pp. 3–40.

[82] D. Fincham, *Choice of timestep in molecular dynamics simulation*, Comput. Phys. Commun. 40 (1986), pp. 263–269.

[83] D. Jefferies and S. Khalid, *Atomistic and coarse-grained simulations of membrane proteins: A practical guide*, Methods 185 (2021), pp. 15–27.

[84] D.C. Rapaport, *The Art of Molecular Dynamics Simulation*, 2nd ed.Cambridge University Press, Cambridge, 2004.

[85] Y. Du, R. Schmid-Fetzer, J. Wang, S. Liu, J. Wang and Z. Jin, *Computational Design of Engineering Materials: Fundamentals and Case Studies*, Cambridge University Press, Cambridge, 2023.

[86] R.B. Shirts, S.R. Burt and A.M. Johnson, *Periodic boundary condition induced breakdown of the equipartition principle and other kinetic effects of finite sample size in classical hard-sphere molecular dynamics simulation*, J. Chem. Phys. 125 (2006), pp. 164102.

[87] A.R. Leach, *Molecular Modelling: Principles and Applications*, Prentice Hall, 2001.

[88] A. Ramaswamy, S. Balasubramanian and M. Rajagopalan, *Chapter 2 - Biomolecular Talks—Part 1: A Theoretical Revisit on Molecular Modeling and Docking Approaches*, in *Molecular Docking for Computer-Aided Drug Design*, M.S. Coumar, ed., Academic Press, 2021, pp. 31–55.

# Chapter III

# Application 1 – Comprehensive Computational Strategies to Identify Novel Curcumin Derivatives Against Pancreatic Cancer

## III.1    Introduction

Building on the insights discussed in Chapter I, it becomes evident that there is a pressing need for the development of selective inhibitors targeting the PI3Kα isoform. PI3Kα plays a pivotal role in promoting tumor growth and metastasis in pancreatic cancer, making it an attractive target for therapeutic intervention. However, the current landscape is marked by a significant lack of isoform-selective inhibitors, with the majority of available options being pan-PI3K inhibitors. These pan-inhibitors, while effective in some contexts, indiscriminately target all class I PI3K isoforms, leading to substantial off-target effects and associated toxicities [1].

Developing selective PI3Kα inhibitors is particularly challenging due to the remarkable structural homology shared by the different PI3K isoforms, which complicates the process of achieving high specificity [2]. Despite these challenges, recent advancements have shown promise. A notable example is Alpelisib (NVP-BYL719), a selective PI3Kα inhibitor that has gained FDA approval for the treatment of HR-positive, HER2-negative breast cancer with PIK3CA mutations [3]. This success highlights the potential of selective PI3Kα inhibition as a therapeutic strategy, though its application in pancreatic cancer remains underexplored.
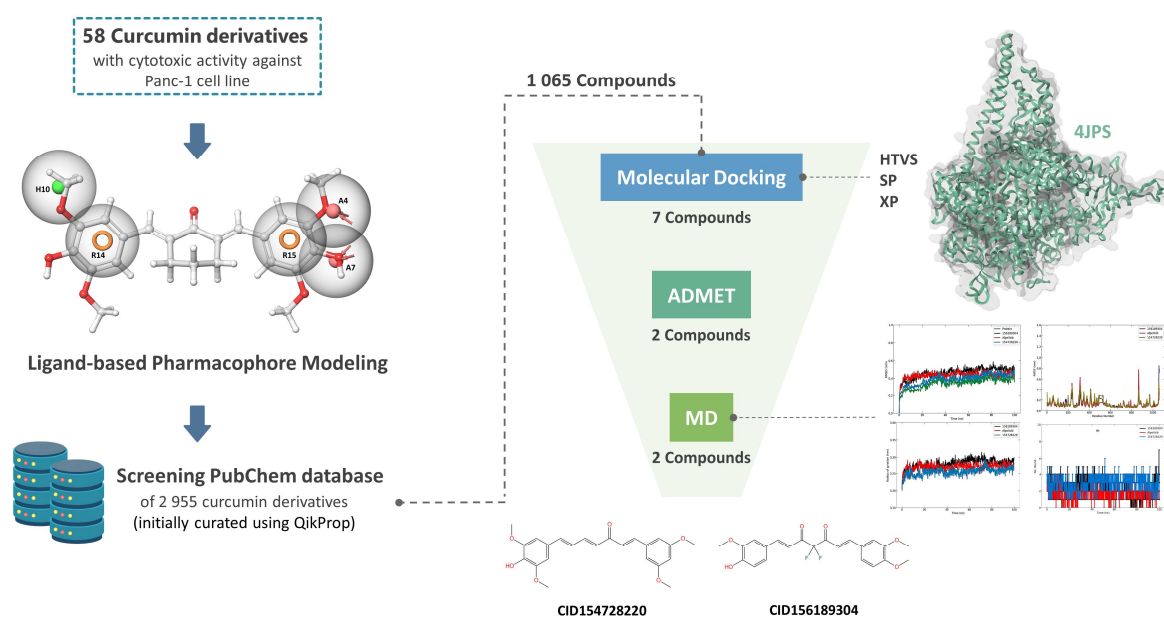


Figure III.1: Schematic representation of the workflow.

In this chapter, we will delve into the computational approaches employed in our research to identify potential curcumin-based therapeutic agents that specifically target PI3Kα for the treatment of pancreatic cancer (Figure III.1). These methods aim to overcome the challenges associated with isoform selectivity and provide new avenues for more effective and safer cancer therapies.

## III.2    Material and methods

All computational studies discussed in this chapter, except for the molecular dynamics simulations, were performed on a laptop equipped with an Intel Core i5 processor (1.60 GHz), 16.0 GB RAM, and a 64-bit Windows 10 operating system.

### III.2.1    Dataset preparation

A dataset of 58 curcumin-related compounds, which have documented anticancer activity against the Panc-1 cell line, was compiled from the literature [4]. The $IC_{50}$ values of these compounds, ranging from 0.16 to 27.86 μM, were converted to $pIC_{50}$ values using the formula: $pIC_{50} = - \log_{10}[IC_{50}]$.

The 2D molecular structures were initially sketched using ChemDraw Ultra (v20.1), followed by 3D conversion and energy optimization in the Maestro interface via the LigPrep module. Partial atomic charges and ionization states at a physiological pH of 7.4 were determined using Epik (v5.4) [5]. Each compound was limited to a single stereoisomer by maintaining specified chirality. Energy minimization was carried out using the OPLS3e force field [6] until achieving an RMSD cutoff of 0.01 Å. The structures obtained were then used for further modeling studies.

### III.2.2    Pharmacophore and 3D-QSAR modeling

Pharmacophore and 3D-QSAR models were developed using the Schrödinger Phase software [7]. The ligands and their $pIC_{50}$ values were input into the Develop Pharmacophore Model panel. Ligands were categorized based on their potencies as active ($pIC_{50} \geq 5.5$), inactive ($pIC_{50} < 5.0$), or moderately active. By employing the default settings, up to 50 conformers were generated for each ligand. The conformers were then aligned using an automated method to identify the best alignment and common features among them.

The pharmacophore sites were developed using the six built-in properties in Phase: hydrophobic group (H), positively charged group (P), hydrogen bond acceptor (A), negatively charged group (N), hydrogen bond donor (D), and aromatic ring (R) [7]. The matching tolerance was set to 2 Å, and minimal distances between features were set to 1 Å. Common features among most active compounds (≥50%) were identified to develop pharmacophore hypotheses using a tree-based partitioning algorithm [7,8]. By restricting the number of sites to four or five, twenty variants were generated. These hypotheses were assessed and ranked according to criteria such as volume, vector, survival scores, site, number of matches, and survival inactivity [7].

For the best scoring pharmacophore hypothesis, a 3D-QSAR model was developed using atom-based PLS regression method. The dataset was divided into training and test sets in a 75:25 ratio, with balanced representation of low, medium, and high activity compounds [9]. Grid spacing was maintained at 1 Å, and variables with |t value| < 2.0 were excluded. The model showed increased predictive power and statistical significance up to five PLS factors. Finally, the model was rigorously validated to ensure its reliability for prospective screening and drug design applications.

### III.2.3 Database preparation and pharmacophore-based screening

To conduct virtual screening, 2,955 curcumin derivatives were obtained from the PubChem repository (https://pubchem.ncbi.nlm.nih.gov/) as 2D SDF files (accessed on November 4, 2022) [10]. Using Phase software (v6.5) [11], these compounds were converted into a 3D database. This involved uploading the structures to the Generate Phase Database panel and processing them through the LigPrep module. Epik (v5.4) was used to adjust the ionization states to a pH of 7.4 and generate a single tautomer for each ligand [5]. Finally, energy minimization was carried out using the OPLS3e force field [6].

After preparation, QikProp (v6.6) was employed to compute pharmacokinetic properties. Initial filtering was based on Lipinski's Rule of Five to select drug-like candidates, followed by further refinement through additional custom filters, including:

$$QPPCaco > 25; \ -1.5 \leq QPlogKhsa \leq 1.5; \ -2 \leq QPlogPo/w \leq 6.5$$
$$-3.0 \leq QPlogBB \leq 1.2; \ -6.5 \leq QPlogS \leq 0.5; \ QPPMDCK > 25$$
$$Percent \ Human \ Oral \ Absorption > 25$$

These criteria helped in identifying compounds with favorable ADME properties and kept the database size manageable. To identify potential inhibitors, the best pharmacophore model was used as a 3D query, requiring structures to match at least four of the five features specified in the hypothesis. Additionally, the selective PI3Kα inhibitor Alpelisib was acquired from PubChem and optimized using the same settings applied during the database preparation.

### III.2.4    Protein preparation

The Protein Data Bank provided the 3D structure of PI3Kα bound to Alpelisib (PDB ID: 4JPS, 2.20 Å) [12], which was prepared using Schrödinger's Protein Preparation Wizard [13]. Bond orders were assigned based on the CCD database, and hydrogen atoms were added. Prime (v6.2) was used to complete missing side chains and loops [14]. Water molecules beyond 5.0 Å from the ligand were deleted. Epik was employed to generate protonation and tautomeric states of the het group at a pH of 7.4. Hydrogen bond assignments were optimized, and PROPKA adjusted water orientations at a pH of 7.4 [15]. Finally, the OPLS3e force field minimized the complex until the RMSD of heavy atoms reached 0.3 Å [6].

### III.2.5    Grid generation and structure-based virtual screening

Following ligand-based pharmacophore screening, the chosen compounds were flexibly docked into the PI3Kα binding pocket (PDB ID: 4JPS) [12]. A receptor grid centered on the co-crystallized ligand was generated, with coordinates X=−1.27, Y=−9.35, and Z=−16.54. Van der Waals radius scaling was set to 1 Å, and a partial charge cutoff of 0.25 was used, with no constraints. The docking protocol was validated by extracting and redocking the co-crystallized ligand into its original binding site.

Docking was conducted using the Virtual Screening Workflow of Glide (v8.9) [16], with three docking levels: HTVS, SP, and XP. At each level, the top 20% of molecules were passed to the next stage. The XP docking results underwent binding free energy calculation and minimization via the MM-GBSA method in the Prime module. The MM-GBSA calculations used the OPLS-2005 force field, the VSGB 2.0 solvation model [17], and rotamer search algorithms [18]. Molecular interactions and binding poses were visualized using the Maestro interface (Schrödinger Release 2020-3) and PyMOL v2.5 software.

### III.2.6    ADMET prediction

The quest for new pharmaceuticals is notoriously protracted and costly. A review of clinical trials from 2010 to 2017 shows that poor drug-like properties account for roughly $10-15\%$ of clinical trial failures, and unmanageable toxicity accounts for another 30% [19]. To overcome these challenges and improve the success rate of drug candidates reaching the market, it is essential to integrate pharmacokinetic assessments early in the drug discovery process. This allows for better resource allocation towards the most promising candidates. We assessed the ADME properties of our compounds using the QikProp module (v6.6) from Schrödinger, and their toxicity profiles were evaluated using the pkCSM online server.

### III.2.7    Molecular dynamics simulations

Using GROMACS (v2020.2), molecular dynamic simulations were conducted over a 100 ns period in an aqueous medium, utilizing the CHARMM36 force field. The systems were prepared in a cubic box filled with TIP3P water molecules, and $Na^+$ and $Cl^-$ ions were introduced to neutralize the overall charge. The protein was placed at the center of the box, ensuring at least 1 nm distance from the box edges to comply with the minimum image convention. Energy minimization was performed via the Steepest Descent algorithm for 5,000 steps, achieving a maximum force below $1,000 \, kJ \, mol^{-1} \, nm^{-1}$ to resolve steric clashes.

The complexes underwent equilibration to stabilize temperature and pressure, using NVT and NPT ensembles. The equilibration process lasted 125 ps at 300.15 K, with positional restraints of $400 \, kJ \, mol^{-1} \, nm^{-2}$ on the backbone and $40 \, kJ \, mol^{-1} \, nm^{-2}$ on the sidechains. A 100 ns production run was then executed in the NPT ensemble at 300.15 K and 1 bar. Temperature regulation was managed using a Nose-Hoover thermostat, while pressure was controlled via the Parrinello-Rahman barostat. The LINCS algorithm was employed to constrain hydrogen bonds, as per the parameters defined by CHARMM-GUI. Additionally, a Velocity rescaling thermostat set at 300 K with a coupling constant of 1 ps was utilized. Trajectory data were recorded every 2 ps throughout the simulation.

## III.3    Results and discussion

### III.3.1    Pharmacophore 3D-QSAR modeling

Out of the 20 hypotheses generated, AAHRR_1 was identified as the best fit, boasting a maximum survival score of 5.596. The hypothesis consists of two hydrogen bond acceptors

(A), one hydrophobic group (H), and two aromatic rings (R) (Figure III.2). The alignment of AAHRR_1 over active and inactive compounds is depicted in Figure III.3. The distances and angles between various sites of AAHRR_1 are provided in the Appendix.
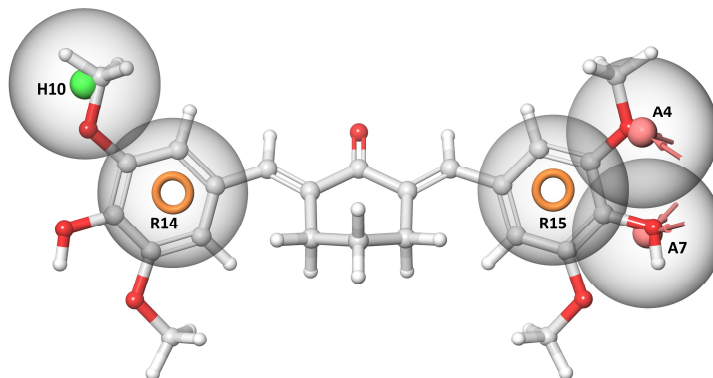


Figure III.2: AAHRR_1 pharmacophore aligned with the reference ligand. Pink sphere with arrows: H-bond acceptor (A); green sphere: hydrophobic group (H); orange circle: aromatic ring (R).
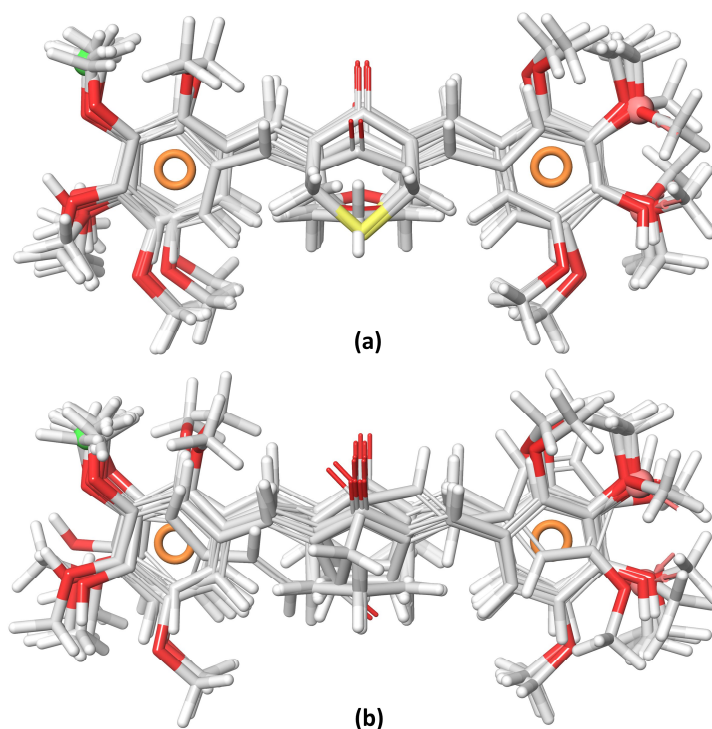


Figure III.3: Mapping of active (a) and inactive (b) compounds onto the AAHRR_1 pharmacophore model.

The results of the atom-based 3D-QSAR model derived from AAHRR_1 are summarized in Table III.1. At PLS factor 5, the model demonstrated a high variance ratio

(F) of 783.8 and a Pearson-r value of 0.988, indicating high statistical significance and a strong correlation between predicted and observed values. Furthermore, the small standard deviation (SD = 0.053) and root mean square error (RMSE = 0.080) suggest that the model is both accurate and reliable in its predictions. The $R^2$ value of 0.990 implies that 99% of the variance in the data is accounted for by the model, and the cross-validated correlation coefficient ($Q^2$) of 0.971 highlights the model's excellent predictive power and reliability when applied to new data. The stability score of 0.577, though moderate, is within acceptable limits and indicates a stable model with reliable predictive performance across different datasets. The chemical structures of the dataset, along with their corresponding experimental and predicted values, are presented in Table III.2.

Table III.1: Statistical parameters of the AAHRR_1 model.

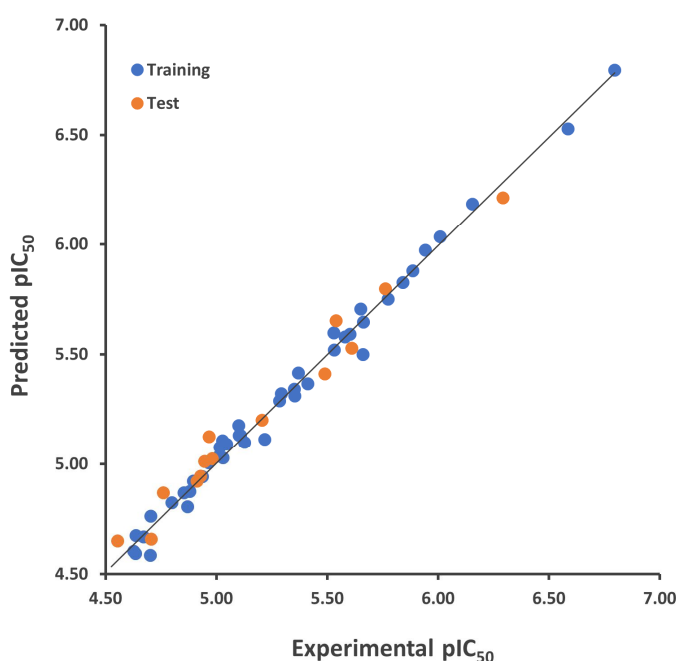| PLS | SD | $R^2$ | Stability | F | P | RMSE | $Q^2$ | Pearson-r |
|-----|-----|-------|-----------|-----|-----|------|-------|-----------|
| 1 | 0.251 | 0.765 | 0.853 | 136.9 | 8.4e-15 | 0.28 | 0.649 | 0.877 |
| 2 | 0.175 | 0.889 | 0.766 | 163.9 | 2.78e-20 | 0.25 | 0.722 | 0.923 |
| 3 | 0.110 | 0.957 | 0.614 | 296.6 | 2.37e-27 | 0.12 | 0.930 | 0.989 |
| 4 | 0.078 | 0.979 | 0.575 | 450.9 | 4.49e-32 | 0.08 | 0.969 | 0.989 |
| 5 | 0.053 | 0.990 | 0.577 | 783.8 | 3.13e-37 | 0.08 | 0.971 | 0.988 |



Figure III.4: Visualization of observed versus Phase-predicted activity using a scatter plot for the training and test set compounds.

Table III.2: Molecular structures of training and test set compounds along with their predicted and actual $pIC_{50}$ values.
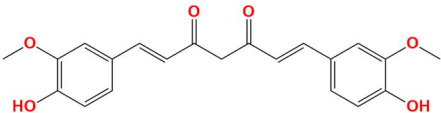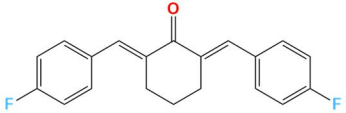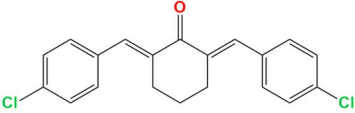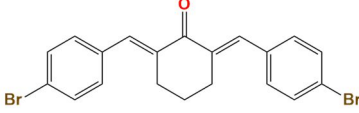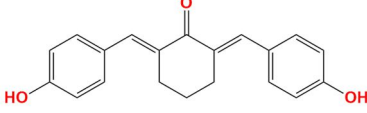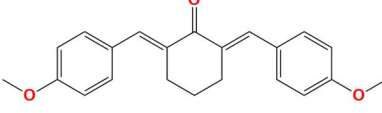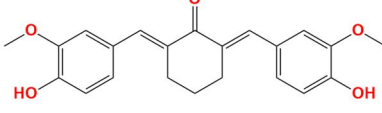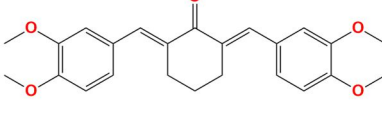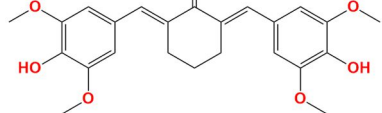
| Ligands | Chemical structure | Experimental activity ($pIC_{50}$) | Predicted activity ($pIC_{50}$) | Residual activity |
|---|---|---|---|---|
| Curcumin |  | 4.626 | 4.602 | −0.023 |
| A1[t] |  | 4.982 | 5.022 | 0.040 |
| A2[t] |  | 4.929 | 4.942 | 0.013 |
| A3[t] |  | 4.912 | 4.921 | 0.009 |
| A4 |  | 4.879 | 4.874 | −0.005 |
| A6 |  | 4.798 | 4.823 | 0.025 |
| A7 |  | 5.354 | 5.311 | −0.043 |
| A8 |  | 5.351 | 5.340 | −0.010 |
| A9 |  | 5.580 | 5.577 | −0.003 |

Table III.2 (Continued)

| Ligands | Chemical structure | Experimental activity (pIC$_{50}$) | Predicted activity (pIC$_{50}$) | Residual activity |
|---------|-------------------|-----------------------------------|--------------------------------|-------------------|
| A10 |  | 5.015 | 5.077 | 0.062 |
| A11 |  | 4.937 | 4.942 | 0.005 |
| B1[t] |  | 4.705 | 4.657 | −0.048 |
| B2 |  | 4.635 | 4.593 | −0.042 |
| B3 |  | 4.702 | 4.583 | −0.119 |
| B4 |  | 4.671 | 4.667 | −0.004 |
| B6 |  | 4.638 | 4.674 | 0.036 |
| B7 |  | 5.013 | 5.038 | 0.025 |
| B8 |  | 4.870 | 4.804 | −0.066 |

Table III.2 (Continued)
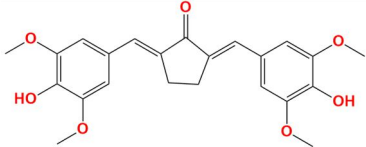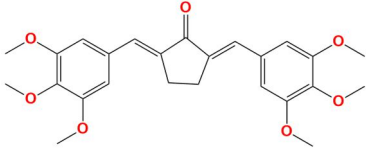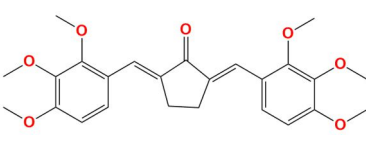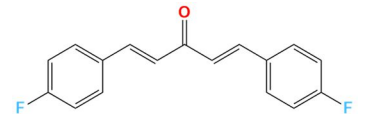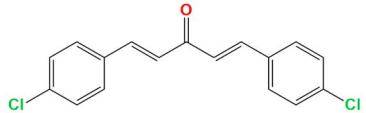
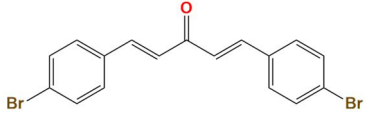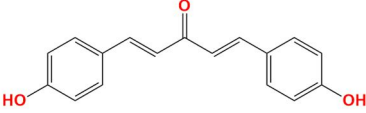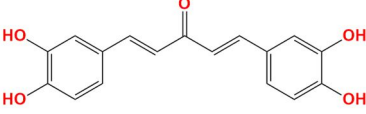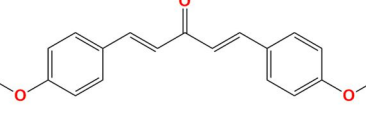| Ligands | Chemical structure | Experimental activity (pIC$_{50}$) | Predicted activity (pIC$_{50}$) | Residual activity |
|---------|-------------------|-----------------------------------|--------------------------------|-------------------|
| B9 |  | 5.108 | 5.131 | 0.023 |
| B10 |  | 4.896 | 4.921 | 0.025 |
| B11 |  | 4.704 | 4.761 | 0.057 |
| C1 |  | 5.029 | 5.026 | −0.003 |
| C2 |  | 4.964 | 5.005 | 0.040 |
| C3 |  | 4.853 | 4.867 | 0.013 |
| C4 |  | 5.028 | 5.107 | 0.079 |
| C5 |  | 5.045 | 5.090 | 0.045 |
| C6$^{t}$ |  | 4.947 | 5.010 | 0.064 |

Table III.2 (Continued)
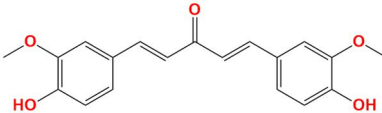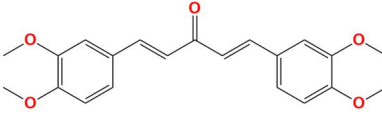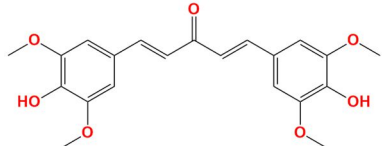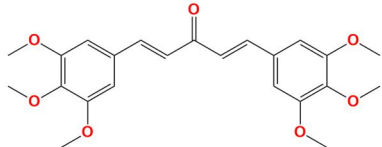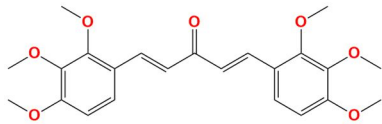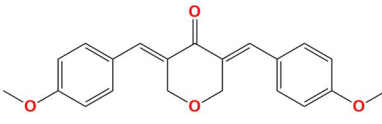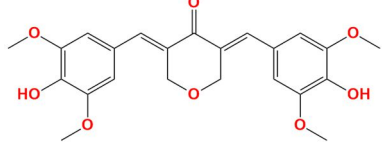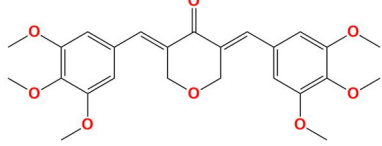
| Ligands | Chemical structure | Experimental activity (pIC$_{50}$) | Predicted activity (pIC$_{50}$) | Residual activity |
|---|---|---|---|---|
| C7 |  | 5.532 | 5.518 | −0.013 |
| C8 |  | 5.664 | 5.645 | −0.018 |
| C9 |  | 5.650 | 5.704 | 0.055 |
| C10 |  | 5.775 | 5.749 | −0.026 |
| C11 |  | 5.293 | 5.320 | 0.027 |
| E4 |  | 5.413 | 5.366 | −0.048 |
| E6 |  | 5.218 | 5.113 | −0.105 |
| E9 |  | 5.842 | 5.825 | −0.017 |
| E10 |  | 6.796 | 6.795 | −0.001 |

Table III.2 (Continued)

| Ligands | Chemical structure | Experimental activity (pIC$_{50}$) | Predicted activity (pIC$_{50}$) | Residual activity |
|---|---|---|---|---|
| E11[t] |  | 5.611 | 5.526 | −0.085 |
| F4 |  | 5.370 | 5.415 | 0.045 |
| F6 |  | 5.101 | 5.176 | 0.075 |
| F9[t] |  | 5.762 | 5.797 | 0.035 |
| F10 |  | 6.585 | 6.527 | −0.058 |
| F11[t] |  | 5.539 | 5.652 | 0.113 |
| AN1 |  | 5.886 | 5.878 | −0.008 |
| AN2 |  | 5.602 | 5.590 | −0.012 |
| AN3 |  | 5.662 | 5.497 | −0.164 |

Table III.2 (Continued)
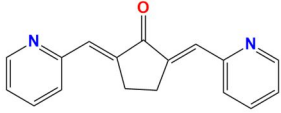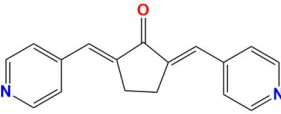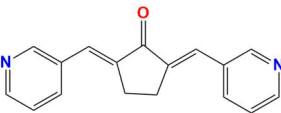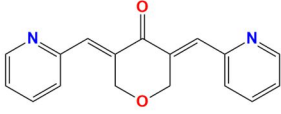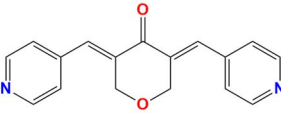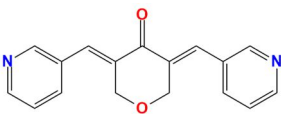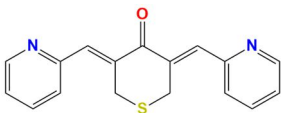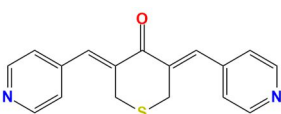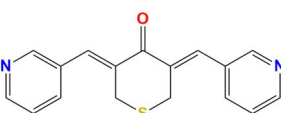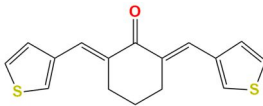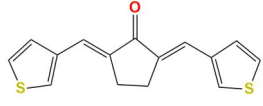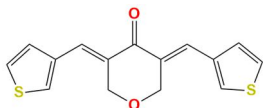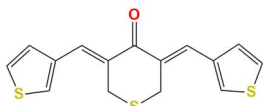
| Ligands | Chemical structure | Experimental activity (pIC$_{50}$) | Predicted activity (pIC$_{50}$) | Residual activity |
|---------|-------------------|-----------------|-----------------|-------------------|
| BN1 | | 5.285 | 5.288 | 0.003 |
| BN2[t] | | 5.206 | 5.201 | −0.005 |
| BN3 | | 5.103 | 5.130 | 0.027 |
| EN1[t] | | 5.489 | 5.410 | −0.079 |
| EN2 | | 6.009 | 6.034 | 0.025 |
| EN3 | | 5.529 | 5.595 | 0.066 |
| FN1 | | 6.155 | 6.184 | 0.029 |
| FN2[t] | | 6.292 | 6.212 | −0.081 |
| FN3 | | 5.943 | 5.972 | 0.029 |

Table III.2 (Continued)

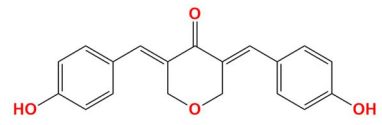| Ligands | Chemical structure | Experimental activity (pIC$_{50}$) | Predicted activity (pIC$_{50}$) | Residual activity |
|---------|-------------------|-----------|-----------|-----------|
| AS[t] |  | 4.760 | 4.868 | 0.108 |
| BS[t] |  | 4.555 | 4.650 | 0.095 |
| ES[t] |  | 4.966 | 5.124 | 0.158 |
| FS |  | 5.126 | 5.098 | −0.027 |

**t**: test set

## III.3.2 Model validation

## III.3.2.1 Enrichment study

To evaluate the effectiveness of the AAHRR_1 model in virtual screening and its ability to distinguish between active and inactive compounds, we performed an enrichment study. The Schrödinger decoy set, comprising 1,000 drug-like compounds, was enriched with sixteen active compounds [22,23], none of which were included in the initial model construction. Following preparation using LigPrep default settings, the dataset was imported into the hypothesis validation tool within the Phase module, a specialized tool used to assess the predictive accuracy of pharmacophore models by testing them against known active and inactive compounds.

The pharmacophore model AAHRR_1 demonstrates robust performance in the enrichment study, as reflected by key metrics. The model achieved an EF1% of 13, indicating that it is 13 times more effective at identifying active compounds within the top

1% of the ranked list compared to a random selection. The Robust Initial Enhancement (RIE) score of 8.59 further corroborates the model's efficacy in early recognition of active compounds. The overall predictive accuracy of the model is exemplified by the Receiver Operating Characteristic (ROC) score of 0.96 and the Area Under the Curve (AUC) of 0.95, both approaching the ideal value of 1.0. These metrics indicate that AAHRR_1 is highly effective in distinguishing active compounds from inactive ones, with particularly strong performance in prioritizing active compounds during the early screening stages.
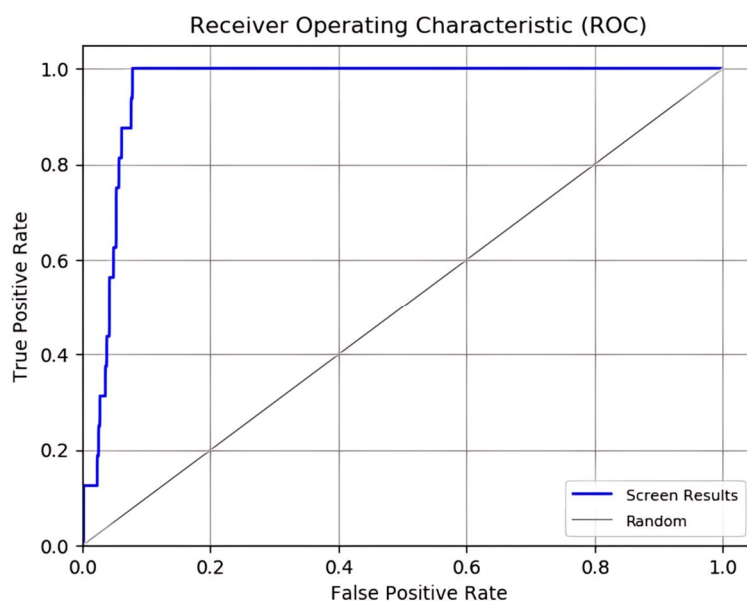


Figure III.5: ROC plot of AAHRR_1 pharmacophore model.

### III.3.2.2 External validation

KNIME, the Konstanz Information Miner, stands as an advanced platform for data analytics and machine learning, offering a comprehensive suite of tools for data integration, processing, analysis, and exploitation. In our study, we made use of the Enalos KNIME nodes, developed by Nova Mechanics Ltd., which provide specialized functionalities in molecular modeling and cheminformatics. Specifically, we employed the Enalos Model Acceptability Criteria KNIME node (v 4.7) to rigorously evaluate the predictive accuracy of our QSAR model [20]. This node integrates all the external validation tests recommended by Tropsha, ensuring a comprehensive performance evaluation [21].

The validation results, illustrated in the right column of Figure III.6, were compared to the statistical benchmarks shown in the left column. The analysis confirmed that the model

met all of Tropsha's criteria. This validation confirms the utility of our QSAR model in drug discovery pipelines, especially for identifying anticancer compounds with similar structural attributes.



| Criterion | Assessment | Result |
|---|---|---|
| R^2 > 0.6 | PASS | R^2 = 0.977 |
| Rcvext^2 > 0.5 | PASS | Rcvext^2 = 0.973 |
| (R^2-R0^2)/R^2 < 0.1 | PASS | (R^2-R0^2)/R^2 = 0.008 |
| (R^2-R'0^2)/R^2 < 0.1 | PASS | (R^2-R'0^2)/R^2 = 0.004 |
| abs(R0^2-R'0^2) < 0.1 | PASS | abs(R0^2-R'0^2) = 0.004 |
| 0.85 < k < 1.15 | PASS | k = 0.996 |
| 0.85 < k' < 1.15 | PASS | k' = 1.004 |

**Model Predictive**

Figure III.6: Results obtained from the Enalos Model Acceptability Criteria KNIME node.

### III.3.2.3 Y-randomization test

To ensure the reliability of our QSAR model, we conducted a Y-randomization test, where the response variable vector was randomly permuted across ten iterations. The findings revealed a marked decrease in $R^2$ and $Q^2$ values in the permuted models compared to the original one, with a $cRp^2$ value greater than 0.5. This indicates that the original model is statistically robust and not influenced by random correlation.

Table III.3: Y-randomization test results.

| Model | R | $R^2$ | $Q^2$ | Model | R | $R^2$ | $Q^2$ |
|---|---|---|---|---|---|---|---|
| Original | 0.995 | 0.990 | 0.971 | Random 6 | 0.780 | 0.608 | 0.537 |
| Random 1 | 0.845 | 0.714 | 0.318 | Random 7 | 0.763 | 0.583 | −0.723 |
| Random 2 | 0.741 | 0.549 | −0.501 | Random 8 | 0.733 | 0.537 | 0.243 |
| Random 3 | 0.840 | 0.705 | −0.587 | Random 9 | 0.803 | 0.644 | −0.381 |
| Random 4 | 0.821 | 0.674 | −0.863 | Random 10 | 0.707 | 0.501 | 0.099 |
| Random 5 | 0.807 | 0.652 | −0.252 | | | | |

Table III.4: Y-randomization parameters for the permuted models.

| Parameters | Model score | Threshold |
|---|---|---|
| Average R | 0.803 | $< R$ |
| Average $R^2$ | 0.651 | $< R^2$ |
| Average $Q^2$ | −0.104 | $< Q^2$ |
| $cR^2p$ | 0.585 | $> 0.50$ |

### III.3.3   3D-QSAR contour map analysis

The atom-based 3D-QSAR results are visualized as clusters of cubes, each representing pharmacophoric regions that influence the biological activity of molecules. Blue cubes denote regions where specific group substitutions enhance activity, whereas red cubes indicate areas where such substitutions diminish activity.

Table III.5 provides a breakdown of atom-type contributions, offering insights into how different atomic properties impact the biological activity of a compound. At PLS factor five, hydrophobic groups (0.469) and electron-withdrawing groups (0.337) exhibit the most substantial contributions, significantly outweighing those from hydrogen-bond donors (0.034) and other factors (0.160). This prominence suggests that hydrophobic and electron-withdrawing properties are the most influential in determining the biological activity of the compounds. Figure III.7 presents a pictorial representation of the contour maps derived from the best QSAR model for the most active (E10) and least active (BS) compounds.

Table III.5: Atom-type fraction contribution.

| Factors | H-bond donor | Hydrophobic/non-polar | Electron-withdrawing | Other |
|---|---|---|---|---|
| 1 | 0.218 | 0.546 | 0.363 | 0.069 |
| 2 | 0.026 | 0.511 | 0.365 | 0.098 |
| 3 | 0.030 | 0.473 | 0.355 | 0.141 |
| 4 | 0.033 | 0.466 | 0.342 | 0.160 |
| 5 | 0.034 | 0.469 | 0.337 | 0.160 |

The hydrogen bond donor contour map reveals a predominance of red cubes in both E10 and BS, suggesting that adding hydrogen bond donors could reduce biological activity. This observation aligns with their relatively low contribution (0.034) to the QSAR model.

For the most potent compound, the hydrophobic map reveals a prominent blue contour surrounding the distal benzene rings, with a notable concentration around the methyl groups of the methoxy substituents. This suggests that the hydrophobic interactions are strong and localized, particularly around these regions, contributing significantly to the compound's potency. In contrast, the least active compound, featuring distal thiophene rings, displays scattered and less concentrated blue cubes across its structure. The presence of red cubes further indicates regions of polarity, suggesting that hydrophobic interactions are weaker and more disrupted compared to E10, contributing to the reduced activity of the compound.
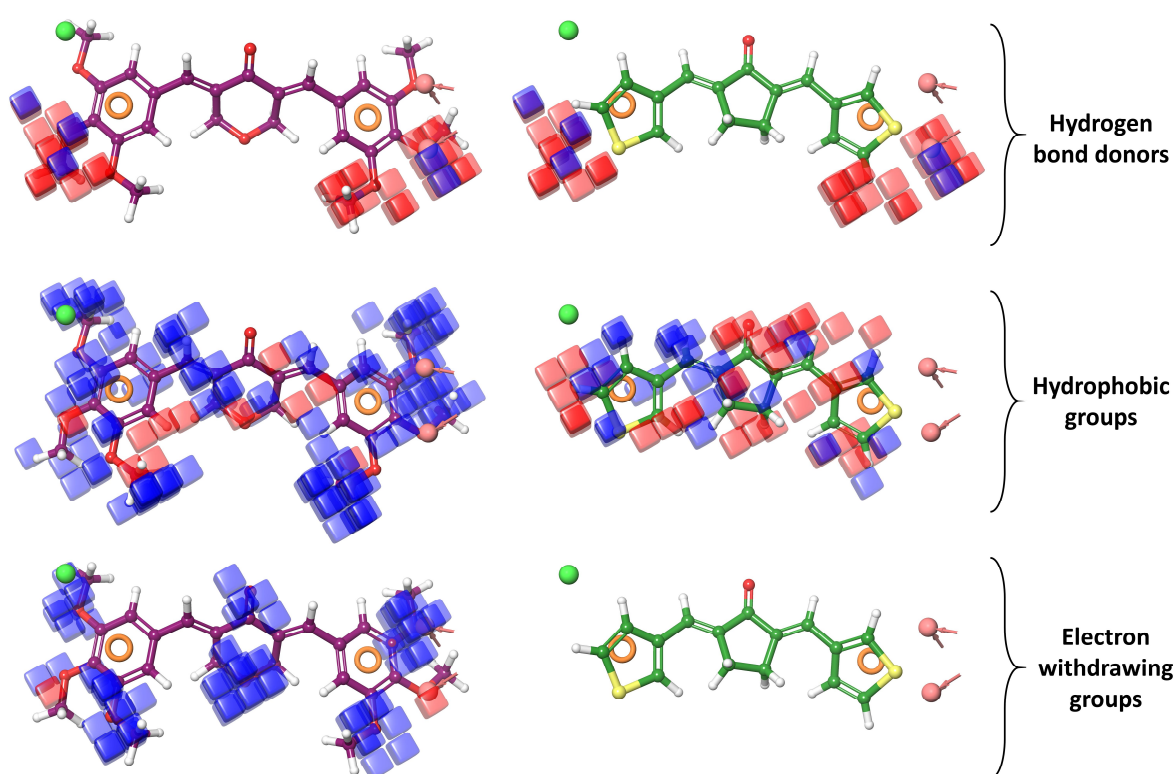


Figure III.7: Contour visualization of H-bond donors, hydrophobic, and electron-withdrawing groups in the most active (plum) and least active (green) compounds.

For the electron-withdrawing contour map, the analysis reveals that E10 exhibits clusters of blue cubes around the methoxy substituents at the meta positions on the phenol rings and around the tetrahydropyran-4-one cyclic linker. This pattern indicates that electron-withdrawing groups at these positions are favorable for enhancing the compound's activity. In contrast, the least active compound, which possesses a cyclopentanone linker and is devoid of methoxy substituents, shows a total absence of blue cubes. This absence suggests

a deficiency in electron-withdrawing groups, correlating with its lower activity and decreased potency.

This comparison demonstrates how the introduction of different cyclic ketones as linkers, and the presence and position of certain substituents such as methoxy groups, can significantly modulate the biological activity of curcumin derivatives. Future work should continue to explore these structural elements to develop more potent and effective therapeutic agents.

### III.3.4 Pharmacophore-based virtual screening

The versatile Phase system (v6.5) from the Schrödinger software suite was integral to our research. It facilitated key processes including pharmacophore modeling, 3D-QSAR development, and the creation and screening of a 3D compound library [7]. Initially, we compiled a dataset comprising 2,955 curcumin derivatives from PubChem. These compounds underwent a preliminary screening process to assess their compliance with Lipinski's rule of five, along with other pharmacokinetic criteria, resulting in a reduced dataset of 1,208 compounds. Subsequently, the AAHRR_1 pharmacophore model was employed as a 3D query to further refine this dataset, requiring that the structures match at least four out of the five pharmacophore features. This rigorous selection process yielded 1,065 compounds, which were then subjected to molecular docking studies for further analysis.

### III.3.5 Molecular docking-based screening

The previously identified 1,065 compounds from the pharmacophore-based virtual screening were processed through a multi-tiered docking protocol using Glide (v8.9). The reliability of the docking procedure was validated by superimposing the co-crystallized ligand with its redocked pose, resulting in a root mean square deviation of 0.137 Å. This low RMSD value demonstrates the high accuracy of the docking process, as illustrated in Figure III.8.
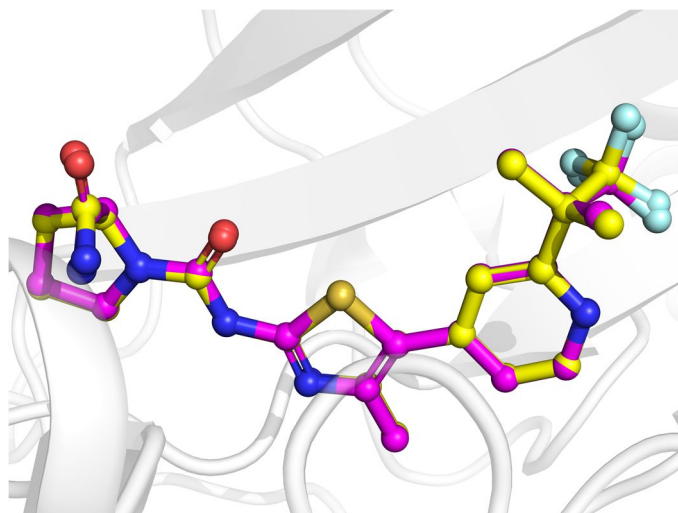
Figure III.8: Superposition of the docked pose (magenta) of Alpelisib with its crystal structure conformation (yellow).

The initial screening phase utilized high-throughput virtual screening (HTVS), which successfully narrowed the pool of compounds to 212 ligands. This was followed by standard precision (SP) docking, further refining the list to 41 ligands. In the final phase, extra precision (XP) docking was employed, resulting in the identification of seven promising hits as potential PI3Kα inhibitors.

The docking scores and free binding energy values for the hit compounds, along with the reference drug Alpelisib, are presented in Table III.6. The hit compounds demonstrated lower docking scores (from −10.926 to −9.882 kcal/mol) compared to Alpelisib (−7.810 kcal/mol), indicating more favorable interactions with the target protein. Additionally, MM-GBSA calculations revealed free binding energy scores ranging from −57.42 to −46.22 kcal/mol, with five compounds showing higher binding affinities to PI3Kα than Alpelisib (−47.31 kcal/mol), supporting their potential as PI3Kα inhibitors.

For curcumin derivatives, the hydroxyl and methoxy substituents are instrumental in forming hydrogen bonds, predominantly engaging with conserved residues such as Gln859, Asp933, and Lys802. Notably, compounds CID118729404 and CID91333351 demonstrate an extended hydrogen bonding network, forming additional interactions with Tyr836, Ser854, and Asp810.

The carbonyl moiety consistently participates in hydrogen bonding with Val851 in all the evaluated compounds. However, an exception occurs with CID91451674, where the interaction is mediated by the hydroxyl group of the β-hydroxymethyl substituent on the α,β-

unsaturated carbonyl system, rather than by the carbonyl oxygen. Notably, the same hydroxyl group forms an additional unique hydrogen bond with Glu849, further distinguishing the binding mode of CID91451674 from that of the other derivatives.

The aromatic rings provide molecular stabilization through π-π stacking interactions, primarily with the indole ring of the hydrophobic residue Trp780. An exception to this pattern is observed with compound CID68556085, which instead engages in π-π interactions with the imidazole ring of the polar residue His855. Furthermore, compounds CID91451674 and CID145605201 exhibit additional π-cation interactions between their phenolic moieties and the guanidinium group of Arg770.

In contrast, Alpelisib exhibits unfavorable steric interactions with Lys802 and Val851 (see Figure III.16; steric clashes are highlighted in orange) and forms only three hydrogen bonds with the Ser854 and Gln859 residues.

Table III.6: Molecular docking results.

| Compounds | XP score (kcal/mol) | MM-GBSA (kcal/mol) | Hydrogen interactions | Hydrophobic interactions |
|---|---|---|---|---|
| CID91451674 | −10.926 | −50.24 | Lys802, Glu849, Gln859, Asp933, Val851 | Trp780, Arg770 |
| CID156189304 | −10.725 | −51.48 | Lys802, Val851, Gln859, Asp933 | Trp780 |
| CID68556085 | −10.523 | −57.42 | Val851, Lys802 | His855 |
| CID118729404 | −10.247 | −56.66 | Ser854, Lys802, Val851, Tyr836, Asp810 | Trp780 |
| CID154728220 | −10.217 | −56.16 | Gln859, Lys802, Asp933, Val851 | Trp780 |
| CID91333351 | −10.184 | −43.60 | Asp933, Val851, Tyr836, Ser854, Asp810, Gln859 | Trp780 |
| CID145605201 | −9.882 | −46.22 | Lys802, Gln859, Val851, Asp933 | Trp780, Arg770 |
| Alpelisib | −7.810 | −47.31 | Ser854, Gln859 | - |
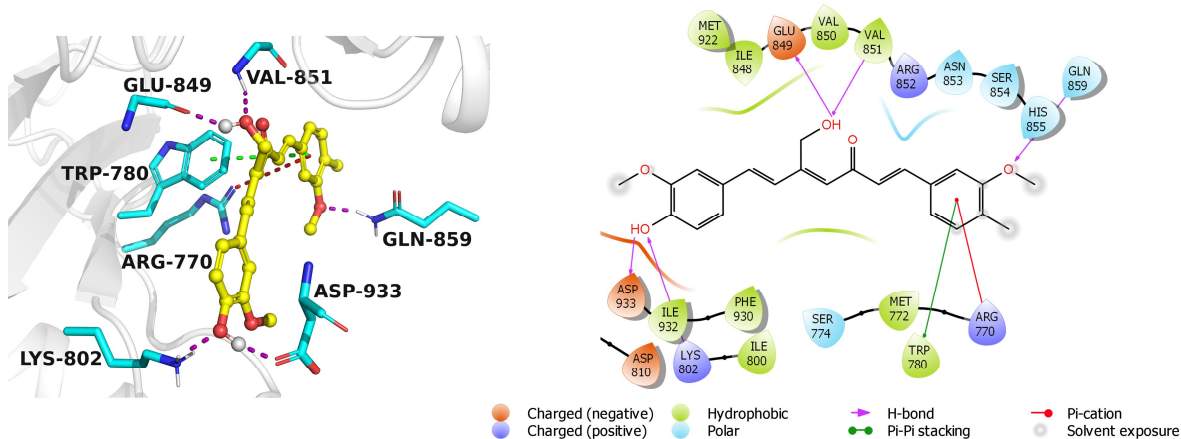
Figure III.9: 2D and 3D ligand interaction diagrams of the CID91451674 compound within the binding pocket of PI3Kα.
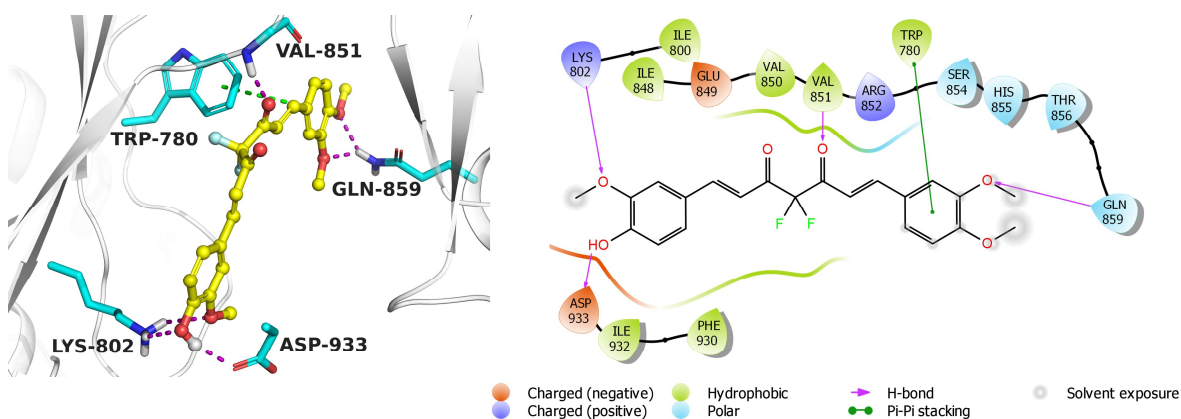


Figure III.10: 2D and 3D ligand interaction diagrams of the CID156189304 compound within the binding pocket of PI3Kα.
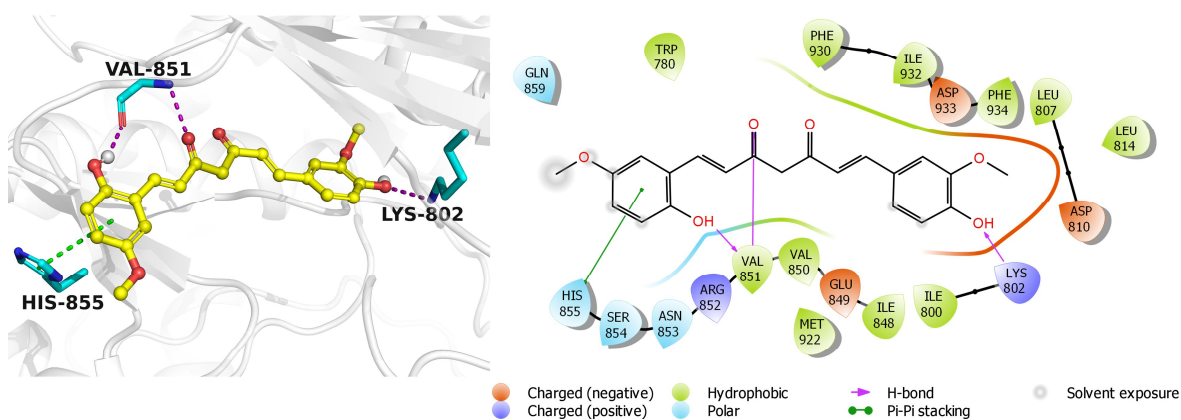


Figure III.11: 2D and 3D ligand interaction diagrams of the CID68556085 compound within the binding pocket of PI3Kα.
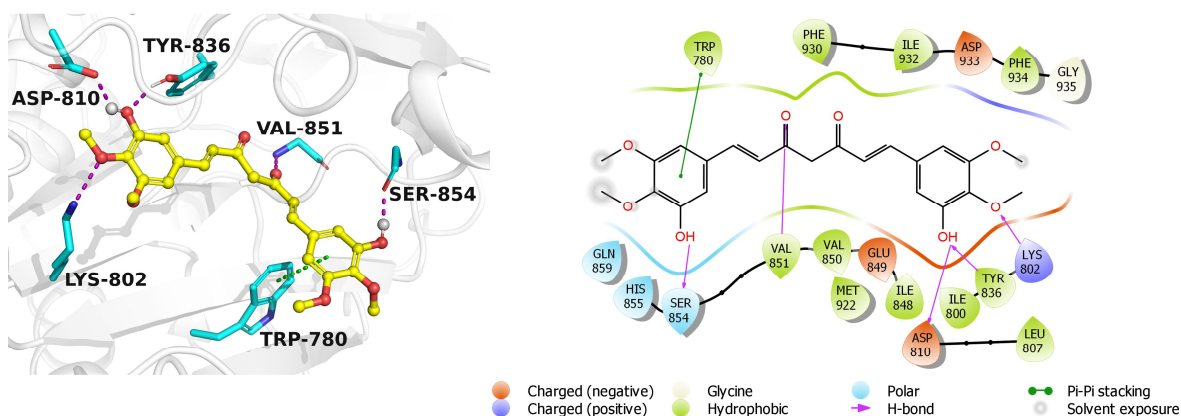
Figure III.12: 2D and 3D ligand interaction diagrams of the CID118729404 compound within the binding pocket of PI3Kα.
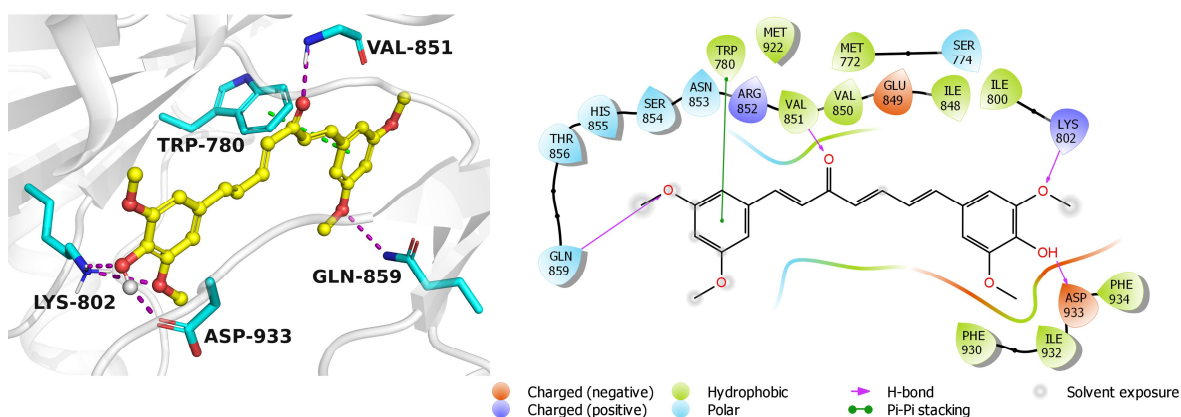


Figure III.13: 2D and 3D ligand interaction diagrams of the CID154728220 compound within the binding pocket of PI3Kα.
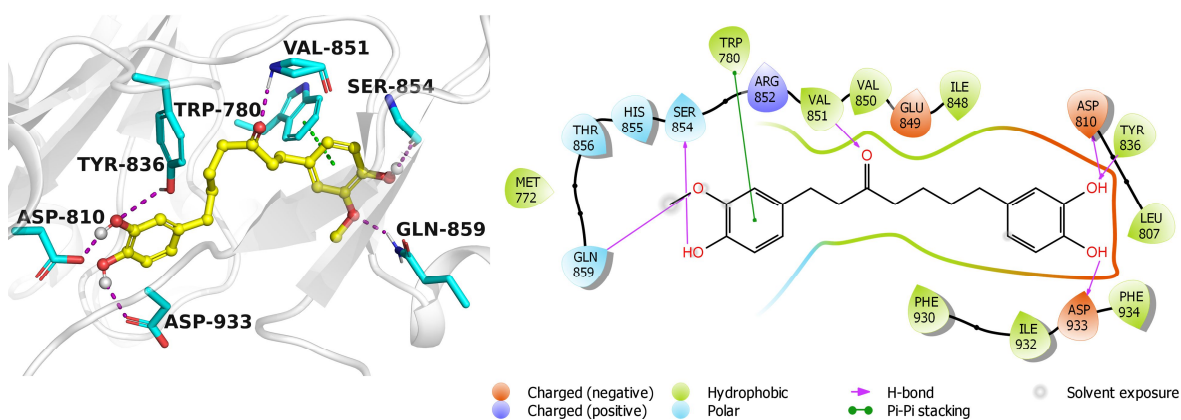


Figure III.14: 2D and 3D ligand interaction diagrams of the CID91333351 compound within the binding pocket of PI3Kα.
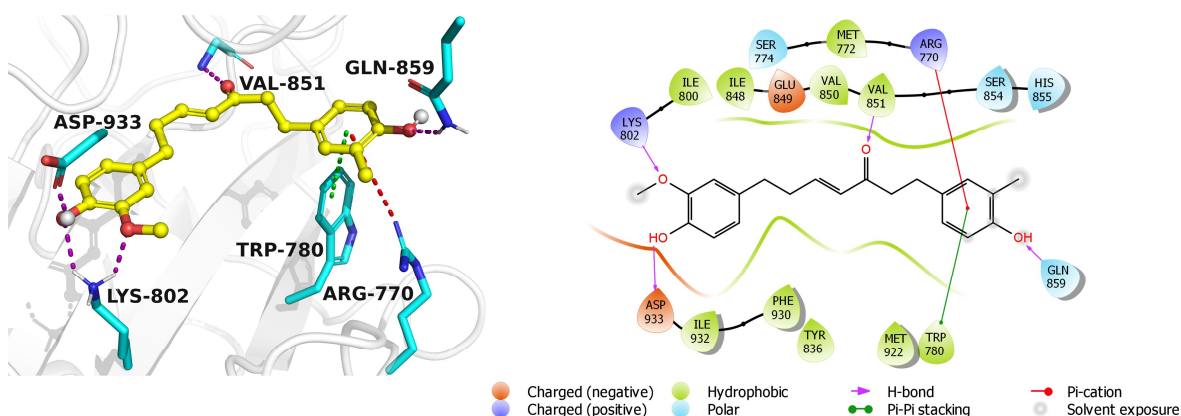
Figure III.15: 2D and 3D ligand interaction diagrams of the CID145605201 compound within the binding pocket of PI3Kα.
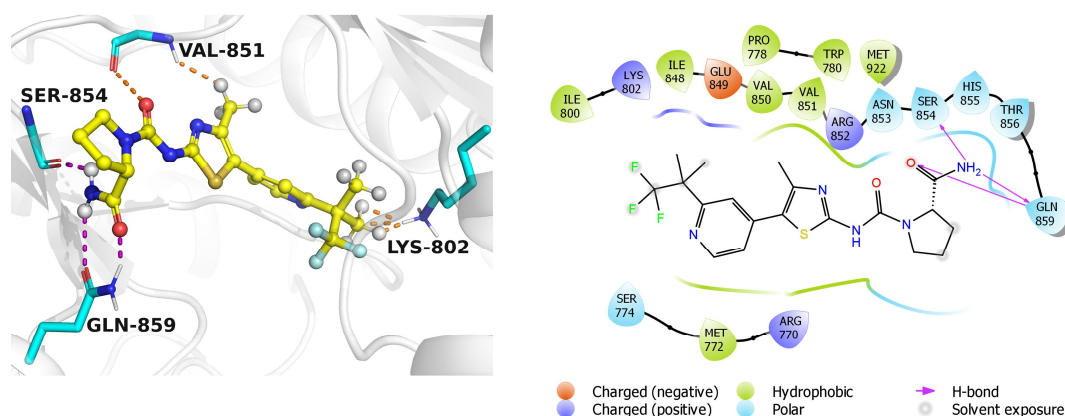


Figure III.16: 2D and 3D ligand interaction diagrams of Alpelisib within the binding pocket of PI3Kα.

### III.3.6    ADMET analysis

In our comprehensive assessment of the pharmacokinetic and toxicity profiles of our hit compounds, we made use of the QikProp module and pkCSM web server. The results, detailed in Tables III.7 and III.8, provide an insightful comparison against the reference drug Alpelisib.

Jorgensen's Rule of Three (RO3) is a critical metric for predicting oral bioavailability, stipulating that a compound should have a predicted aqueous solubility (QPlogS) greater than −5.7, predicted Caco-2 cell permeability (QPPCaco) above 22 nm/s, and fewer than seven primary metabolites [28]. Compounds CID91451674, CID156189304, and CID154728220 demonstrated high permeability rates, with QPPCaco values exceeding 500

nm/s, indicating strong potential for oral absorption. However, CID154728220 violated the RO3 with a QPlogS of −5.839, and several compounds exceeded the acceptable number of primary metabolites. Nevertheless, none of the compounds fell outside the 95% range for known drug properties (#stars = 0), suggesting good overall drug-likeness. The compounds demonstrated higher oral absorption percentages compared to Alpelisib with four of them exhibiting a 100% absorption rate.

The QPlogBB parameter predicts the brain/blood partition coefficient and is crucial for assessing the ability of compounds to cross the blood-brain barrier (BBB). QPlogBB values typically range from −3.0 to −1.2, with higher values indicating better potential for BBB penetration. The CNS activity scores complement this by indicating the predicted central nervous system activity, ranging from −2 (inactive) to 2 (active). Alpelisib, with a QPlogBB of −0.800 and a CNS value of −1, is predicted to have high brain permeability and low CNS activity. The other compounds, with QPlogBB values between −1.421 and −2.248, show lower brain permeability and no predicted CNS activity.

Table III.7: Predicted ADME properties of the studied compounds.

| Compounds | #stars | ROF | QPlogS | QPPCaco | #metab | CNS | QPlogBB | QPPMDCK | QPlogKhsa | Oral Absorption |
|---|---|---|---|---|---|---|---|---|---|---|
| CID91451674 | 0 | 0 | −5.558 | 565.372 | 5 | −2 | −1.715 | 267.087 | 0.357 | 100.000 |
| CID156189304 | 0 | 0 | −5.382 | 531.440 | 4 | −2 | −1.527 | 474.679 | 0.198 | 100.000 |
| CID68556085 | 0 | 0 | −4.304 | 209.891 | 5 | −2 | −2.049 | 91.518 | −0.043 | 84.952 |
| CID118729404 | 0 | 0 | −4.656 | 201.969 | 7 | −2 | −2.248 | 87.790 | −0.020 | 85.917 |
| CID154728220 | 0 | 0 | −5.839 | 984.135 | 5 | −2 | −1.421 | 486.231 | 0.503 | 100.000 |
| CID91333351 | 0 | 0 | −3.552 | 122.734 | 8 | −2 | −2.038 | 51.242 | 0.080 | 80.401 |
| CID145605201 | 0 | 0 | −5.675 | 418.638 | 8 | −2 | −1.622 | 193.020 | 0.546 | 100.000 |
| Alpelisib | 0 | 0 | −4.279 | 131.765 | 6 | −1 | −0.800 | 927.116 | −0.149 | 78.751 |

#stars :  Number of property values outside the 95% range of known drugs
ROF :  Number of violations of Lipinski's rule of five
QPlogS : Predicted aqueous solubility
QPPCaco : Predicted apparent Caco-2 cell permeability
#metab : Number of likely metabolic reactions
CNS : Predicted central nervous system activity
QPlogBB : Predicted brain/blood partition coefficient
QPPMDCK : Predicted apparent MDCK cell permeability
QPlogKhsa : Prediction of binding to human serum albumin

To further evaluate the permeability of these compounds, the QPPMDCK parameter is used, reflecting permeability through MDCK (Madin-Darby Canine Kidney) cell membranes [29]. QPPMDCK values range from less than 25 nm/sec (poor permeability) to greater than 500 nm/sec (high permeability). CID156189304 and CID154728220, with QPPMDCK values of 474.679 and 486.231 nm/sec, respectively, indicate moderate to high permeability and good oral absorption. Alpelisib has a much higher QPPMDCK value of 927.116 nm/sec, suggesting excellent membrane permeability and potential BBB penetration, consistent with its high QPlogBB value. While high permeability benefits drug absorption, it also raises concerns about unintended BBB penetration and potential off-target effects, necessitating a balanced approach in drug design.

QPlogKhsa predicts the binding affinity to human serum albumin (HSA), with an acceptable range of −1.5 to 1.5. All compounds fall within this range, indicating adequate binding affinity to HSA which supports balanced systemic distribution and bioavailability.

Table III.8: Predicted toxicity profiles of the studied compounds.

| Compounds | Ames toxicity | MTD (log mg/kg/day) | hERG I inhibitor | hERG II inhibitor | $LD_{50}$ (mol/kg) | LOAEL (log mg/kg_bw/day) | Hepatotoxicity | Skin sensitization |
|---|---|---|---|---|---|---|---|---|
| CID91451674 | No | −0.047 | No | Yes | 2.095 | 2.013 | No | No |
| CID156189304 | No | 0.171 | No | No | 2.293 | 1.375 | No | No |
| CID68556085 | No | −0.455 | No | Yes | 1.983 | 1.598 | No | No |
| CID118729404 | No | 0.257 | No | Yes | 1.685 | 2.119 | No | No |
| CID154728220 | No | 0.503 | No | No | 2.157 | 2.041 | No | No |
| CID91333351 | No | −0.017 | No | Yes | 2.102 | 2.731 | No | No |
| CID145605201 | No | 0.344 | No | Yes | 2.004 | 1.301 | Yes | No |
| Alpelisib | No | 0.042 | No | No | 2.326 | 0.524 | Yes | No |

MTD : Maximum Tolerated Dose
hERG I : Human Ether-à-go-go-Related Gene Channel (Type I)
hERG II : Human Ether-à-go-go-Related Gene Channel (Type II)
$LD_{50}$ : Median Lethal Dose
LOAEL : Lowest Observed Adverse Effect Level

Maximum Tolerated Dose (MTD) is the highest dose of a drug that does not cause unacceptable side effects or overt toxicity. Alpelisib has an MTD of 0.042, indicating

moderate tolerance. Compounds CID156189304 (0.171), CID118729404 (0.257), CID154728220 (0.503), and CID145605201 (0.344) show higher MTD values, suggesting they can be administered at higher doses with fewer side effects, thus potentially safer for therapeutic use. In contrast, CID68556085 (−0.455), CID91451674 (−0.047), and CID91333351 (−0.017) exhibit lower MTD values, indicating higher toxicity at lower doses compared to Alpelisib.

Median Lethal Dose ($LD_{50}$) measures the dose required to kill 50% of a test population, reflecting acute toxicity. Most compounds have $LD_{50}$ values comparable to Alpelisib (2.326), with CID156189304 (2.293) and CID154728220 (2.157) being particularly close, indicating similar acute toxicity profiles.

The Lowest Observed Adverse Effect Level (LOAEL) is the minimum dose at which adverse effects are detectable. Among the compounds analyzed, Alpelisib exhibits the lowest LOAEL at 0.524, indicating that it can induce adverse effects at relatively low doses. In contrast, our hit compounds demonstrate substantially higher LOAEL values, suggesting a more favorable safety profile for potential long-term administration.

None of the compounds are Ames toxic or primary hERG I inhibitors, minimizing concerns about mutagenicity and major cardiac toxicity. However, some compounds, including CID91451674, CID68556085, CID118729404, CID91333351, and CID145605201, are flagged as hERG II inhibitors, indicating potential cardiotoxicity risks that require further exploration.

Alpelisib is predicted to be hepatotoxic, a characteristic shared only with CID145605201 among the compounds tested. This highlights a potential risk for liver toxicity that requires careful monitoring in clinical use. None of the compounds, including Alpelisib, show predicted skin sensitization, indicating a low risk of allergic skin reactions.

Based on the ADME properties and toxicity profiles, compounds CID156189304 and CID154728220 emerge as the most promising alternatives to Alpelisib. They exhibit excellent oral absorption rates, moderate to high permeability through MDCK cell membranes, and adequate binding affinity to HSA. These properties coupled with a reduced risk of CNS side effects and better overall safety profiles, make them strong candidates for further molecular dynamics (MD) simulations to evaluate their therapeutic potential.

### III.3.7    Molecular dynamics simulations

Molecular dynamics (MD) simulations offer insights into the behavior of protein-ligand complexes at the atomic level. In this study, we conducted MD simulations to examine the structural dynamics and stability of PI3Kα in complex with three ligands: CID154728220, CID156189304, and the benchmark inhibitor Alpelisib. The simulations spanned 100 ns, enabling us to observe interactions and conformational changes within each complex. We analyzed key parameters such as Root Mean Square Deviation (RMSD), Root Mean Square Fluctuation (RMSF), Radius of Gyration (Rg), and hydrogen bonding to gain a deeper understanding of the molecular behavior of these complexes.

The RMSD results indicate that the deviations of the protein-ligand complexes are minimal, with less than a 0.2 nm difference from the apoprotein. This consistency in deviation across all systems suggests that the complexes achieve a stable conformation throughout most of the simulation. The relative stability compared to the unbound protein implies that ligand binding does not significantly perturb the protein's overall structural integrity. Among the ligands, CID154728220 exhibited the most favorable stability, as evidenced by its lower RMSD profile compared to the reference compound Alpelisib. Although CID156189304 demonstrated a generally stable RMSD trajectory, it had a slightly elevated mean RMSD value relative to CID154728220, suggesting marginally higher conformational flexibility.

The RMSF analysis elucidates the flexibility of protein residues during MD simulations. Higher RMSF values indicate regions of the protein with greater flexibility, while lower RMSF values denote more rigid, stable regions. The three ligands show similar RMSF patterns, indicating that the core structure and flexibility profile of the protein are largely maintained regardless of the bound ligand. The majority of residues exhibit RMSF values below 0.2 nm, suggesting overall structural stability across all three complexes. CID154728220 complex demonstrated the highest residual stability among the three, with minimal fluctuations in most of its amino acid residues. This indicates that CID154728220 maintains a more stable interaction with the protein, contributing to lower overall flexibility and greater structural rigidity in the complex compared to Alpelisib and CID156189304.
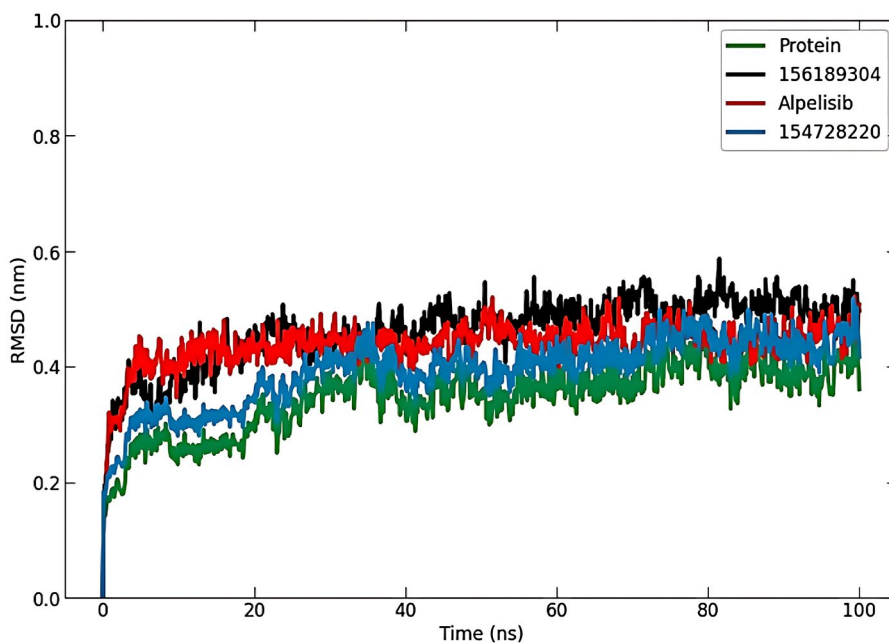
Figure III.17: RMSD analysis of the backbone atoms of PI3Kα and its complexes.
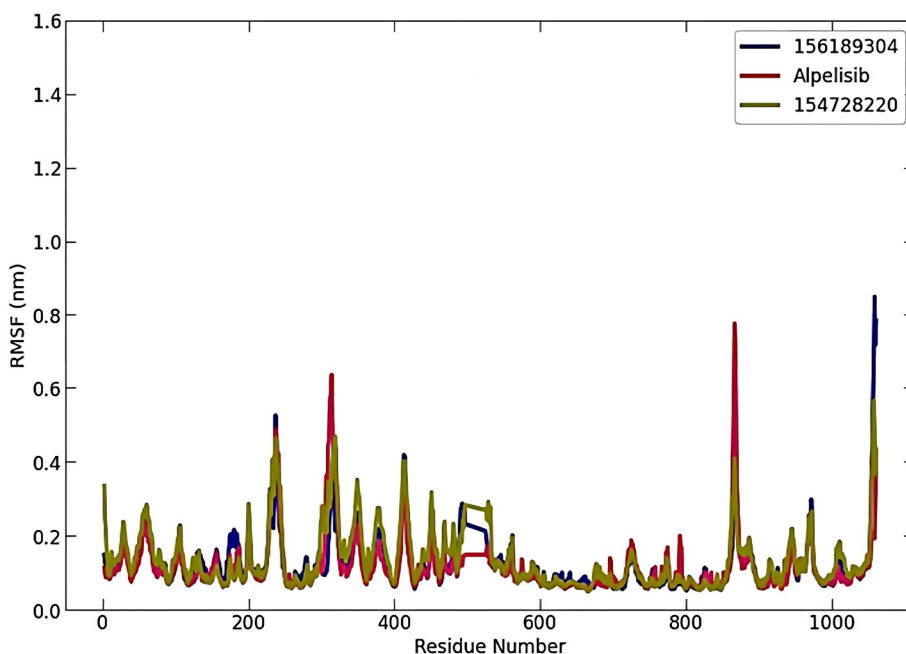


Figure III.18: RMSF per residue of PI3Kα in complex with CID156189304, CID154728220, and Alpelisib.

The Radius of Gyration (Rg) offers valuable insights into the overall compactness and conformational changes of the protein throughout the simulation. Initially, all three complexes exhibit a rapid increase in Rg during the first few nanoseconds, reflecting a

transient expansion in the protein's structure as it adjusts to ligand binding. Following this phase, the Rg values stabilize and fluctuate within a narrow range (approximately 3.20 to 3.32 nm), which indicates that the protein retains structural stability for the remainder of the simulation. Among the ligands, CID156189304 demonstrates the highest Rg values on average, implying that this ligand induces a slightly more expanded protein conformation. This could suggest weaker or less compact interactions within the binding pocket. Conversely, CID154728220 exhibits consistently lower Rg values, indicating that it promotes a more compact protein structure, potentially due to stronger interactions or more effective stabilization of the protein-ligand complex. Alpelisib displays intermediate Rg values, fluctuating between those of CID154728220 and CID156189304.
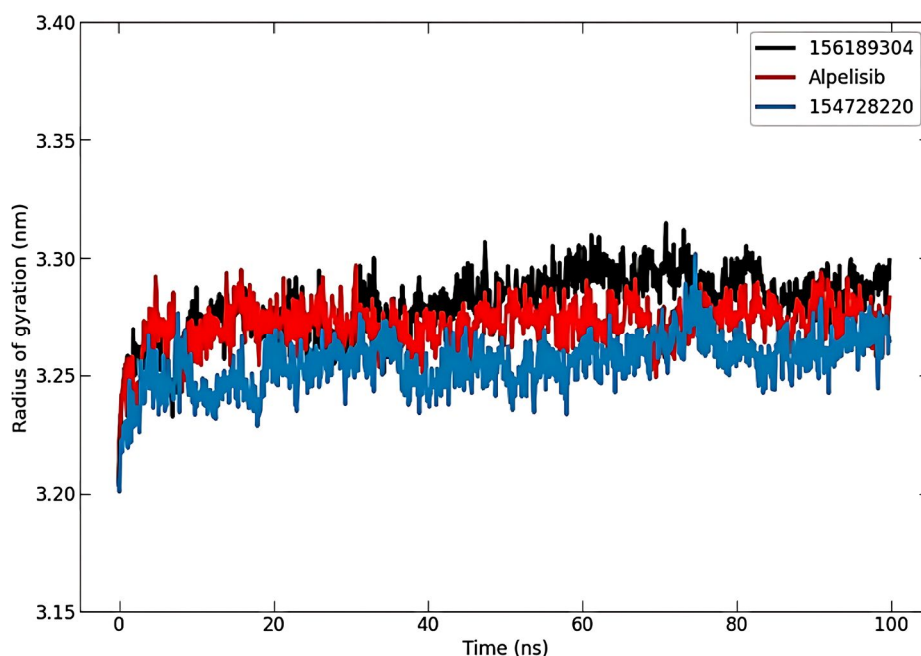


Figure III.19: Radius of gyration of protein-ligand complexes over 100 ns simulation time.

Hydrogen bond analysis reveals that the two selected compounds (CID156189304 and CID154728220) form more hydrogen bonds with the target at any given point in time compared to the standard. They maintain an average of three bonds, while the standard (Alpelisib) maintains two. This indicates a higher degree of interaction between these two compounds and the target compared to Alpelisib, suggesting that they may be more potent binders of PI3Kα.
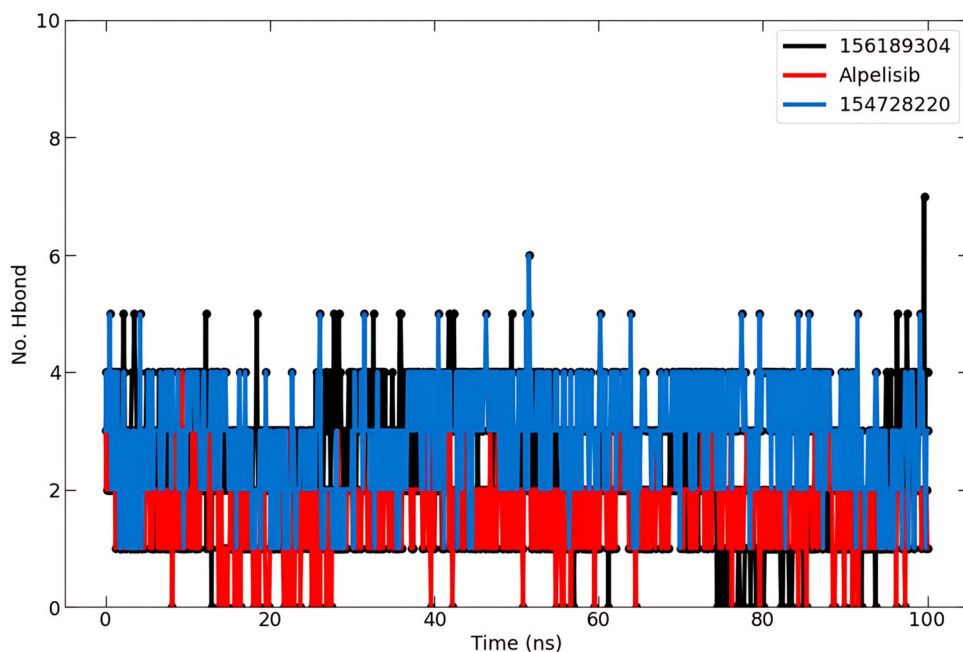
Figure III.20: Number of hydrogen bonds between protein-ligand complexes during 100 ns.
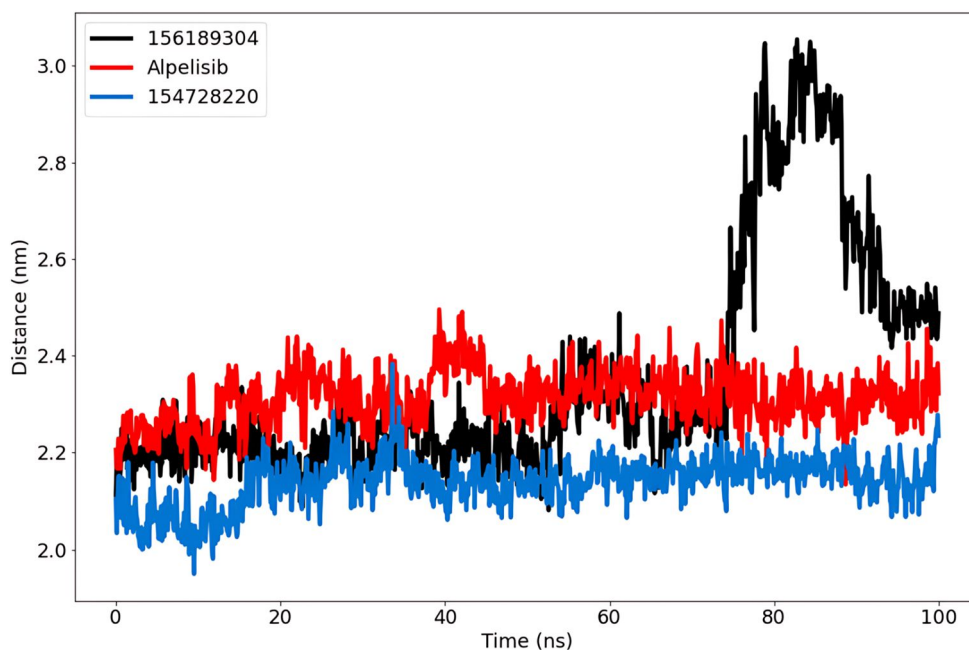


Figure III.21: Averaged ligand-protein distance for PI3Kα in complex with CID154728220, CID156189304, and Alpelisib.

The distance graph illustrates the fluctuations in the distance between the ligands and PI3Kα throughout the 100 ns MD simulation. CID154728220 consistently maintains the closest proximity to PI3Kα, indicating a stable interaction and suggesting strong binding affinity. This stability, characterized by minimal fluctuations, implies that CID154728220

may have the most favorable binding mode with a high degree of binding stability. Alpelisib, while also showing relatively stable binding, experiences more noticeable fluctuations compared to CID154728220. These variations may represent minor conformational adjustments within the protein-ligand complex. Nevertheless, the overall interaction between Alpelisib and PI3Kα remains consistent. CID156189304, on the other hand, displays a significant increase in distance around 75 ns, peaking around 85 ns before partially returning to its original proximity. This pronounced shift suggests potential ligand dissociation or a notable change in its binding conformation, indicating reduced binding stability over time. Such behavior could impact its efficacy, particularly in scenarios requiring sustained interaction.

Principal Component Analysis (PCA) was calculated using the Bio3D program in R. All principal components captured 56.6% of the structural variance in the protein. The dynamic cross-correlated motions (DCCM) of protein residues were also calculated, with colors ranging from red (positively correlated motions) to white (no correlation) to blue (negatively correlated motions).

The binding free energies of the ligand-protein complexes were estimated using the Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) method, as implemented in the g-mmpbsa tool. CID154728220 exhibited the most favorable binding free energy (−147.109 kJ/mol), followed closely by CID156189304 (−144.742 kJ/mol). Notably, both compounds demonstrated binding energies approximately 60 kJ/mol more favorable than Alpelisib (−87.010 kJ/mol), highlighting their stronger interactions with the PI3Kα active site.

Table III.9: Binding free energy and its components for PI3Kα complexed with CID154728220, CID156189304, and Alpelisib.

| Complex | $\Delta G$ | van der Waal energy | Electrostatic energy | Polar solvation energy | SASA energy |
|---|---|---|---|---|---|
| CID154728220 | −147.109 ± 24.711 | −184.329 ± 10.421 | −125.605 ± 41.048 | 185.351 ± 22.471 | −22.526 ± 1.210 |
| CID156189304 | −144.742 ± 25.004 | −150.743 ± 24.573 | −146.282 ± 56.807 | 172.024 ± 52.630 | −19.741 ± 2.592 |
| Alpelisib | −87.010 ± 25.084 | −185.684 ± 11.547 | −21.833 ± 28.418 | 141.941 ± 15.622 | −21.435 ± 1.092 |
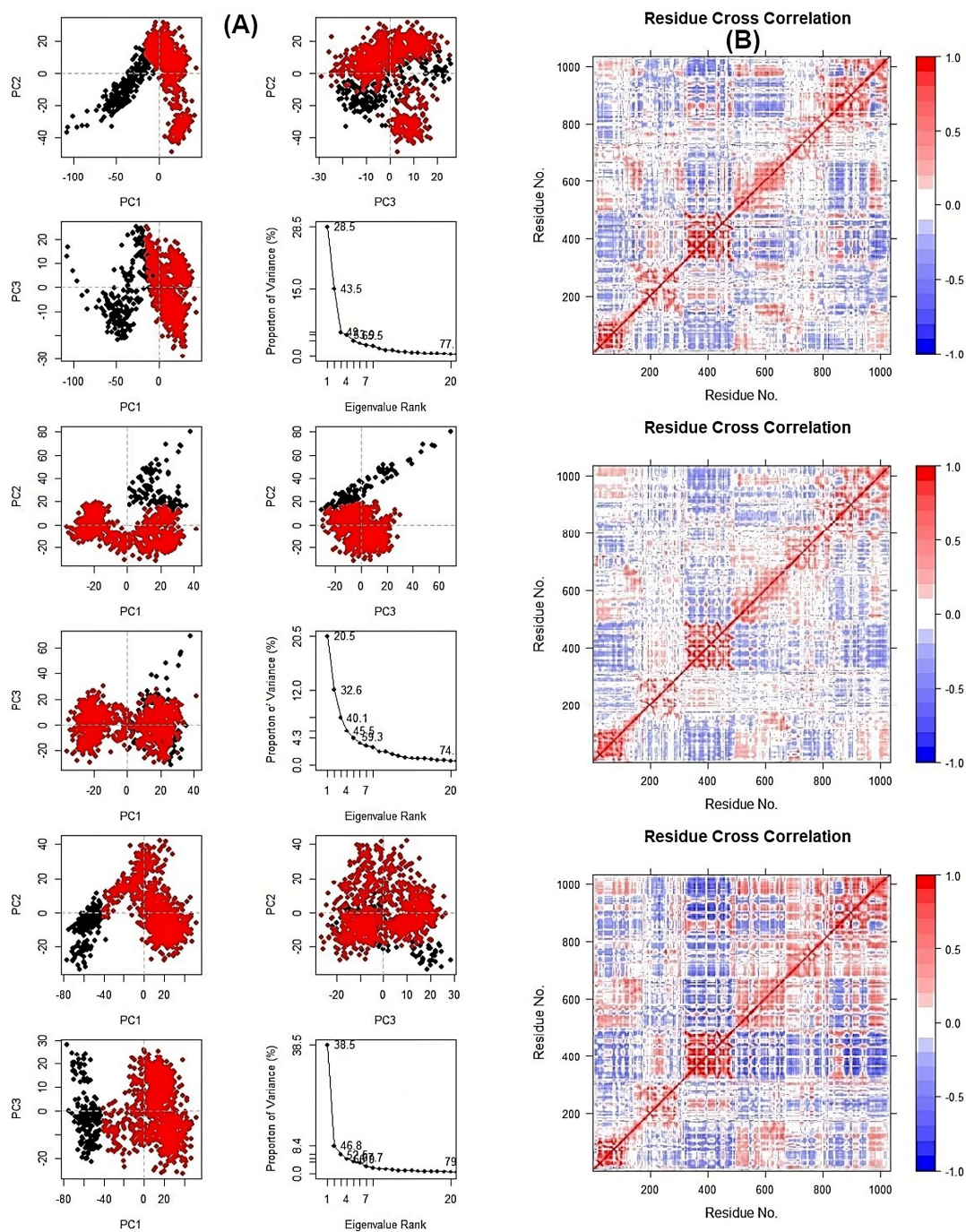
Figure III.22: Principal component analysis (a) and dynamic cross-correlated motions (b) of protein-ligand complexes.

## III.4   Conclusion

This study demonstrates the efficacy of an integrated computational approach in identifying novel curcumin-based compounds as potential PI3Kα inhibitors for pancreatic cancer treatment.

The developed pharmacophore hypothesis, AAHRR_1, demonstrated the highest survival score and exhibited strong enrichment metrics, indicating its robust predictive power in distinguishing between active and inactive compounds. The atom-based 3D-QSAR model derived from this pharmacophore exhibited strong statistical performance, including an $R^2$ value of 0.990 and a cross-validated $Q^2$ of 0.971. The model's reliability and robustness were further confirmed using the Y-randomization test and external validation.

The virtual screening protocol, employing the AAHRR_1 pharmacophore model followed by molecular docking, successfully identified promising curcumin derivatives with potentially higher binding affinities to PI3Kα than the reference inhibitor, Alpelisib. Among the top-ranked hits, two compounds (CID154728220 and CID156189304) exhibited favorable ADMET profiles, underscoring their potential as drug candidates.

Molecular dynamics simulations further validated these findings, demonstrating the stability of the ligand-protein complexes and providing atomic-level insights into binding mechanisms and conformational dynamics over time. CID154728220 emerged as the most promising PI3Kα inhibitor, closely followed by CID156189304. Both compounds are strong candidates for further optimization and experimental validation.

# References

[1] M. Zhang, H. Jang and R. Nussinov, *PI3K inhibitors: review and new strategies*, Chem. Sci. 11 (2020), pp. 5855–5865.

[2] S. Srivastava, A. Vengamthodi, I. Singh, B.S. Choudhary, M. Sharma and R. Malik, *Determination of comprehensive in silico determinants as a strategy for identification of novel PI3Kα inhibitors*, Struct. Chem. 30 (2019), pp. 1761–1778.

[3] P. Narayan, T.M. Prowell, J.J. Gao, L.L. Fernandes, E. Li, X. Jiang et al., *FDA Approval Summary: Alpelisib Plus Fulvestrant for Patients with HR-positive, HER2-negative, PIK3CA-mutated, Advanced or Metastatic Breast Cancer*, Clin. Cancer Res. 27 (2021), pp. 1842–1849.

[4] X. Wei, Z.-Y. Du, X. Zheng, X.-X. Cui, A.H. Conney and K. Zhang, *Synthesis and evaluation of curcumin-related compounds for anticancer activity*, Eur. J. Med. Chem. 53 (2012), pp. 235–245.

[5] J.C. Shelley, A. Cholleti, L.L. Frye, J.R. Greenwood, M.R. Timlin and M. Uchimaya, *Epik: a software program for pK a prediction and protonation state generation for drug-like molecules*, J. Comput. Aided Mol. Des. 21 (2007), pp. 681–691.

[6] K. Roos, C. Wu, W. Damm, M. Reboul, J.M. Stevenson, C. Lu et al., *OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules*, J. Chem. Theory Comput. 15 (2019), pp. 1863–1874.

[7] S.L. Dixon, A.M. Smondyrev, E.H. Knoll, S.N. Rao, D.E. Shaw and R.A. Friesner, *PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results*, J. Comput. Aided Mol. Des. 20 (2006), pp. 647–671.

[8] D.K. Yadav, S. Kumar, M.K. Teli and M.-H. Kim, *Ligand-based pharmacophore modeling and docking studies on vitamin D receptor inhibitors*, J. Cell. Biochem. 121 (2020), pp. 3570–3583.

[9] A. Golbraikh and A. Tropsha, *Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection*, J. Comput. Aided Mol. Des. 16 (2002), pp. 357–369.

[10] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He et al., *PubChem 2019 update: improved access to chemical data*, Nucleic Acids Res. 47 (2019), pp. D1102–D1109.

[11] S.L. Dixon, A.M. Smondyrev and S.N. Rao, *PHASE: A Novel Approach to Pharmacophore Modeling and 3D Database Searching*, Chem. Biol. Drug Des. 67 (2006), pp. 370–372.

[12] P. Furet, V. Guagnano, R.A. Fairhurst, P. Imbach-Weese, I. Bruce, M. Knapp et al., *Discovery of NVP-BYL719 a potent and selective phosphatidylinositol-3 kinase alpha inhibitor selected for clinical evaluation*, Bioorg. Med. Chem. Lett. 23 (2013), pp. 3741–3748.

[13] G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, *Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments*, J. Comput. Aided Mol. Des. 27 (2013), pp. 221–234.

[14] M.P. Jacobson, D.L. Pincus, C.S. Rapp, T.J.F. Day, B. Honig, D.E. Shaw et al., *A hierarchical approach to all-atom protein loop prediction*, Proteins 55 (2004), pp. 351–367.

[15] M. Rostkowski, M.H.M. Olsson, C.R. Søndergaard and J.H. Jensen, *Graphical analysis of pH-dependent properties of proteins predicted using PROPKA*, BMC Struct. Biol. 11 (2011), pp. 6.

[16] T.A. Halgren, R.B. Murphy, R.A. Friesner, H.S. Beard, L.L. Frye, W.T. Pollard et al., *Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening*, J. Med. Chem. 47 (2004), pp. 1750–1759.

[17] J. Li, R. Abel, K. Zhu, Y. Cao, S. Zhao and R.A. Friesner, *The VSGB 2.0 model: a next generation energy model for high resolution protein structure modeling*, Proteins 79 (2011), pp. 2794–2812.

[18] M.F. Khan, G. Verma, P. Alam, M. Akhter, M.A. Bakht, S.M. Hasan et al., *Dibenzepinones, dibenzoxepines and benzosuberones based p38α MAP kinase inhibitors: Their pharmacophore modelling, 3D-QSAR and docking studies*, Comput. Biol. Med. 110 (2019), pp. 175–185.

[19] D. Sun, W. Gao, H. Hu and S. Zhou, *Why 90% of clinical drug development fails and how to improve it?*, Acta Pharm. Sin. B 12 (2022), pp. 3049–3062.

[20] G. Verma, M.F. Khan, W. Akhtar, M.M. Alam, M. Akhter, O. Alam et al., *Pharmacophore modeling, 3D-QSAR, docking and ADME prediction of quinazoline based EGFR inhibitors*, Arab. J. Chem. 12 (2019), pp. 4815–4839.

[21] A. Golbraikh and A. Tropsha, *Beware of q2!*, J. Mol. Graph. Model. 20 (2002), pp. 269–276.

[22] H.A. Abou-Zied, B.G.M. Youssif, M.F.A. Mohamed, A.M. Hayallah and M. Abdel-Aziz, *EGFR inhibitors and apoptotic inducers: Design, synthesis, anticancer activity and docking studies of novel xanthine derivatives carrying chalcone moiety as hybrid molecules*, Bioorganic Chem. 89 (2019), pp. 102997.

[23] M. Hisham, B.G.M. Youssif, E.E.A. Osman, A.M. Hayallah and M. Abdel-Aziz, *Synthesis and biological evaluation of novel xanthine derivatives as potential apoptotic antitumor agents*, Eur. J. Med. Chem. 176 (2019), pp. 117–128.

# Chapter IV

# Application 2 – Targeting KRAS G12D Using Fragment-Based Drug Discovery

## IV.1 Introduction

The development of the first KRAS G12C inhibitors represents a significant breakthrough in targeting the previously "undruggable" KRAS protein. Sotorasib, marketed as Lumakras by Amgen, received FDA approval on May 28, 2021, as the first targeted therapy for adult patients with KRAS G12C-mutated locally advanced or metastatic non-small cell lung cancer [1]. Following its lead, Adagrasib, developed by Mirati Therapeutics and marketed as Krazati, also gained approval on December 12, 2022. Both inhibitors covalently bind to the KRAS G12C mutation, locking the protein in its inactive GDP-bound state and effectively inhibiting its oncogenic signaling [2].

### 30 057 Fragments
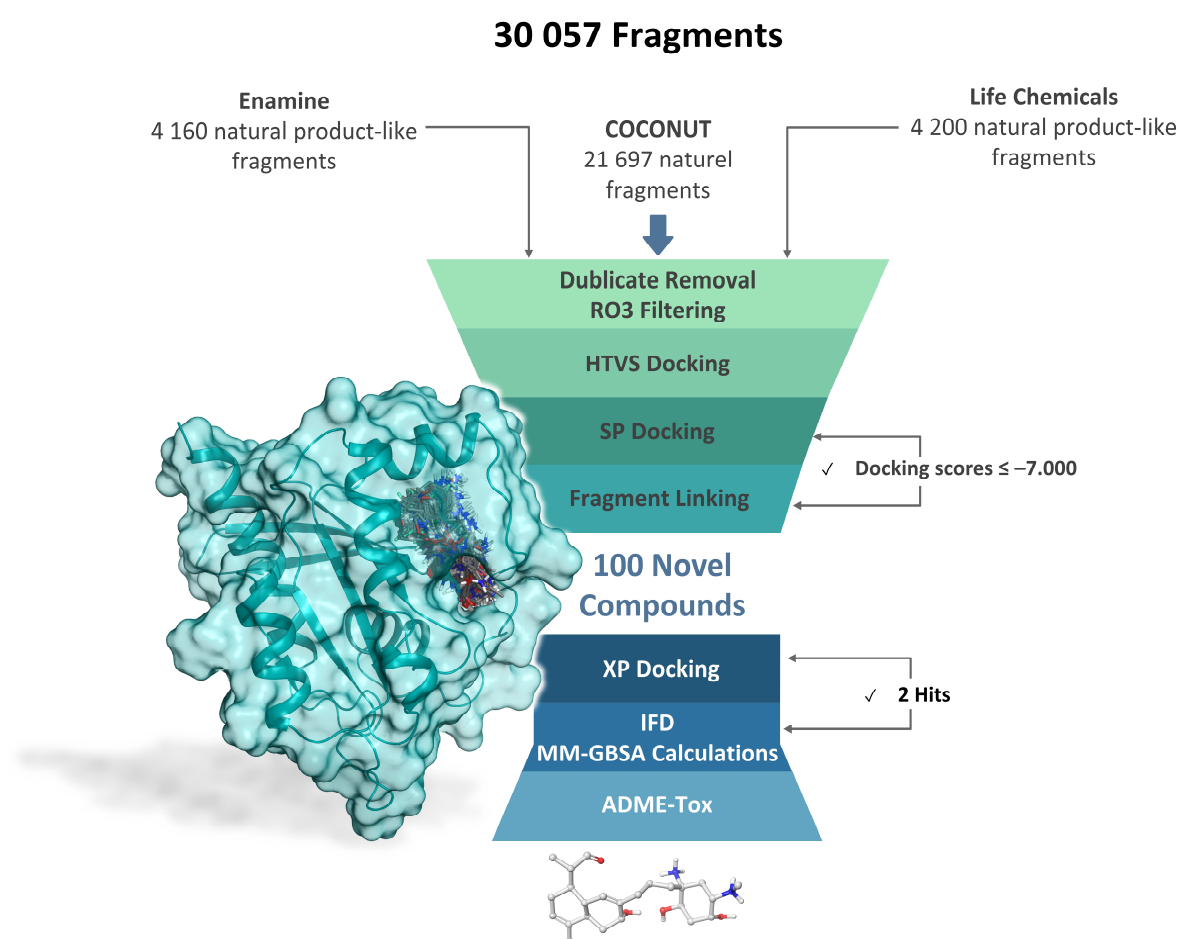


Figure IV.1: Schematic representation of the workflow.

Unfortunately, KRAS G12C inhibitors are not effective against other KRAS mutations such as G12D and G12V due to the distinct chemistries of these variants. Nevertheless, the clinical success of Sotorasib and Adagrasib has provided valuable insights, accelerating the design of inhibitors specifically tailored to target a broader spectrum of mutant KRAS

isoforms. A promising candidate emerging from these efforts is MRTX-1133, a noncovalent, potent, and selective inhibitor designed to target KRAS G12D. MRTX-1133 has shown great promise in preclinical studies and is currently undergoing clinical development for the treatment of solid tumors, including pancreatic and colorectal cancers [3,4].

In an attempt to address the unmet need for a wider range of KRAS inhibitors and to expand treatment options for pancreatic cancer, this chapter discusses the discovery and development of a potential KRAS G12D inhibitor using fragment-based drug discovery (Figure IV.1).

## IV.2    Material and methods

All computational studies discussed in this chapter were conducted on a laptop equipped with an Intel Core i5 processor (1.60 GHz), 16.0 GB RAM, and a 64-bit Windows 10 operating system.

### IV.2.1    Protein preparation

The crystal structure of KRAS G12D bound to MRTX-1133 (PDB ID: 7RPZ) [4], with a resolution of 1.30 Å, was retrieved from the Protein Data Bank (PDB) (https://www.rcsb.org/). The structure was prepared using the Protein Preparation Wizard of the Schrödinger suite [5]. The preparation steps included bond order assignment, addition of hydrogen atoms, removal of water molecules and cofactors, and the generation of het states at a pH of 7.4. Additionally, missing side chains and loops were modeled using Prime [6]. Following this, the complex underwent restrained minimization employing the OPLS3e force field [7] until it reached an RMSD cut-off of 0.30 Å. The prepared complex was employed to generate a receptor grid, with the co-crystallized ligand serving as the centroid. The X, Y, and Z grid coordinates were set at 2.62, 5.19, and −23.01, respectively.

### IV.2.2    Fragment libraries preparation

The incorporation of natural product scaffolds into fragment-based drug discovery offers the potential to explore chemically validated regions of space that are often inaccessible to conventional synthetic molecules [8,9]. To exploit this opportunity, we compiled a diverse set of 30,057 natural and natural product-like fragments sourced from Life Chemicals, Enamine, and COCONUT databases.

After eliminating duplicates using Canvas software [10], we applied the Rule of Three filtration criteria (MW≤ 300 Da, logP≤ 3, HBD≤ 3, HBA≤ 3) to identify fragments suitable for FBDD [11]. This process refined our dataset to 20,545 unique fragments. Each fragment was then prepared using the Schrödinger LigPrep module, with ionization states generated at a pH of 7.4 using Epik [12]. Structures were constrained to a maximum of four stereoisomers, followed by energy minimization employing the OPLS3e force field [7].

### IV.2.3    Fragment screening

Using the Virtual Screening Workflow of the Glide module [13], the prepared fragments were initially docked into the KRAS G12D binding site using the high-throughput virtual screening (HTVS) mode. Subsequently, the recovered fragments from the HTVS were subjected to a second round of docking, using the more accurate standard precision (SP) mode [14].

### IV.2.4    Fragment linking

The top fragments with SP docking scores ≤−7.000, underwent three rounds of linking using the "Combine Fragments" tool within the Schrödinger suite [15]. This tool facilitates the direct joining of fragments that are positioned at various regions of the target binding site, enabling the design of novel compounds with favorable interactions. Typically, the tool connects the fragments by identifying feasible bonds that can be formed between them [16]. Default parameters were applied, including a maximum bond angle deviation of 15°, atom-atom distance capped at 1 Å, a minimum centroid distance between fragments set at 2 Å, and a fragment atom count not exceeding 200.

### IV.2.5    Molecular docking

The newly formed compounds resulting from the fragment linking process, along with MRTX-1133, were prepared using the Ligprep module, following the same parameters employed in the initial fragment preparation. These compounds were then subjected to molecular docking within the KRAS G12D binding site, utilizing the extra precision (XP) mode of the Glide module [13]. Finally, the docking results were visualized and analyzed using the Maestro interface (Schrödinger Release 2020-3) and PyMOL v2.5 software.

## IV.2.6  Induced Fit Docking and MM-GBSA calculations

For a better understanding of the protein-ligand interactions of the docked compounds, we performed flexible docking using the Induced Fit Docking (IFD) module of the Schrödinger suite. The IFD extended sampling protocol was employed, which generates up to 80 poses using automatic docking settings. The IFD protocol begins with an initial docking using Glide with softened van der Waals potentials to generate a diverse set of ligand poses. For each of these poses, Prime is employed to predict and optimize the side-chain conformations of residues within 5 Å of the ligand to accommodate the ligand binding. Both the ligand and the nearby receptor residues are then minimized to reflect the induced fit conformation. Finally, each ligand is re-docked into its corresponding low-energy protein structures, and the resulting complexes are scored and ranked using a combined scoring function known as IFDScore, which integrates GlideScore and the Prime energy terms [17]. The IFDScore is calculated as:

$$\text{IFDScore} = 1.0 \times \text{Prime Energy} + 9.057 \times \text{GlideScore} + 1.428 \times \text{Glide Ecoul}$$

The retrieved poses for each ligand were subjected to free binding energy calculations using the Molecular mechanics with generalized Born and surface area solvation (MM-GBSA) method.

## IV.2.7  ADMET prediction

The ADMET properties of the compounds were predicted using the SwissADME, Protox 3.0, and Pred-hERG web tools. SwissADME was employed to evaluate the pharmacokinetic properties, drug-likeness, and medicinal chemistry friendliness of the compounds [18]. Key parameters such as water solubility, gastrointestinal absorption, blood-brain barrier permeability, and cytochrome P450 interactions were assessed. Protox 3.0 was used to predict the toxicity of the compounds, including $LD_{50}$ values and various toxicity endpoints such as hepatotoxicity, carcinogenicity, mutagenicity, and cytotoxicity [19]. Pred-hERG was employed to predict the potential for cardiac toxicity through hERG channel inhibition [20].

## IV.3  Results and discussion

### IV.3.1  Fragments screening and linking

Initially, we obtained over thirty thousand natural and natural product-like fragments from Life Chemicals, COCONUT, and Enamine libraries. These fragments were subjected

to a screening process to eliminate duplicates, followed by filtering based on the Rule of Three criteria using Canvas software. This process resulted in 20,545 distinct fragments. The filtered fragments were processed using the LigPrep module, generating 72,176 conformations. These conformations were initially screened using the HTVS mode of the Glide module. From the top 40,627 fragments, further docking was performed using the SP mode, identifying 4,529 fragments with SP docking scores ≤−7.000 kcal/mol. Subsequently, the selected fragments underwent a three-round linking process using the "Combine Fragments" tool, leading to the development of 100 entirely novel compounds. The newly designed structures were then processed with the LigPrep module, using the same parameters as in the initial fragment preparation, before being submitted to the XP docking protocol.

### IV.3.2    Analysis and visualization of molecular docking results

Molecular docking was conducted using the Glide module in extra precision (XP) mode to evaluate the binding affinity of the newly designed compounds within the KRAS G12D binding site. To validate the docking protocol, MRTX-1133 was initially extracted from its binding site and subsequently redocked into the same position. The superposition of the docked pose onto the crystallographically determined pose resulted in a RMSD value of 1.223 Å, as shown in Figure IV.2. This value, falling below 2 Å, confirms the accuracy and reliability of our docking procedure. The docking scores and binding interactions of the best docked pose of each compound are represented in Table IV.1. Both Hit1 and Hit2 exhibit higher binding affinities compared to MRTX-1133, with Hit1 demonstrating the most favorable docking score of −13.881 kcal/mol.
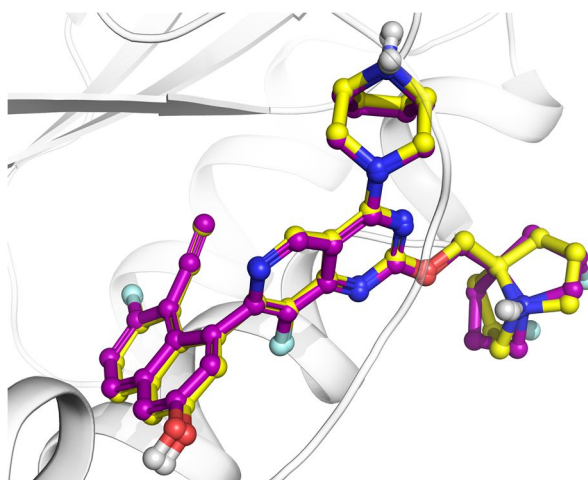


Figure IV.2: The alignment of the docked pose (magenta) of MRTX-1133 with its crystallographic conformation (yellow).

Table IV.1: Molecular docking results.

| Compounds | XP docking score (kcal/mol) | Hydrogen bond interactions | Hydrophobic interactions | Salt bridges |
|---|---|---|---|---|
| Hit1 | −13.881 | Hie95, Gly60, Thr58, Lys16, Asp12 | - | Asp12, Glu62 |
| Hit2 | −13.601 | Asp12, Lys16, Thr58, Ala59, Gly60, Arg68 | - | Asp12 |
| MRTX-1133 | −12.891 | Arg68, Asp69, Tyr96, Hie95 | Tyr64 | Asp12 |

MRTX-1133 establishes a complex network of interactions within the binding pocket of KRAS G12D, forming four hydrogen bonds with Arg68, Asp69, Tyr96, and Hie95 residues. Additionally, it engages in two π-π stacking interactions with the phenol ring of Tyr64 and forms a salt bridge with the carboxylate of Asp12.
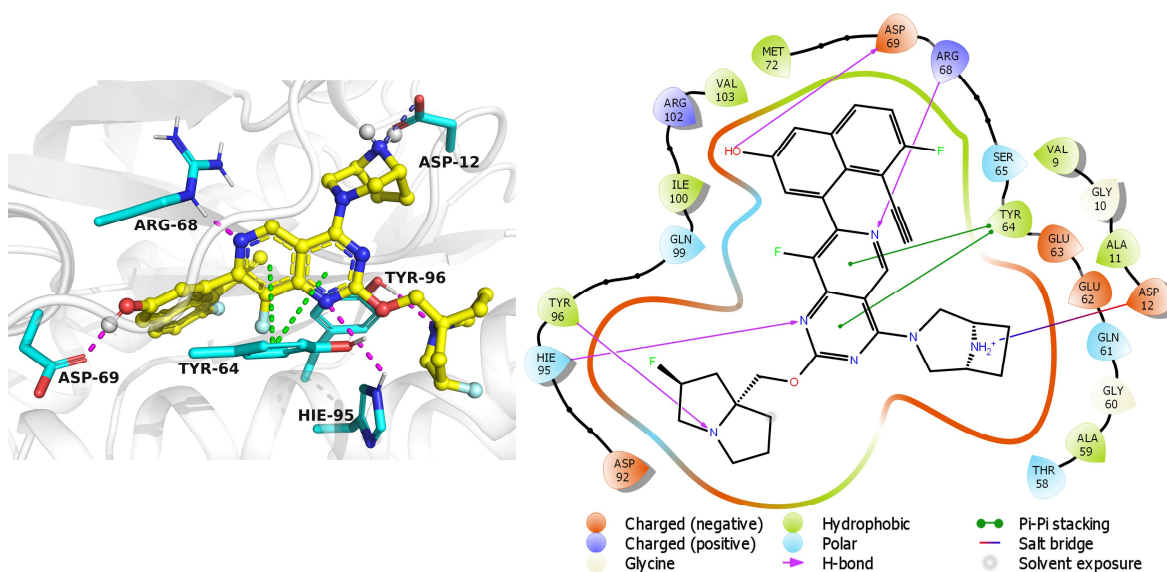


Figure IV.3: 2D and 3D ligand interaction diagrams of MRTX-1133 within the binding pocket of KRAS G12D.

For the novel compounds, hydrogen bonding interactions appear to be essential in stabilizing the ligand-protein complexes. In particular, Hit1 forms six hydrogen bonds with the Hie95, Gly60, Thr58, Lys16, and Asp12 residues. Likewise, Hit2 establishes six hydrogen bonds with the Asp12, Lys16, Thr58, Ala59, Arg68, and Gly60 residues.

Notably, the designed compounds effectively form a salt bridge interaction with Asp12, a feature previously recognized as crucial for the selective targeting of KRAS G12D and for

eliciting anti-tumor effects in preclinical pancreatic cancer models [21]. Additionally, Hit1 forms a second salt bridge interaction with Glu62, which could explain its improved binding affinity compared to MRTX-1133.
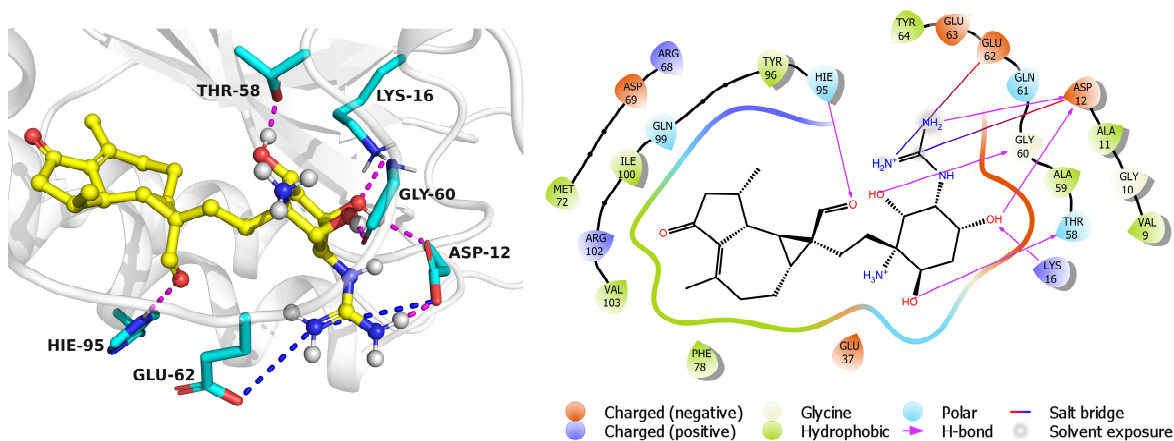


Figure IV.4: 2D and 3D ligand interaction diagrams of Hit1 within the binding pocket of KRAS G12D.
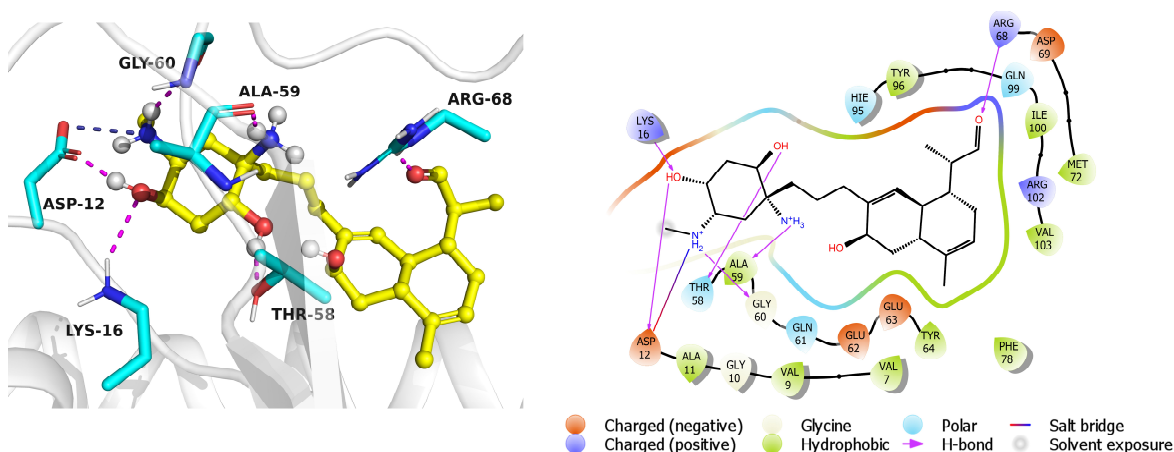


Figure IV.5: 2D and 3D ligand interaction diagrams of Hit2 within the binding pocket of KRAS G12D.

### IV.3.3 IFD and MM-GBSA analysis

Induced Fit Docking (IFD) allows for the flexibility of both the ligand and the receptor during the docking process, accommodating conformational changes and resulting in more accurate predictions of binding poses. This method helps to simulate a more realistic interaction scenario by considering the induced fit effects, where the binding site adapts to the ligand. The IFDScore indicates how well a ligand binds to its receptor, considering both

the ligand binding affinity and the induced fit effects. Lower (more negative) IFDScore values suggest better binding interactions. In this case, Hit1 and Hit2 show significantly lower IFDScore values (−7945.033 and −7991.433 kcal/mol, respectively) compared to MRTX-1133 (−6956.122 kcal/mol). This indicates that the novel compounds have stronger predicted binding interactions with the receptor, suggesting a better fit and more favorable binding poses.

The MM-GBSA results further support the IFDScore findings, where Hit1 and Hit2 have lower MM-GBSA scores (−87.766 and −89.150 kcal/mol, respectively) compared to MRTX-1133 (−72.723 kcal/mol). This suggests stronger and more stable binding interactions of these hits with KRAS G12D than those of MRTX-1133.

Table IV.2: IFD and MM-GBSA Scores.

| Compounds | IFDScore (kcal/mol) | MM-GBSA (kcal/mol) |
| --- | --- | --- |
| Hit1 | −7945.033 ± 40.258 | −87.766 ± 2.742 |
| Hit2 | −7991.433 ± 28.324 | −89.150 ± 4.978 |
| MRTX-1133 | −6956.122 ± 48.534 | −72.723 ± 4.765 |

The interaction fingerprints analysis shows that all compounds maintained their key interactions over several poses, with numerous novel contacts observed. Unlike MRTX-1133, the newly designed compounds demonstrated robust interactions with Lys16, Thr58, Gln61, Glu62, and Glu63 residues. This distinct binding pattern likely accounts for their high affinity towards the KRAS protein.

All ligands exhibited significant emerging interactions with Val9, Gly10, Arg68, Met72, Tyr96, Gln99, Ile100, and Val103. These residues are likely critical in enhancing the binding stability and specificity of the compounds. Identifying these interaction hotspots offers new insights into optimizing drug design, as targeted modifications to these residues could potentially improve binding affinity and specificity, leading to the development of more effective therapeutic agents.
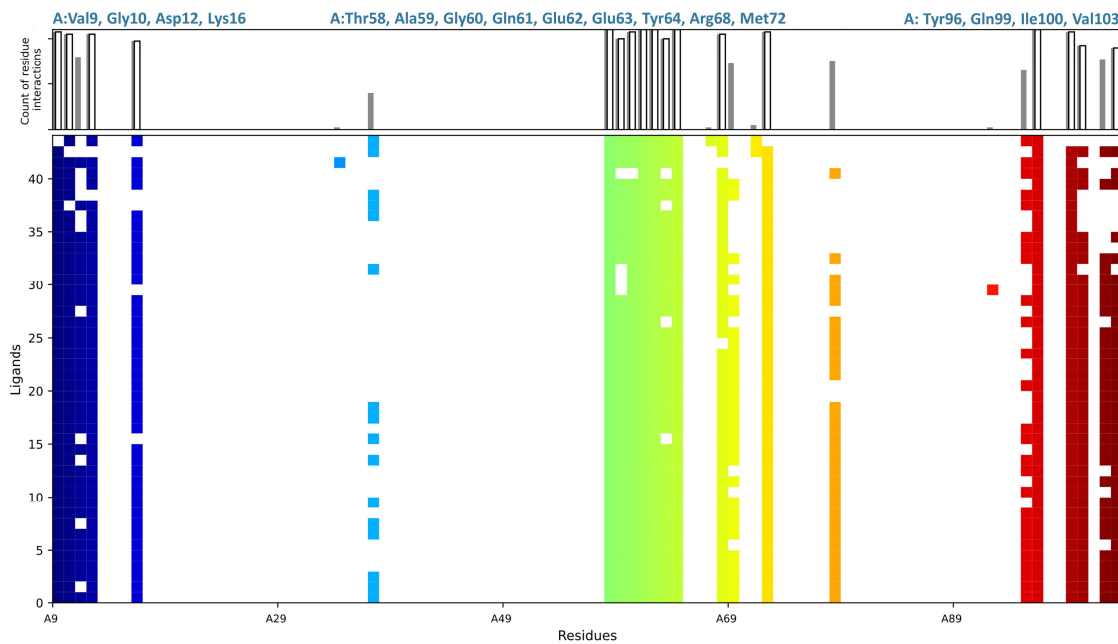
Figure IV.6: Structural interaction fingerprints of Hit1.
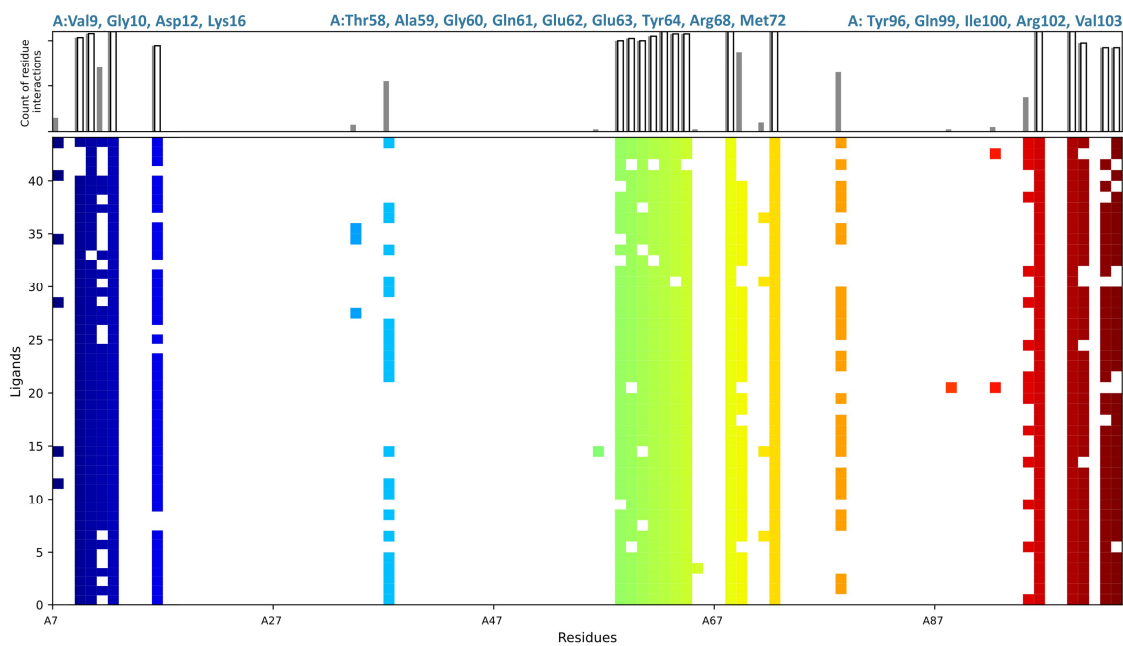


Figure IV.7: Structural interaction fingerprints of Hit2.
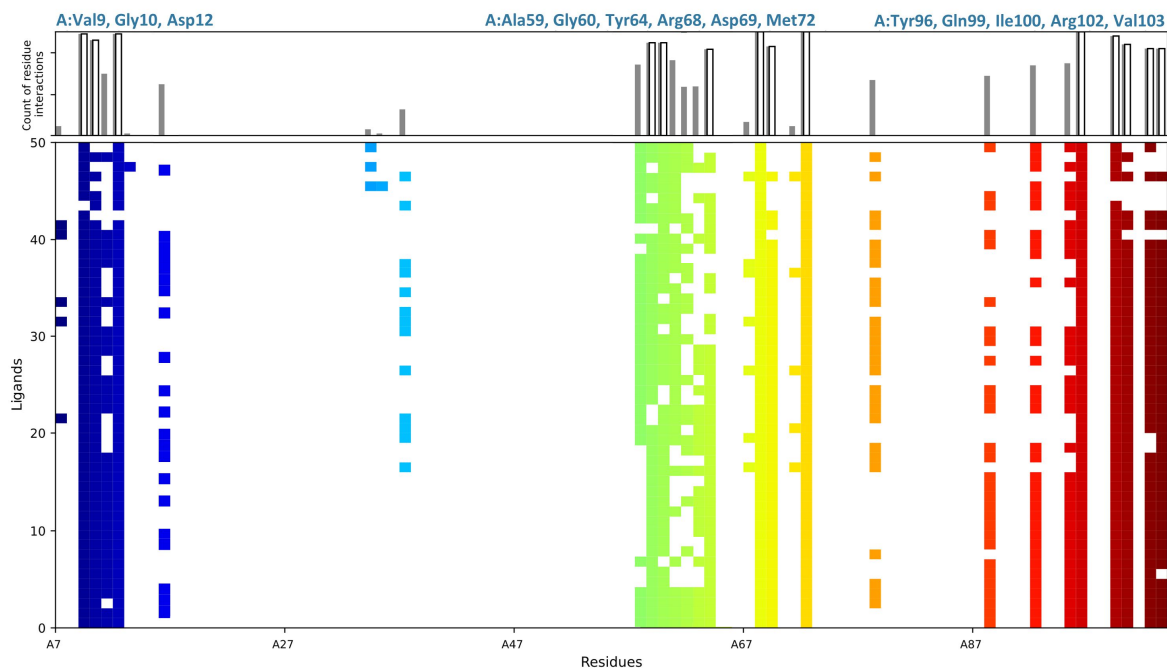
Figure IV.8: Structural interaction fingerprints of MRTX-1133.

## IV.3.4 ADMET analysis

To provide a comprehensive understanding of the ADMET properties of our compounds, we employed SwissADME, Protox 3.0, and Pred-hERG web servers. The results are presented in Table IV.3 and Table IV.4.

Hit1 has the highest TPSA (186.11 Å²), followed by Hit2 (122.01 Å²) and MRTX-1133 (92.42 Å²). Higher TPSA usually correlates with poor cell membrane permeability, which is reflected in Hit1's low GI absorption. Due to its high polarity and potentially lower oral bioavailability, Hit1 may be more suitable for alternative administration routes.

In terms of lipophilicity, Hit1 has the lowest logP value (0.94), indicating it is less lipophilic compared to Hit2 (3.17) and MRTX-1133 (4.62). While higher lipophilicity can enhance membrane permeability, it may also reduce water solubility, explaining why MRTX-1133 has poor solubility compared to the highly soluble Hit1 and Hit2.

None of the compounds are capable of permeating the blood-brain barrier (BBB), making them suitable for non-central nervous system (CNS) applications. However, Hit2 has been identified as a substrate for P-glycoprotein (P-gp), which may reduce its bioavailability by facilitating its active efflux from cells. This limitation can be addressed by co-administering P-gp inhibitors, such as verapamil or cyclosporin A, which could

effectively inhibit the efflux process and thereby increase the intracellular concentration of the compound, enhancing its therapeutic efficacy [22].

Table IV.3: List of the ADME properties of the studied compounds.

| Properties | | Hit1 | Hit2 | MRTX-1133 |
|---|---|---|---|---|
| **Physico-chemical properties** | MW (g/mol) | 450.57 | 422.60 | 602.65 |
| | Heavy atoms | 32 | 30 | 44 |
| | Aromatic heavy atoms | 0 | 0 | 20 |
| | Rotatable bonds | 6 | 7 | 5 |
| | H-bond acceptors | 5 | 4 | 8 |
| | H-bond donors | 7 | 5 | 3 |
| | TPSA ($Å^2$) | 186.11 | 122.01 | 92.42 |
| **Lipophilicity** | Log $P_{o/w}$ (ilogP ) | 0.94 | 3.17 | 4.62 |
| **Water solubility** | Log S (ESOL) | Very soluble | Very soluble | Poorly soluble |
| **Pharmacokinetics** | GI absorption | Low | High | High |
| | BBB permeant | No | No | No |
| | P-gp substrate | No | Yes | No |
| | CYP1A2 inhibitor | No | No | No |
| | CYP2C19 inhibitor | No | No | No |
| | CYP2C9 inhibitor | No | No | Yes |
| | CYP2D6 inhibitor | No | No | No |
| | CYP3A4 inhibitor | No | No | No |
| | Log $K_P$ (cm/s) | −10.44 | −8.95 | −6.32 |
| **Druglikeness** | Lipinski violations | 1 | 0 | 1 |
| | Ghose violations | 1 | 1 | 3 |
| | Veber violations | 1 | 0 | 0 |
| | Bioavailability Score | 0.55 | 0.55 | 0.55 |
| **Medicinal Chemistry** | Synthetic accessibility | 5.73 | 5.72 | 6.45 |

MW : Molecular Weight

TPSA : Topological Polar Surface Area

Log Po/w :  Predicted octanol/water partition coefficien

Log S :  Predicted aqueous solubility

BBB : Blood-Brain Barrier

P-gp : P-Glycoprotein

Kp: Skin Permeability Coefficient

Regarding cytochrome P450 enzyme inhibition, neither Hit1 nor Hit2 shows inhibitory activity against any of the major CYP enzymes. In contrast, MRTX-1133 inhibits CYP2C9, which may raise the risk of drug-drug interactions, potentially complicating its therapeutic use in combination with other medications metabolized by this enzyme. Skin permeability values show that Hit1 has the lowest permeability coefficient ($-10.44$ cm/s), indicating it is less likely to be absorbed through the skin compared to Hit2 ($-8.95$ cm/s) and MRTX-1133 ($-6.32$ cm/s).

In terms of drug-likeness, Hit1 violates one criterion each from the Lipinski, Ghose, and Veber rules, whereas MRTX-1133 fails to meet one Lipinski criterion and three of Ghose's rules. In contrast, Hit2 does not violate any of Lipinski's or Veber's rules but violates one of Ghose's criteria. This suggests that Hit2 is generally well within the acceptable range for drug-like properties, making it the most promising candidate among the three compounds.

All compounds have the same bioavailability score (0.55), indicating moderate bioavailability. Hit1 and Hit2 have similar synthetic accessibility scores (5.73 and 5.72, respectively), indicating moderate ease of synthesis. In contrast, MRTX-1133 has a higher score (6.45), suggesting it might be more challenging to synthesize due to its complexity.

Table IV.4: Comparative toxicity profiles of the studied compounds.

| Classification | Target | Hit1 | Hit2 | MRTX-1133 |
|---|---|---|---|---|
| Oral toxicity | $LD_{50}$ (mg/kg) | 1300 | 9000 | 2500 |
| | Toxicity class | 4 | 6 | 5 |
| Organ toxicity | Hepatotoxicity | Inactive | Inactive | Inactive |
| Toxicity endpoints | Carcinogenicity | Inactive | Inactive | Inactive |
| | Mutagenicity | Inactive | Inactive | Inactive |
| | Cytotoxicity | Inactive | Inactive | Inactive |
| | hERG blockage | Non-blocker | Non-blocker | Blocker |

$LD_{50}$ : Median Lethal Dose
hERG : Human Ether-à-go-go-Related Gene

Oral toxicity data show that Hit2 has the highest $LD_{50}$ (9000 mg/kg), indicating lower acute toxicity compared to Hit1 (1300 mg/kg) and MRTX-1133 (2500 mg/kg). None of the compounds show hepatotoxicity, and all are inactive regarding carcinogenicity,

mutagenicity, and cytotoxicity. However, MRTX-1133 is a blocker of the hERG channel, posing a significant risk of cardiotoxicity.

## IV.4    Conclusion

The findings presented in this chapter underscore the potential of fragment-based drug discovery as an effective approach for developing novel inhibitors targeting challenging oncogenic proteins such as KRAS G12D. Through a systematic process of fragment screening, linking, and optimization, two compounds were identified, both displaying higher binding affinities to KRAS G12D compared to the reference compound MRTX-1133.

The induced fit docking (IFD) results revealed that both hits exhibit significantly lower IFDScore values than MRTX-1133, indicating a more favorable fit within the KRAS G12D binding site. This enhanced interaction is further supported by the MM-GBSA results, which show that the novel compounds form more stable complexes with the target protein compared to the reference drug.

These compounds exhibit several significant advantages over MRTX-1133 in terms of ADMET properties, particularly regarding solubility, drug-likeness, and toxicity profiles. Hit2 emerges as the most promising candidate due to its high gastrointestinal absorption, adherence to drug-likeness rules, and lower acute toxicity, positioning it as a safer option for therapeutic application. While Hit1 exhibits lower GI absorption due to its high polarity, it remains a viable candidate for alternative routes of administration. Importantly, neither Hit1 nor Hit2 demonstrated significant toxicity concerns, such as hepatotoxicity, carcinogenicity, mutagenicity, or hERG channel inhibition, distinguishing them from MRTX-1133, which carries a higher risk of cardiotoxicity due to its hERG channel blocking potential.

Overall, the improved pharmacokinetic profiles, reduced toxicity, and enhanced binding interactions of our novel compounds suggest that they are promising alternatives to MRTX-1133, with Hit2 showing particular potential for further optimization and clinical development.

# References

[1] E.C. Nakajima, N. Drezner, X. Li, P.S. Mishra-Kalyani, Y. Liu, H. Zhao et al., *FDA Approval Summary: Sotorasib for KRAS G12C Mutated Metastatic NSCLC*, Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res. 28 (2022), pp. 1482–1486.

[2] S. Dhillon, *Adagrasib: First Approval*, Drugs 83 (2023), pp. 275–285.

[3] R.B. Kargbo, *Targeting KRASG12D Mutations: Discovery of Small Molecule Inhibitors for the Potential Treatment of Intractable Cancers*, ACS Med. Chem. Lett. 14 (2023), pp. 1041–1042.

[4] X. Wang, S. Allen, J.F. Blake, V. Bowcut, D.M. Briere, A. Calinisan et al., *Identification of MRTX1133, a Noncovalent, Potent, and Selective KRASG12D Inhibitor*, J. Med. Chem. 65 (2022), pp. 3123–3133.

[5] G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju and W. Sherman, *Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments*, J. Comput. Aided Mol. Des. 27 (2013), pp. 221–234.

[6] M.P. Jacobson, D.L. Pincus, C.S. Rapp, T.J.F. Day, B. Honig, D.E. Shaw et al., *A hierarchical approach to all-atom protein loop prediction*, Proteins 55 (2004), pp. 351–367.

[7] K. Roos, C. Wu, W. Damm, M. Reboul, J.M. Stevenson, C. Lu et al., *OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules*, J. Chem. Theory Comput. 15 (2019), pp. 1863–1874.

[8] B. Over, S. Wetzel, C. Grütter, Y. Nakai, S. Renner, D. Rauh et al., *Natural-product-derived fragments for fragment-based ligand discovery*, Nat. Chem. 5 (2013), pp. 21–28.

[9] T. Rodrigues, D. Reker, P. Schneider and G. Schneider, *Counting on natural products for drug design*, Nat. Chem. 8 (2016), pp. 531–541.

[10] K.B. Lokhande, P. Ghosh, S. Nagar and K. Venkateswara Swamy, *Novel B, C-ring truncated deguelin derivatives reveals as potential inhibitors of cyclin D1 and cyclin E using molecular docking and molecular dynamic simulation*, Mol. Divers. 26 (2022), pp. 2295–2309.

[11] P. Kirsch, A.M. Hartman, A.K.H. Hirsch and M. Empting, *Concepts and Core Principles of Fragment-Based Drug Design*, Molecules 24 (2019), pp. 4309.

[12] J.C. Shelley, A. Cholleti, L.L. Frye, J.R. Greenwood, M.R. Timlin and M. Uchimaya, *Epik: a software program for pK a prediction and protonation state generation for drug-like molecules*, J. Comput. Aided Mol. Des. 21 (2007), pp. 681–691.

[13] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz et al., *Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy*, J. Med. Chem. 47 (2004), pp. 1739–1749.

[14] D. Ramírez and J. Caballero, *Is It Reliable to Use Common Molecular Docking Methods for Comparing the Binding Affinities of Enantiomer Pairs for Their Protein Target?*, Int. J. Mol. Sci. 17 (2016), pp. 525.

[15] A.A. Alzain, F.A. Elbadwi and F.O. Alsamani, *Discovery of novel TMPRSS2 inhibitors for COVID-19 using in silico fragment-based drug design, molecular docking, molecular dynamics, and quantum mechanics studies*, Inform. Med. Unlocked 29 (2022), pp. 100870.

[16] D.M. Teli, B. Patel and M.T. Chhabria, *Fragment-based design of SARS-CoV-2 Mpro inhibitors*, Struct. Chem. 33 (2022), pp. 2155–2168.

[17] D.K. Lokwani, A.P. Sarkate, K.S. Karnik, A.P.G. Nikalje and J.A. Seijas, *Structure-Based Site of Metabolism (SOM) Prediction of Ligand for CYP3A4 Enzyme: Comparison of Glide XP and Induced Fit Docking (IFD)*, Molecules 25 (2020), pp. 1622.

[18] A. Daina, O. Michielin and V. Zoete, *SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules*, Sci. Rep. 7 (2017), pp. 42717.

[19] P. Banerjee, E. Kemmler, M. Dunkel and R. Preissner, *ProTox 3.0: a webserver for the prediction of toxicity of chemicals*, Nucleic Acids Res. 52 (2024), pp. W513–W520.

[20] R.C. Braga, V.M. Alves, M.F.B. Silva, E. Muratov, D. Fourches, L.M. Lião et al., *Pred-hERG: A Novel web-Accessible Computational Tool for Predicting Cardiac Toxicity*, Mol. Inform. 34 (2015), pp. 698–701.

[21] Z. Mao, H. Xiao, P. Shen, Y. Yang, J. Xue, Y. Yang et al., *KRAS(G12D) can be targeted by potent inhibitors via formation of salt bridge*, Cell Discov. 8 (2022), pp. 1–14.

[22] Md.L. Amin, *P-glycoprotein Inhibition for Optimal Drug Delivery*, Drug Target Insights 7 (2013), pp. 27–34.

# General Conclusion

This thesis addresses the challenge of pancreatic cancer by applying computational strategies to discover novel therapeutic candidates, focusing on both direct and indirect KRAS inhibition. By integrating the chemical diversity of natural products with advanced in silico techniques, this work uncovers promising avenues for the development of effective treatments.

In the first study, the objective was to identify curcumin derivatives as potential inhibitors of the PI3Kα isoform. To achieve this, a ligand-based pharmacophore model was developed to capture the essential features for biological activity. The AAHRR_1 hypothesis emerged as the best promising model, characterized by two aromatic rings, two hydrogen bond acceptors, and one hydrophobic group. This pharmacophore was used to construct an atom-based 3D-QSAR model, which demonstrated high predictive accuracy with an $R^2$ value of 0.990 and a $Q^2$ of 0.971. The 3D-QSAR model was thoroughly verified by external validation and Y-randomization tests, confirming its robustness and reliability. Contour map analysis further highlighted key factors for biological activity, such as hydrophobic interactions and electron-withdrawing properties, providing valuable insights for future drug development. Following an enrichment analysis which confirmed the AAHRR_1 pharmacophore's ability to discriminate between active and inactive compounds, the model was used to screen a dataset of 1,208 curcumin derivatives, identifying 1,065 compounds that aligned with at least four out of the five pharmacophore features. Further refinement through structure-based virtual screening, employing a multi-step molecular docking protocol, narrowed this pool to seven promising hits with docking scores ranging from −10.926 to −9.882 kcal/mol. ADMET analysis highlighted two of these compounds, CID154728220 and CID156189304, for their favorable pharmacokinetic properties, and both were chosen for molecular dynamics (MD) simulations to further investigate their structural stability. The MD simulations confirmed that CID154728220 and CID156189304 were the most promising candidates, demonstrating structural stability and strong binding interactions with PI3Kα. These findings highlight the potential of these compounds as novel PI3Kα inhibitors for anti-pancreatic cancer therapies.

The second study aimed to directly target the KRAS G12D mutant using a fragment-based drug discovery approach. Over thirty thousand natural and natural product-like fragments were screened, resulting in the identification of 4,529 fragments with docking scores ≤−7.000 kcal/mol. These fragments were subjected to a rigorous linking process, ultimately leading to the generation of 100 novel compounds. Among these, two promising hits emerged, exhibiting higher binding affinities compared to the reference compound MRTX-1133, with docking scores of −13.881 and −13.601 kcal/mol. Validation through Induced Fit Docking (IFD) and MM-GBSA calculations further confirmed their robust binding interactions and greater stability with KRAS G12D. Interaction fingerprints analysis revealed that, unlike MRTX-1133, Hit1 and Hit2 consistently maintained interactions with residues such as Lys16, Thr58, Gln61, Glu62, and Glu63 across multiple poses. These stable interactions likely contribute to their high affinity for KRAS G12D. Additionally, novel interactions with residues such as Val9, Gly10, Arg68, Met72, Tyr96, Gln99, Ile100, and Val103 were identified, suggesting that these hotspots may enhance binding stability and specificity. Future studies could focus on optimizing these interactions to further improve therapeutic efficacy. In the ADMET-Tox analysis, Hit2 presented the best balance of pharmacokinetic properties, safety, and drug-likeness, making it the most promising candidate for further development as a targeted therapy for pancreatic cancer.

Future studies should focus on the experimental validation of these compounds to confirm their therapeutic potential. Additionally, exploring other KRAS-mediated signaling pathways, such as RAF/MEK/ERK or RalGDS, could provide new targets for overcoming KRAS-driven oncogenesis in pancreatic cancer, offering a broader scope for therapeutic development.

# Appendix

Table 1: Intersite distances between the pharmacophoric sites of AAHRR_1.

| Site 1 | Site 2 | Distance |
|--------|--------|----------|
| A4 | A7 | 2.72 |
| A4 | H10 | 14.79 |
| A4 | R14 | 12.55 |
| A4 | R15 | 2.80 |
| A7 | H10 | 15.44 |
| A7 | R14 | 12.67 |
| A7 | R15 | 2.77 |
| H10 | R14 | 3.72 |
| R15 | H10 | 12.72 |
| R15 | R14 | 10.11 |

Table 2: Intersite angles between the pharmacophoric sites of AAHRR_1.

| Atom 1 | Atom 2 | Atom 3 | Angle |
|--------|--------|--------|-------|
| A7 | A4 | H10 | 98.8 |
| A7 | A4 | R14 | 86.3 |
| A7 | A4 | R15 | 60.2 |
| H10 | A4 | R14 | 12.5 |
| H10 | A4 | R15 | 38.6 |
| R14 | A4 | R15 | 26.1 |
| A4 | A7 | H10 | 71.1 |
| A4 | A7 | R14 | 81.3 |
| A4 | A7 | R15 | 61.3 |
| H10 | A7 | R14 | 10.2 |
| H10 | A7 | R15 | 9.8 |
| R14 | A7 | R15 | 20.0 |
| A4 | H10 | A7 | 10.0 |
| A4 | H10 | R14 | 46.9 |
| A4 | H10 | R15 | 7.9 |
| A7 | H10 | R14 | 36.9 |
| A7 | H10 | R15 | 2.1 |
| R14 | H10 | R15 | 39.0 |
| A4 | R14 | A7 | 12.4 |
| A4 | R14 | H10 | 120.6 |
| A4 | R14 | R15 | 7.0 |
| A7 | R14 | H10 | 132.9 |
| A7 | R14 | R15 | 5.4 |
| H10 | R14 | R15 | 127.6 |
| A4 | R15 | A7 | 58.5 |

Table 2 (Continued)

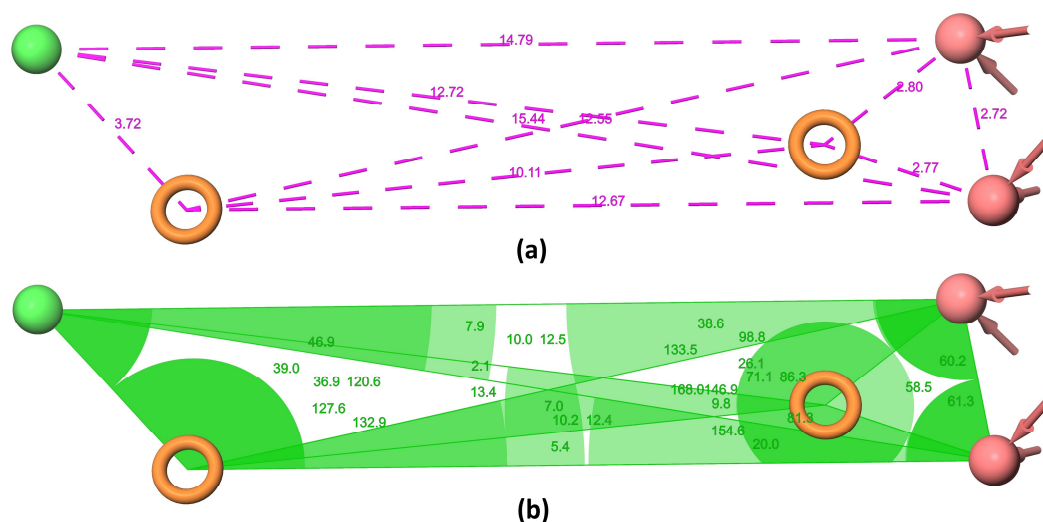| Atom1 | Atom2 | Atom3 | Angle |
|-------|-------|-------|-------|
| A4 | R15 | H10 | 133.5 |
| A4 | R15 | R14 | 146.9 |
| A7 | R15 | H10 | 168.0 |
| A7 | R15 | R14 | 154.6 |
| H10 | R15 | R14 | 13.4 |



Figure 1: The intersite distances (a) and angles (b) among the pharmacophoric points of AAHRR_1.

Table 3: IFDScore and MM-GBSA results for Hit1, Hit2, and MRTX-1133.

| Hit1 | | Hit2 | | MRTX-1133 | |
|------|------|------|------|-----------|------|
| **IFDScore** | **MM–GBSA** | **IFDScore** | **MM–GBSA** | **IFDScore** | **MM–GBSA** |
| −7999.17 | −87.648 | −7999.17 | −89.352 | −6999.17 | −74.469 |
| −7995.07 | −85.066 | −7995.07 | −84.338 | −6995.07 | −76.578 |
| −7992.76 | −87.667 | −7992.76 | −83.177 | −6992.76 | −76.273 |
| −7992.15 | −85.055 | −7992.15 | −91.909 | −6992.15 | −76.282 |
| −7989.84 | −85.126 | −7989.84 | −83.215 | −6989.84 | −77.477 |
| −7988.15 | −82.394 | −7988.15 | −83.955 | −6988.15 | −69.547 |
| −7984.99 | −87.409 | −7984.99 | −83.562 | −6984.99 | −70.711 |
| −7983.65 | −86.591 | −7983.65 | −86.074 | −6983.65 | −70.006 |
| −7981.33 | −86.643 | −7981.33 | −85.402 | −6981.33 | −69.943 |
| −7981.13 | −88.191 | −7981.13 | −85.132 | −6981.13 | −70.263 |
| −7978.54 | −85.867 | −7978.54 | −88.289 | −6978.54 | −69.898 |
| −7977.83 | −85.880 | −7977.83 | −86.279 | −6977.83 | −70.181 |
| −7977.51 | −85.836 | −7977.51 | −85.104 | −6977.51 | −70.608 |

Table 3 (Continued)

| Hit1 | | Hit2 | | MRTX-1133 | |
|---|---|---|---|---|---|
| **IFDScore** | **MM-GBSA** | **IFDScore** | **MM-GBSA** | **IFDScore** | **MM-GBSA** |
| −7976.76 | −84.740 | −7976.76 | −85.152 | −6976.76 | −70.508 |
| −7970.22 | −84.778 | −7970.22 | −89.701 | −6970.22 | −73.231 |
| −7967.4 | −84.704 | −7967.4 | −88.795 | −6967.4 | −74.129 |
| −7960.52 | −87.818 | −7960.52 | −85.172 | −6960.52 | −79.535 |
| −7959.39 | −85.223 | −7959.39 | −85.817 | −6959.39 | −75.848 |
| −7957.23 | −84.842 | −7957.23 | −87.829 | −6957.23 | −79.770 |
| −7956.51 | −86.506 | −7956.51 | −88.911 | −6956.51 | −75.914 |
| −7954.28 | −84.881 | −7954.28 | −86.910 | −6954.28 | −78.697 |
| −7952.66 | −86.793 | −7952.66 | −88.040 | −6952.66 | −76.848 |
| −7951.15 | −84.328 | −7951.15 | −92.832 | −6951.15 | −77.649 |
| −7949.7 | −88.106 | −7949.7 | −92.061 | −6949.7 | −75.872 |
| −7949.06 | −90.041 | −7949.06 | −92.721 | −6949.06 | −71.362 |
| −7948.71 | −89.142 | −7948.71 | −93.180 | −6948.71 | −77.290 |
| −7942.59 | −87.564 | −8053.37 | −91.355 | −6942.59 | −79.200 |
| −7942.03 | −95.566 | −8043.57 | −97.819 | −6942.03 | −71.920 |
| −7941.24 | −90.785 | −8042.86 | −96.282 | −6941.24 | −78.333 |
| −7938.56 | −92.739 | −8035.04 | −94.773 | −6938.56 | −71.103 |
| −7931.73 | −90.377 | −8025.47 | −96.009 | −6931.73 | −71.265 |
| −7931.71 | −90.818 | −8024.34 | −96.292 | −6931.71 | −77.339 |
| −7919.6 | −94.335 | −8021.42 | −91.396 | −6919.6 | −70.701 |
| −7910.64 | −88.440 | −8019.58 | −88.227 | −6910.64 | −73.670 |
| −7910.43 | −90.685 | −8017 | −100.303 | −6910.43 | −71.865 |
| −7910.14 | −89.836 | −8014.92 | −87.055 | −6910.14 | −77.639 |
| −7902.01 | −90.839 | −8012.74 | −93.238 | −6902.01 | −72.511 |
| −7898.11 | −87.043 | −8008.85 | −89.277 | −6898.11 | −79.342 |
| −7895.78 | −88.753 | −8007.9 | −86.165 | −6895.78 | −77.286 |
| −7890.34 | −87.485 | −8005.01 | −97.033 | −6890.34 | −68.121 |
| −7881.18 | −89.454 | −8004.93 | −98.475 | −6881.18 | −63.239 |
| −7865 | −87.548 | −8004.28 | −79.277 | −6865 | −71.384 |
| −7854.86 | −88.117 | −8003.31 | −84.730 | −6854.86 | −64.463 |
| −7839.79 | −90.066 | −8002.74 | −82.002 | −6839.79 | −64.190 |
| - | - | - | - | −7053.37 | −66.669 |
| - | - | - | - | −7043.57 | −76.849 |
| - | - | - | - | −7042.86 | −61.907 |
| - | - | - | - | −7035.04 | −68.393 |
| - | - | - | - | −7025.47 | −65.109 |
| - | - | - | - | −7024.34 | −64.750 |