Computational approaches to identify potential inhibitors of fibroblast activation protein for cancer therapeutics

Thesis submitted in partial fulfilment of the requirement for the degree

Of

Master of Science degree in Biotechnology

By

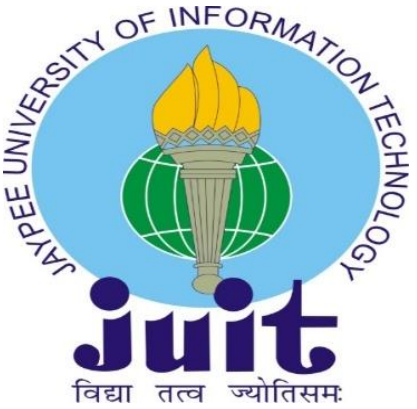
Aman Parashar

Roll no. 217828

Under the Supervision Of

Dr. Raj Kumar

Assistant Professor



Department of Biotechnology & Bioinformatics

Jaypee University of Information Technology Waknaghat, Solan 173234

Himachal Pradesh

SELF - DECLARATION

I, Aman Parashar, student of M.Sc. Biotechnology, Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh do declare that work reported in M.Sc. thesis entitled "Computational approaches to identify potential inhibitors of fibroblast activation protein for cancer therapeutics" submitted in the Department of Biotechnology & Bioinformatics, Jaypee University of Information Technology, Waknaghat is an authentic record of my work carried out over a period from June 2022 to May 2023 under the supervision of Dr. Raj Kumar Assistant Professor, Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Solan, Himachal Pradesh. I also authenticate that I have carried out the above mentioned project work under the proficiency stream. The matter embodied in the report has not been submitted for the award of any other degree or diploma.

Aman Parashar, 217828

Date:

Department of Biotechnology and Bioinformatics,

Jaypee University of Information Technology,

Place: Solan, Himachal Pradesh

CERTIFICATE

This is to certify that the work reported in the M.Sc. Biotechnology thesis entitled "Computational approaches to identify potential inhibitors of fibroblast activation protein for cancer therapeutics", submitted by Mr. Aman Parashar (217828) at Jaypee University of Information Technology, Waknaghat, India, is a bonafide record of her original work carried out from July 2022 to May 2023 under my supervision. This work has not been submitted elsewhere for any other degree or diploma.

Dr. Raj Kumar

Assistant Professor

Date:

Department of Biotechnology and Bioinformatics,

Jaypee University of Information Technology,

Place: Solan, Himachal Pradesh

ACKNOWLEDGMENT

First and foremost, I would like to express my sincere gratitude to the Department of Biotechnology and Bioinformatics, JUIT for giving me consistent support and guidance and support throughout the project process.

With proud privilege and a profound sense of gratitude, I acknowledge my indebtedness to my guide Dr. Raj Kumar, Assistant Professor, at the Jaypee University of Information Technology for his valuable guidance, suggestions, constant encouragement, and cooperation. I am been extremely fortunate to have a guide, who consistently provided me with the liberty to explore my potential and for providing me such a good lab environment.

Aman Parashar

217828

TABLE OF CONTENTS

List of abbreviations

CAF	Cancer Associated Fibroblast
FAP	Fibroblast Activation Protein
CRC	Colorectal Cancer
PRC	Pancreatic Cancer
VS	Virtual Screening
RMSD	Root Mean Square Deviation
SBDD	Structure Based Drug Designing
LBDD	Ligand Based Drug Designing
CADD	Computer Aided Drug Designing
DPP	Dipeptidyl Peptidase
MDS	Molecular Dynamics Simulation
DS	Discovery Studio
SI	Selectivity index
WHO	World Health Organization
GLOBOCAN	Global Cancer Observatory

List of Figures

List of Tables

Abstract

Cancer is the leading cause of mortality worldwide, accounting for nearly 10 million deaths in 2020, according to the World Health Organisation. Breast cancer, lung cancer, colon and rectal cancer, prostate cancer, non-melanoma skin cancer, and stomach cancer were the most prevalent new cancer cases in 2020. Many target cells, receptors, enzymes, and molecular markers are being identified for pharmacological targets that can suppress tumour growth, and one of them is fibroblast activation protein (FAP), which is overexpressed in many forms of cancer and hence can be an effective drug target. FAP has its resemblance to other family member proteins such as DPPIV and PREP. Therefore, it is challenging to find inhibitors for FAP in order to lessen the unwanted side effects. This study is primarily concerned with identifying inhibitors that inhibit FAP. This is accomplished through the use of computational approaches such as molecular docking and virtual screening. First, the accuracy of LibDock program is validated and then the compounds with the low IC50 value are selected and docked, which are then used as reference compounds. By virtual screening the compounds are screened out based on the LibDock Score and they are further analysis based on their pose, orientation and molecular interaction. The study resulted in identifying 3 hits compounds, namely, MSID001315, MSID001017 and MSID001261 and therefore can be utilized as potential inhibitor compound for targeting FAP. For further studies MDS and *in-vitro* testing can be performed on these molecules.

Keywords:

Fibroblast activation protein (FAP), Virtual screening, Molecular dynamics simulation, Cancer, Molecular docking

CHAPTER 1: INTRODUCTION

Introduction

Cancer is one of the most widespread causes of fatalities worldwide, yet its impact differs between developed and developing nations. This may be due to numerous causes involved, such as the ageing and expanding population, the increasing socioeconomic development, and many others [1]. For instance, males with colorectal and prostate cancer are more likely to live in locations with high Human Development Index (HDI) than women with cervical cancer [2]. When treating cancer, there are various challenges to overcome, such as targeting cancer stem cells (CSCs), cancer epigenetic profiling is lacking, difficulties with detecting type of cancer, traditional chemotherapeutic agents' constraints, and lack of available efficient biomarkers for determining cancer prognosis and diagnosis [3]. Therefore, WHO has recommended developing a comprehensive strategy for treating cancer. Scientists have investigated a variety of selective and efficient biomarkers in this area. For example, the most popular tumour marker for HCC is plasma alpha fetoprotein (AFP) [4], while the different serum markers (proPSA, tPSA, PSADT, EPCA, and EPCA-2, PSAV), tissue markers (AMACR, methylation GSTP1), and urine marker (DD3PCA3/UPM-3) are the biomarkers for prostate cancer [5].

FAP is well recognised for being present in stromal fibroblasts, epithelial tumours, including pancreatic adenocarcinoma, oesophageal cancer, breast cancer, and oesophageal cancer [6]. FAP, previously termed as F19 cell surface antigen, is a glycoprotein discovered in cultured fibroblasts in 1986 using the monoclonal antibody (mAb) F19 [7]. The F19 cell surface antigen was termed the FAP in 1994. Generally, the expression of FAP is not observed in tissue of resting fibroblasts, but it may be forced to produce in non-transformed, active stromal fibroblasts. Endometrial cells are generally the only cells that contain FAP protein. FAP is also expressed during wound healing, cirrhosis, rheumatoid arthritis, osteoarthritis, and pulmonary fibrosis, all of which are linked with active stroma. Depending on the body's location, developmental conditions including embryogenesis, tissue maintenance, wound healing, and cancer, fibroblasts can be triggered by a variety of chemical signals. For instance, NHDFs and HPMFs gets activated by the MDA-MB-231 breast cancer cell line, causing them to develop into CAFs with altered shape, gene expression, and motility most notably the ability to speed up breast tumour growth [8].

FAP plays an important role in tumorigenic metastasis and invasion by contributing in angiogenesis, antitumor immunological dysregulation, and ECM (Extracellular Matrix) synthesis, deposition, and remodelling. Therefore, has the potential to be a new candidate for active cancer treatment, especially when paired with chemotherapy. Finding an

effective FAP inhibitor through the traditional drug development approach, on the other hand, is difficult due to structural similarities with DPPIV, PREP, and other family members, and it will also take a significant amount of time and money [9]. There are currently no FDA-approved drugs on the market that target the FAP receptor.

Computational techniques to drug development have gained substantial traction in recent years, providing cost-effective and time-efficient strategies for discovering prospective drug candidates. Molecular docking and virtual screening techniques are used in this study because they are useful tools for selecting hit molecules that can efficiently bind with the FAP active site and limit its enzymatic activity. It is feasible to interrupt tumour growth by inhibiting the FAP receptor, hence decreasing tumour progression. As a result, it may be useful in the treatment of cancer and may boost efficacy when combined with established anticancer therapies. The molecular docking technique will aid in the prediction of ligand binding interactions within the active site. On the other hand, a huge number of chemicals with desirable binding features to the target can be filtered out utilizing the virtual screening technique. This study has uses these computational tools to prioritise and choose hit compounds which may have the potential to disrupt the signalling pathways involved in disease progression.

CHAPTER 2: REVIEW OF LITERATURE

2.1 FAP in different cancers

FAP is considered as a crucial indicator of cancer associated fibroblast (CAF), and it might influence the extracellular matrix's composition and structure to help explain some of the CAFs tumour-promoting behaviours. In pancreatic adenocarcinoma, oesophageal cancer, and breast cancer, the expression of FAP is also observed in cancer cells, increasing their proliferation and tumorigenicity [10]. Healthy adult tissue does not express FAP, with the exception of a subpopulation of pancreatic islet alpha cells that produce glucagon in the course of tissue remodelling as well as in foetal mesenchymal tissue in the time of embryogenesis. Stromal fibroblasts are detected in over 90% of epithelial malignancies. However, it has been demonstrated in other studies that the expression of FAP might also be observed in other cell types outside of stromal fibroblasts, including some epithelial malignant cells [6]. It is discovered to be present in a wide variety of epithelial malignancies, including those of the colon, ovarian, lung, breast, bladder, and pancreatic cancer [7]. In human serum, a soluble variant of fibroblast activation protein without the transmembrane domain was recently track down as an antiplasmin-cleaving enzyme (APCE), which cleaves the Pro12-Asn13 link of Met-a2AP, converting it to the more active form Asn-a2AP and suppressing fibrinolysis.

Because fibroblasts are the major site of FAP expression, it is hoped that a better knowledge of the function of CAFs would lead to novel therapies to improve outcomes in the many malignancies in which FAP is overexpressed [11].

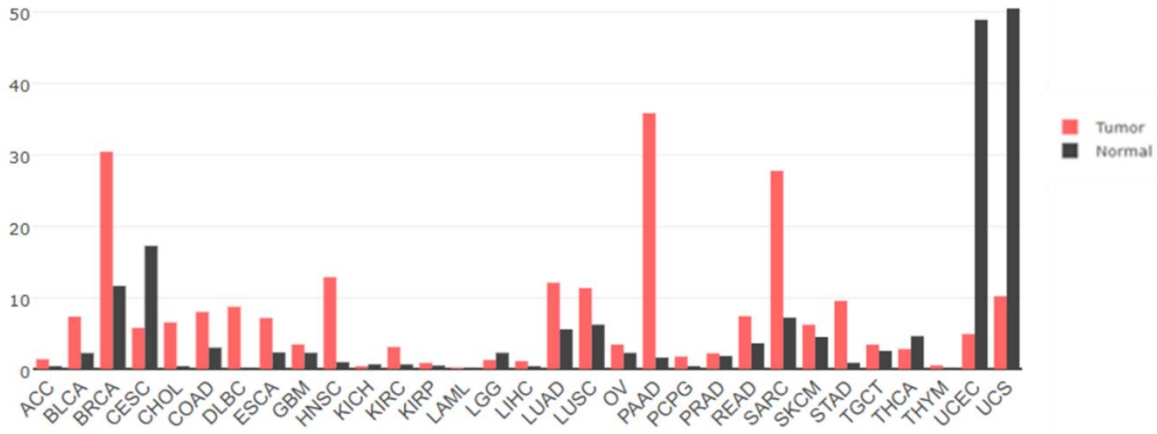


Figure 1: Expression of FAP in different types of cancer ***Source – GEPIA***

ACC: Adrenocortical carcinoma; BLCA: Bladder Urothelial Carcinoma: BRCA: Breast invasive carcinoma; CESC: Cervical squamous cell carcinoma and endocervical adenocarcinoma: CHOL: Cholangio carcinoma: COAD: Colon adenocarcinoma; DLBC: Lymphoid Neoplasm Diffuse Large B-cell Lymphoma: ESCA: Esophageal carcinoma: GBM: Glioblastoma multiforme: HNSC: Head and Neck squamous cell carcinoma; KICH: Kidney Chromophobe; KIRC: Kidney renal clear cell carcinoma; KIRP: Kidney renal papillary cell carcinoma: LAML: Acute Myeloid Leukemia; LGG: Brain Lower Grade Glioma; LIHC: Liver hepatocellular carcinoma; LUAD: Lung adenocarcinoma; LUSC: Lung squamous cell carcinoma; MESO: Mesothelioma; OV: Ovarian serous cystadenocarcinoma; PAAD: Pancreatic adenocarcinoma; PCPG: Pheochromocytoma and Paraganglioma; PRAD: Prostate adenocarcinoma; READ: Rectum adenocarcinoma; SARC: Sarcoma; SKCM: Skin Cutaneous Melanoma; STAD: Stomach adenocarcinoma; TGCT: Testicular Germ Cell Tumors; THCA: Thyroid carcinoma; THYM: Thymoma; UCEC: Uterine Corpus Endometrial Carcinoma; UCS: Uterine Carcinosarcoma; UVM: Uveal Melanoma

2.1.1 Breast cancer

Breast cancer develops over a period of time and involves a variety of cell types; hence, finding effective treatments for it is a global problem [12]. According to the WHO 2020 report, 6,85,000 people died worldwide and 2.3 million women were diagnosed with breast cancer. Numerous factors are responsible or can increase the risk of developing breast cancer, including sex, family history, gene mutations, ageing, estrogen, and unhealthy lifestyle. BRCA1/2 [13], HER2, Epidermal Growth Factor Receptor (EGFR) [14], c-Myc [15], and Ras [16] are genes that have been linked to breast cancer.

The contribution of FAP to breast cancer has been identified in numerous research. For instance, the F19 monoclonal antibody allowed for the detection of FAP in the reactive stroma of breast cancer [7]. FAP- increases proliferation likely via controlling the FAK pathway, and breast cancer patients with poor prognoses and low patient survival had higher levels of FAP-expression [6]. When NHDF was activated by MDA-MB-231 conditioned media, five genes—PDPN, FAP, ACTA, DDR2, and PDGFRB—were expressed more than two times as much, and PDPN and FAP were considerably elevated, leading to fibroblast activation [8].

2.1.2 Gastric cancer

Gastric cancer is the third most frequent cancer that causes death globally, according to GLOBOCAN 2018 statistics. Every year, more than a million new instances of stomach cancer are discovered worldwide. Males had a higher rate of gastric cancer mortality. There are many risk factors of developing gastric cancer, including the loss of one copy of the CDH1 gene [17], *Helicobacter pylori* infection [18], smoking [19], and alcohol consumption. According to a study by Okada K *et al.* [20], FAP-immuno-reactivity was identified in diffuse types and intestinal types of gastric cancer when detected by immunoblotting, together with varying degrees of protein expression [6]. Additionally, the Wnt/-catenin signalling pathway has been associated with the progression of gastric cancer by FAP-derived from CAFs, which starts EMT [21].

2.1.3 Colorectal cancer

As per WHO in 2020, colorectal is the third most frequent cancer in the world with around 2 million cases. Numerous risk factors, such as those that are sporadic (which account for 70% of CRC cases), familial (which account for 25% of CRC cases), and genetic (which account for 5% of CRC cases), have an impact on the extremely complex and varied aetiology of CRC [22]. There are three main, separate genetic routes that have been proposed for CRC. Chromosomal rearrangements, aneuploidy karyotype and the frequent loss-of-heterozygosity (LOH) at tumour suppressor gene loci are the first 3 effects of the CIN (Chromosomal instability) pathway. The second pathway is the MSI (Microsatellite instability) pathway, which is brought on by DNA MMR gene failure, and the third pathway is the CIMP (CpG island methylator phenotype) pathway, which is characterised by a high level of CpG island methylation [23].

FAP expression in the stroma of human colorectal tumours and other epithelial malignancies has been proven. FAP increased the biological phenotype of aggressiveness in human colorectal cancer. Because of the inhibition of enzymatic activity of FAP in these tumours the cancer growth is reduced to a great extent, and it has also been hypothesised that FAP activity aids in the formation of human colorectal tumour xenografts. Higher levels of FAP and a lower overall survival rate are observed in patients with established metastatic illness [24]. The study by Knuchel *et al.* is particularly relevant to the findings here because it demonstrated that

adhesion is due to FGF-2, FGFR, and integrin ßvß5, and that fibroblasts promote the cell elongation and motility through direct cell-cell interactions in CRC [8].

2.1.4 Pancreatic cancer

With over 459 000 new cases and 432000 fatalities, PRC ranks as the seventh most prevalent cancer death cause for both men and women globally, according to GLOBOCAN 2018 figures. Obesity, type II diabetes, and smoking are risk factors for the development of PRC [25]. Other risk factors include the Peutz-Jeghers syndrome (which shoot up the risk of developing PRC by 35%), germline mutations in CDKN2A (which shoot up the risk of developing PRC by 17%), and genes essential for DNA repair. Additionally, people with hereditary pancreatitis syndromes (mutations in SPINK1 and PRSS1), have a 40% overall risk of developing PRC [26]. TP53 [27], SMAD4 [28], and BRAF [29] are a few of the genes in charge of causing PRC. The mitogen-activated protein kinase (MAPK) pathway is activated by a mutation in either of these genes, which causes pancreatic cancer [30]. Some of the pathways that are dysregulated in PRC include phosphatidylinositol-4, 5-bisphosphate 3-kinase (PI3K), the epidermal growth factor receptor (EGFR) pathway, and the hedgehog (HH) route [30]. The serum biomarker CA19-9, which has been linked to pancreatic cancer and is well-established and validated, has a sensitivity of 79–81% and a specificity of 82–90% [102].

According to numerous studies, the FAP, which is highly expressed in CAFs of epithelial malignancies, is a drug target and is linked to desmoplasia and a poor prognosis [31]-[34]. PATSOURAS *et al.*,[33], Lo *et al*.,[35] and Kawase *et al*.,[36] by utilising different approaches, have discovered the FAP expression in both stromal and tumorous cells in PDAC. Kawase *et al.* [36] has identified the two mechanism for the higher expression of FAP in PDAC. In the first mechanism, FAP-expressing fibroblasts increased the cell invasive capacity brought on by APCE, a soluble version of FAP, via altering the different components (glycoprotein, collagen, etc.) of extracellular matrix (ECM), even if the cells are not in direct contact. In a different mechanism, FAP-expressing fibroblasts encourage Rb phosphorylation, which in turn triggers cell cycle activation in pancreatic cancer cells.

2.1.5 Oral cancer

A malignant neoplasia that develops in the oral cavity is oral cancer, often known as squamous cell carcinoma (OSCC). In comparison to women, men are two to three times more likely to acquire mouth cancer [37]. Two key risk factors, smoking [38][39] and alcohol use [40], are present in 90% of oral cancer patients. Oral cancer can also be caused by other risk factors such the human papillomavirus [40] and ultraviolet (UV) exposure. There are number of biomarkers for oral cancer detection like IL-8 [41] [42], TNF-α and IL-1 [42], Cancer antigen 125 (CA125) [43], CD59 and Profilin [44], Haptoglobin and Complement C3 [45].

In OSCC, the FAP expression was observed in cancerous cells which was identified using immunofluorescence and western blotting technique. FAP regulates OSCC cell proliferation, invasion and migration via the Ras-ERK and PTEN/PI3K/Akt signalling pathways [46]. The FAP was further seen to be elevated by a process known as nemosis, and because its expression is constant in tumour cells, it can be utilised as a prognostic indicator [47].

2.1.6 Cervical cancer

Cervical cancer, the fourth most well established female malignancy worldwide, poses a severe danger to world's health and accounted for 90% of the 2,70,000 fatalities from the disease in 2015 [48]. Based on WHO, two strains (16 and 18) of human papillomavirus (HPV) cause over 50% of high-grade cervical pre-cancers. There are several proteins that have been proposed as cervical cancer biomarkers, including CDKN2A, Ki-67, PDL1, exportin-5, myeloid leukaemia cell proliferation protein (Mcl1) and cellular-FLICE-like inhibitory protein (c-FLIP) [49]. Not many studies have been published which can give us a detailed insight into the role of FAP in cervical cancer. Jin *et al*. reported the FAP expression in the epithelial tumour cells of cervical carcinoma [50].

2.1.7 Ovarian cancer

With over 240,000 new cases as of 2018, ovarian cancer ranked sixth among all female cancers globally [51]. Since EOC symptoms are non-specific, approximately 75% of patients receive an advanced stage diagnosis [52]. BRCA1 & BRCA2 mutation [53], Family history [54], Lynch syndrome [55], Uninterrupted ovulation cycles [56], Endometriosis, dietary factors [57], Ethnicity/race are some of the risk factors for ovarian cancer. Ovarian cancer expresses FAP,

although benign tumours and those with limited malignant potential do not [7]. In epithelial ovarian cancer, P. Mhawech-Fauceglia *et al.* has observed the expression of FAP on stromal and tumour cells [58] while the M. Zhang *et al*. has observed the FAP expression on tumour and mesothelial cells [59]. Both *in-vitro* and in an animal model, FAP encourages the growth, adherence, invasion, and migration of ovarian cancer cells [60]. Additionally, a reduced survival rate was associated with patients with ovarian cancer due to the over-expression of FAP [61].

2.1.8 Brain cancer

The electrical circuitry that regulates the functioning of the brain is composed of a number of cells, which includes neurons and astrocytes, giving neurons the support which is require to function properly. Astrocytoma, which are tumours that arise from astrocytes, are perhaps the most common kind of brain cancer in adults. Men are affected slightly more than women. The WHO divides astrocytoma into two categories: grade 1 brain tumours, which are the most benign, and grade 4 brain tumours, which are the most malignant [62]. Three crucial signalling pathways are dysregulated in glioblastoma, a grade IV astrocytoma. The pathways include the p53 route, which was affected in 87% of instances, the RTK/RAS/PI3K pathway, which was affected in 88% of cases, and the RB pathway, which was affected in 78% of cases [63]. Blood-brain barrier and lack of specificity of the potential harmful drug therapies have made pharmacological treatments for brain cancer notably ineffective [64].
Stem cells, stromal cells, macrophages and astrocytes all express FAP [64]. FAP is seen to be upregulated in most of the in most of the grade IV tumours. In glioblastoma, the TWIST transcription factor is thought to be a target of the FAP gene because it encourages invasion of mesenchymal-like cells that emerge because of TWIST-driven epithelial to mesenchymal cell transformation [65].

2.2 FAP in other diseases

2.2.1 Interstitial lung diseases (ILDs)

In lung fibroblasts that were at rest, there was no FAP protein at all. Immunofluorescence staining further supported the TGF-beta dependent elevation of FAP; hence, it can be concluded that FAP may assess the profibrotic activity of ILDs, which may help with early identification and the choice of a suitable treatment window [66].

2.2.2 Rheumatoid arthritis (RA)

RA is an immune-mediated disease that affects around 0.46% of the world's population. Fibroblast-like synoviocytes (FLSs) are thought to be extensively dispersed in the synovium and to be crucial in RA. The ability of FLSs to proliferate is significantly impacted by FAP overexpression. FAP expression contributed to the control of the cell cycle by cytokines like P53 and P21 and freed the contact inhibition between cells [67]. When examined alongside with normal articular cartilage, the presence of FAP is much greater in cartilage from individuals with osteoarthritis. Patients' cartilage exhibits a substantial increase in FAP expression, which implies that FAP may assist in cartilage deterioration. The expression of FAP is observed on chondrocytes in cartilage tissue, and the region surrounding the chondrocytes degrades first [68].

2.2.3 Heart

Cardiac fibroblasts are crucial for preserving the heart's structural integrity. Although excessive and prolonged activation of fibroblasts, which is a defining characteristic of the pathological remodelling of the heart, can increase tissue fibrosis, which is a crucial regulator of cardiac wound healing following damage [69]. Tumour necrosis factor - alpha (TNF-α), which is produced by macrophages, induces the production of FAP by human aortic smooth muscle cells (HASMC), and the extent of macrophage infiltration is correlated with the production of FAP. The weakening of the fibrous cap and formation of plaque are both related to FAP. Antibodies that neutralise FAP collagenolytic activity prevent FAP from cleaving collagen in the fibrous caps after it is produced, demonstrating that FAP is involved in the degradation of collagen in thin cap fibroatheromata [70].

2.3 Structural Analysis of FAP

FAP is 760 amino acid type 2 transmembrane serine proteases that belongs to the DPP family and share almost 50% homology with DPPIV, the family member with which it is most closely related [71]. In humans, the FAP gene is located on chromosome 2q23. Four enzymes that are part of the DPP protein family are —DPP8, FAP, DPP4 and DPP9, which can hydrolyse a prolyl bond located two amino acids from a protein's N-terminus [72]. DPPIV, FAP, DPP8, and DPP9 have comparable enzyme activity and are physically conserved, but because of their different expression and localization patterns, they are anticipated to fulfil a variety of functions

[73]. In this protein family, the catalytic triad-containing α-β-hydrolase domain is highly conserved [73]. This category of proteases is intriguing in several aspects of biology and can be used as a potential therapeutic targets because of the unique catalytic capacity that results from the proline (Pro) residue's cyclic form [72]. It was also shown that DPP4 cleaves and inactivates glucagon-like peptide-1 *in-vivo*, mutating insulin production by the pancreas, it was designated as a therapeutic target for type 2 diabetes.

FAP contains a trans-membrane domain (20 amino acid), cytoplasmic tail (6 amino acid) and extracellular domain (734 amino acid) [74]. The extracellular domain comprises of two separate domains (Figure 3): the 8 β-propeller domain (residues 54–492) that serves as a substrate selection gate and the α/β-hydrolase domain (residues 27–53 and 493–760) [75]. Because of the way the blades are arranged in the propeller domain, a central pore is created, making the active site accessible. A cavity created between the hydrolase domain and the -propeller domain allows for another method of accessing the active site [75]. The catalytic triad, which consists of the residues Ser624, Asp702, and His734, shapes the active site, which is located in a large cavity at the intersection of the α/β -hydrolase domains and β-propeller domains (shown in Figure 4). Only under particular circumstances and by modifying its structural domain can FAP be assembled as a dimer and activated. Another step in the process of creating a functioning FAP is N-linked glycosylation modification [76].

FAP is unique among its family members in that it has both DPP and endopeptidase activities. FAP's Endopeptidase activity is limited to the post-Pro bond following the glycine-Pro (GlyPro) motif, and the FAP's DPP activity permits it to cleave two amino acids off the N-terminus of a protein after a proline (Pro) residue. Synthetic libraries were used to explore the activities of FAP, and the results showed that endopeptidase substrates were strongly preferred by FAP for cleavage [77]. Proteolytic breakdown of ECM components increases angiogenesis and/or cancer cell movement, and is likely related to the gelatinase activity of FAP when it is expressed in ECM remodelling [73]. Additionally, its overexpression is linked to a bad prognosis and it is dramatically enhanced at sites of active tissue remodelling, such as wound healing and fibrosis, at sites of inflammation including arthritis and in atherosclerotic plaques [73] [77].

The S1 specificity pocket (Figure 2) is formed by the side chain of Tyr625, Val650, Trp653, Tyr656, Tyr660, and Val705 which form a hydrophobic environment. This site optimally accommodate proline (pro) residue. The hydrophobic S2 pocket (Figure 2), characterised by

residues Arg123, Phe350, Phe351, Tyr541, Pro544, Tyr625, and Tyr660, may accommodate large hydrophobic and aromatic residues.

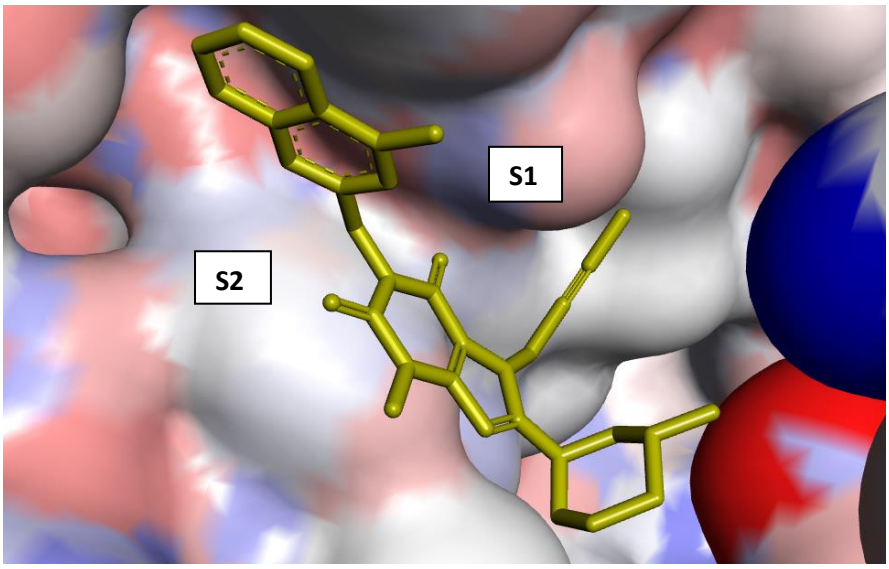


Figure 2: S1 and S2 pocket around the linagliptin (yellow colour)

Due to the inclusion of Ala657 in its structure, endopeptidase activity is shown in FAP, and this causes the active site of FAP to be less acidic.

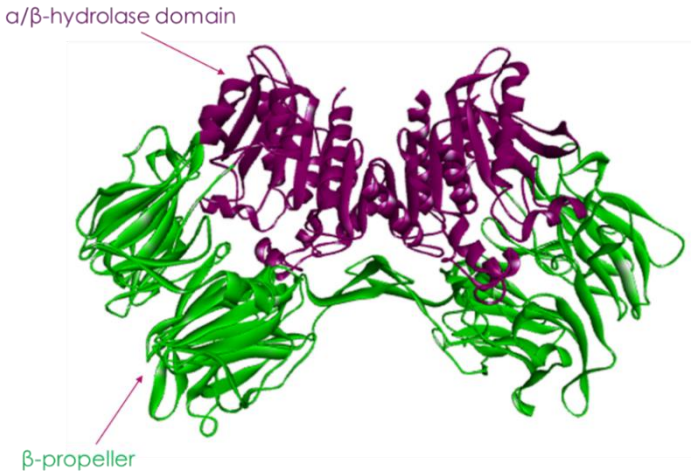


Figure 3: Structure of FAP representing the β-propeller domain in green colour and the α/β - hydrolase domain in magenta colour

A thorough analysis of FAP and DPP4 showed that DPP4 has a greater prominent negatively charged pocket owing to the occurrence of Asp663 (similar residue in FAP: Ala657), which is

in immediate proximity to Glu206. A structural analysis of DPP4 and FAP active sites revealed similar S2-S2 specificity pockets. DPP4's Asp663 moves Glu206 farther into the active site, exposing both acid oxygen atoms to the S2 pocket. This implies that all three acidic residues may help to identify and bind the free N terminal of peptide substrates.

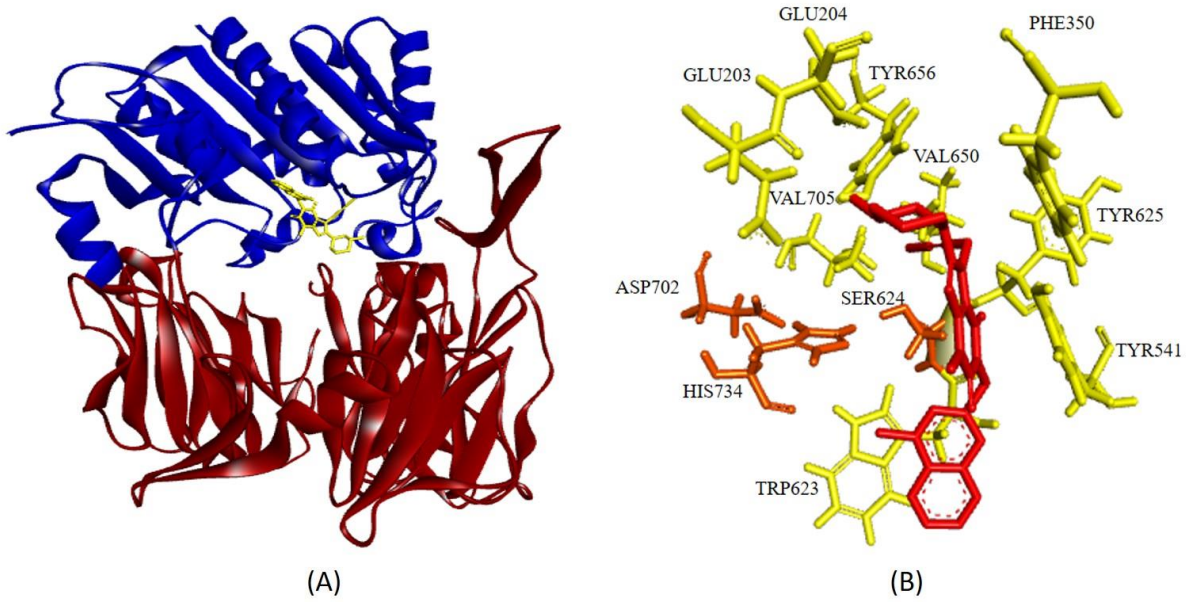


Figure 4: Structure of human FAP with the bound inhibitor in yellow and Catalytic & β-propeller domain in blue and red respectively (A), while the active site residue represented in stick model and the catalytic triad are highlighted in orange (B)

2.4 Computer-aided drug designing (CADD)

Drug discovery remains the most critical task for the scientific community. The entire drug discovery process, shown in Figure 5, typically takes between 10 and 15 years to complete and costs between US$800 million [78][79] and US$1.8 billion [80]. As a result, introducing a new medicine to the market is not only expensive but also requires a lot of time and labour. Rapid advancements in combinatorial chemistry and high-throughput screening (HTS) technology accelerated the discovery process in the 1990s, but they also made it harder to discover new molecular entities due to low hit rates and low conversion of hit to lead compounds, which prompted the creation of new techniques for removing unappealing compounds [79]. There have been several documented successes with recently produced drugs, including amprenavir for HIV infection treatment, zanamivir for influenza infection treatment, and dorzolamide for the treatment of cystoid macular edema [79].

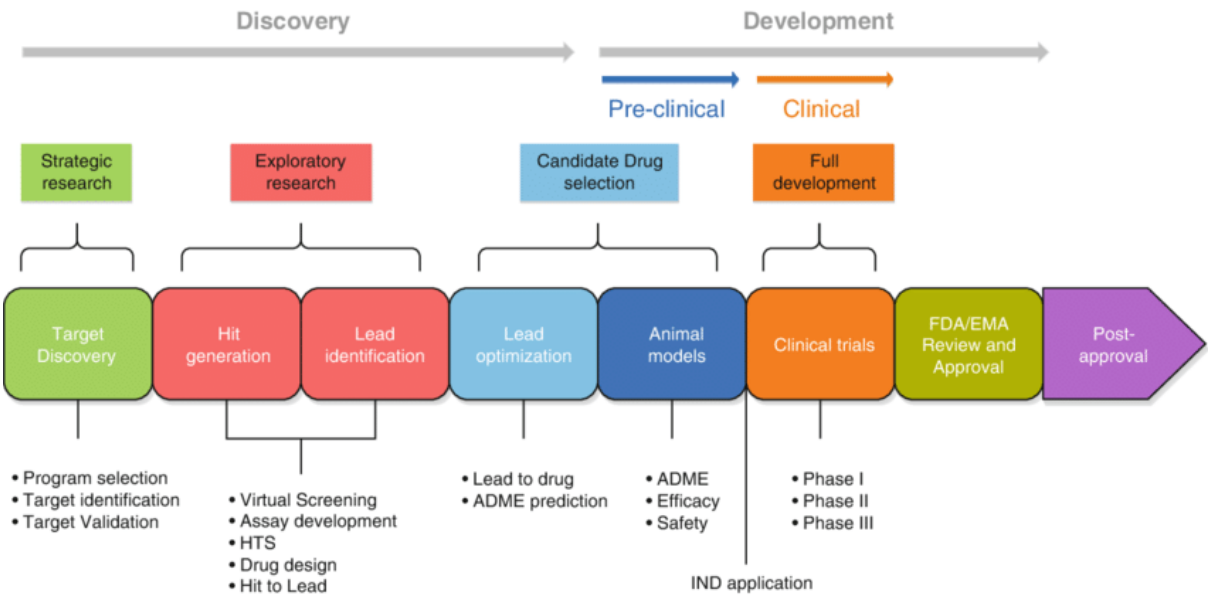


Figure 5: Drug discovery process

Drug development often uses CADD to screen sizable chemical libraries, build digital archives for the study of chemical interactions, enabling lead compounds to be improved by enhancing their biological features, evaluate possible lead compounds, etc. [79]. The systematic collection of biological data and systematic organisation of that data into related databases has been crucial to the contemporary drug development process. A variety of digital archive for small molecules have been compiled like PubChem (https://pubchem.ncbi.nlm.nih.gov/), which imparts information regarding 115 million compounds, Available Chemicals Directory (ACD) (https://www.psds.ac.uk/acd), serves as a central resource for docking studies, ZINC (https://zinc.docking.org/) holds over 230 million purchasable compounds in ready-to-dock formats and many others database like DrugBank (https://go.drugbank.com/), LIGAND, BindingDB (https://www.bindingdb.org/) and ChemDB (https://cdb.ics.uci.edu/) are available.

CADD approaches are classified as ligand-based (LB) or structure-based (SB) based on the accessibility and utilisation of the target structure. The overall workflow of CADD is described in figure 6. Database screening techniques are frequently utilised for hit identification, whereas hit optimization could be accomplished using a variety of techniques. Researchers are constantly updating and adopting new CADD processes with improved degrees of precision and reliability.

In SBDD, the target protein's binding site is utilised to create and assess ligands which are dependent on their anticipated interactions with the protein binding site [81]. Target identification is therefore a crucial phase in SBDD. For the creation of protein structures, a

variety of approaches have been used, including molecular dynamics simulation, homology modelling, X-ray crystallography, Cryo-Electron microscopy, and NMR. There are various drugs on the global marketplace that have been recognised by SBDD method. The main success story of SBDD is the development of FDA-approved medications that suppress HIV-1. Additionally, other drugs discovered using the SBDD approach include the antibiotic norfloxacin, the thymidylate synthase inhibitor, raltitrexed, and amprenavir, which may suppress the HIV protease [82]. The *de-novo* method and the VS method are two ways to categorise SBDD [83]. In VS, databases containing millions of molecules are computationally screened against the target proteins. Docking is used to screen chemical libraries, and ligands are sorted based on their affinity for binding and the best matches from the screening are subsequently submitted to *in-vitro* testing while using a *de-novo* technique, chemical structures of small molecules are created that bind to the target binding cavity with high affinity.

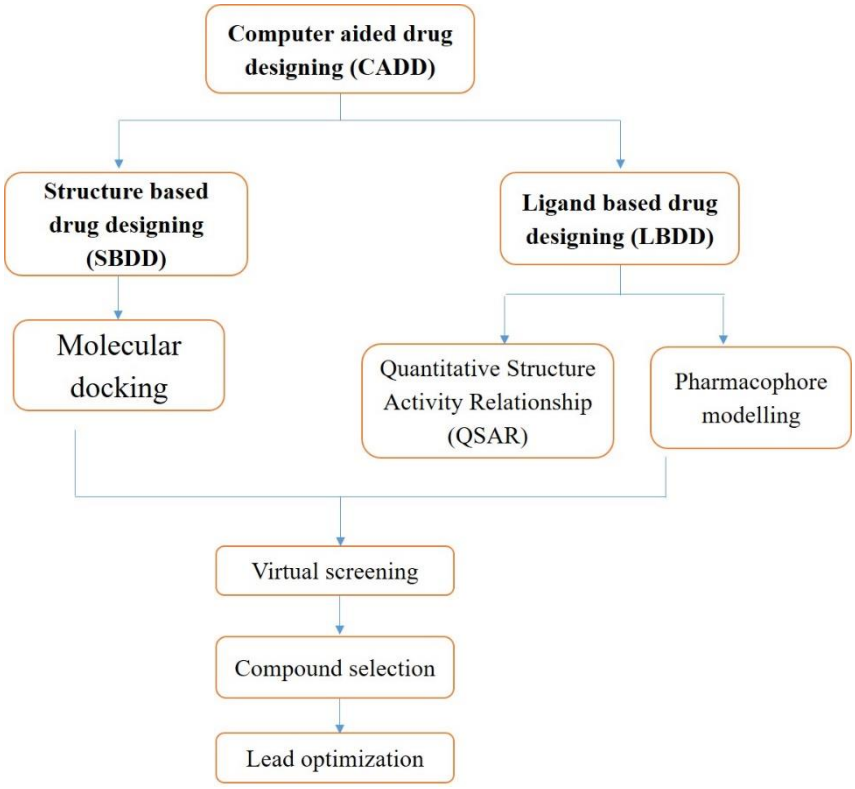


Figure 6: Workflow of computer aided drug design

The LBDD is used when the target structure is unknown and the data from a group of ligands that act on the desired target could be utilised to determine the important physicochemical characteristic (molecular descriptors) and structural that drives the biological activity of that target molecule [84]. Quantitative structure-activity relationships (QSARs) and pharmacophore-based approaches are typical LBDD strategies.

2.5 Molecular docking

During the last three decades, the demands of structural molecular biology and SBDD have spurred the growth of the discipline of molecular docking. [85]. The term "molecular docking" refers to a simulation method that combines a search algorithm for generating possible binding modes of a ligand into its receptor with a scoring function that rates them. Specifically, it seeks to assess the viable binding geometries of a suspected ligand with a target whose 3D structure is known. This approach offers insights into molecular recognition and is helpful in drug development and medicinal chemistry [86]. CADD now includes docking as a crucial component [86]. Docking is an effective method for lead optimisation since it allows for VS of large chemical libraries, rates the results, and proposes structural concepts to assess how the ligands can hinder the target's activity. [85]. Docking usually occurs between a small molecule and a target macromolecule, which is often referred by the term "ligand-protein docking", but protein-protein docking is also gaining popularity. The term "target" can be used to describe a protein, DNA, or RNA macromolecule [85]. The protein's 3D structure must be understood in order to produce the various poses of the ligands within the protein binding area which are ranked using various scoring function. Targets and ligands can be treated in a semi-flexible or static way with traditional docking approaches, although this is difficult. This result in the development of a number of docking methods which take into consideration for the flexibility of protein ligands and their variety of docking states [86]. Dock, AutoDock, Gold, and Glide are some of the docking software or programs that are used and because of their speed and simplicity, these programs are widely used in modern drug discovery approaches. New docking programmes, such as the EUDOC algorithm, SEED, and MM, have been released in the last year in addition to attempts to enhance well-known docking programs, such as AutoDock, DOCK, FlexX, FLOG, GOLD, GREEN, ICM, LUDI, QXP, and SLIDE [87].

It must be remembered that examining different docking methods or programs is challenging, and since there are indications that a few docking methods work more effectively with specific classes of targets than others, it is advised to try a few to see which one(s) best suits their demand. This can be done using a process called Re-Docking. "Re-docking" is the process of removing the ligand from a target complex of a known crystal structure, docking it back into the target, and then repeating the process. Then, researchers must determine how well their docking settings and methodologies can re-dock a variety of ligands to their target of interest. The root mean square deviation (RMSD) of the ligand's atoms in docked and crystallographic conformations is usually used as a measure of docking success if it is less than the rather

arbitrary threshold of 2Å [85]. Nowadays, it is common practice to utilize molecular docking simulation to investigate how various nanoparticles interact with proteins and nucleic acids. This aids in understanding their biological action's mechanism as well as in foretelling any prospective harm [88].

Molecular docking consists primarily of two stages: sampling conformations/orientations and a scoring function.

2.5.1 Scoring functions

Scoring functions serve as a pose selector i.e. they are used to evaluate the quality of docking poses [89], which are generated through the free energy perturbation (FEP) approach, which determines the relative binding free energies by gently mutating a ligand from one state to another.

A scoring function has three main objectives. The first is to clearly distinguish between the experimentally observed binding modes and all other poses that the search algorithm has investigated. The second step is to distinguish between active and inactive ligands. The third objective is to correctly rank the compounds according to their potency [89]. However, correctly predicting binding affinity and identifying active molecules among a variety of ligands remain difficult problems [90] [91].

There are three basic types of scoring functions: [92]

1. **Empirical scoring functions**

   This functions are indeed the sum of extensive empirical energy such as van der Waals, hydrogen bond, entropy, electrostatic, hydrophobicity, and so forth [93] and can be written as

$$\Delta G_{binding} = C\,1\Delta G_{vd} + C2\Delta G_{Hbond} + C3\Delta G_{entropy}$$

   Where the ci terms are the weighting coefficients obtained from the corresponding ΔG term.

   These descriptors are then weighted by regression method to reproduce the binding affinity with a training set. The key differences between the empirical functions relate to the quantity and kind of terms taken into account as well as the protein-ligand

complexes from the training set that were employed during the parameterization procedure. LUDI, GlideScore, ChemScore, $PLANTS_{CHEMPLP}$ etc. are some of the example of empirical scoring functions.

2. **Force-field based scoring functions**

   This scoring function calculates the potential energy of the system's bonded and non-bonded components using a molecular mechanics force field. The primary issues with force field-based scoring functions are overestimation of interactions between charged atoms and atomic collisions caused by high repulsion at close range due to the Lennard-Jones potential [89]. GoldScore, AutoDock are some of the examples of force field-based scoring function.

3. **Knowledge-based scoring functions**

   To avoid deriving weights from experimental data, statistical analysis is used in this scoring function to obtain potential based on interatomic contact between atoms. This approach has a propensity to accurately reflect the complexities that accompany more specific binding effects. However, the quantity and variety of the experimental structures utilized greatly influence this strategy [89]. GOLD/ASP and DrugScore [94] are examples of knowledge-based scoring functions.

Although all three scoring functions have produced positive results, there is still no agreement on which one is better for docking displays [87]. Finding and using the scoring function that is most suited to the topic being studied is the recommended technique most frequently to boost the effectiveness and dependability of molecular docking investigations.

2.5.2 Sampling

Molecular docking is a computational approach used for determining the best conformation of a receptor-ligand complex. So, every docking method utilizes one or more particular search algorithms, which are used to determine the most likely shapes of a protein-ligand complex. The global minima will be equal to the experimental receptor-ligand configuration, and local minima will be equivalent to different binding modes if the energy function adequately models the system's thermodynamics. Present docking methods rely on estimations since accounting for entropic effects is typically not simple. By light of this, there is no assurance that the global minimum examined by a docking approach adheres to the local minima. The majority of

algorithms treat the ligand as flexible while treating the protein as rigid, meaning that while sp3 bonds can spin, bond lengths and angles are maintained constant. Degree of flexibility of the molecule divides the docking into following types:

1. **Rigid docking:**

   This procedure is similar to the "lock-key" paradigm, in which both the ligand and the protein are solid (inflexible) structures with just 3 rotational and 3 translational degrees of freedom considered during sampling.

2. **Flexible docking:**

   Both protein and ligand are considered flexible counterparts in this technique, and so have the largest degree of freedom, increasing the computational effort necessary to complete a docking.

3. **Semi-flexible docking:**

   Only one molecule, the ligand, is flexible in this manner, whereas the protein is rigid. As a result, the ligand has a degree of freedom in addition to the six translational and rotational ones.

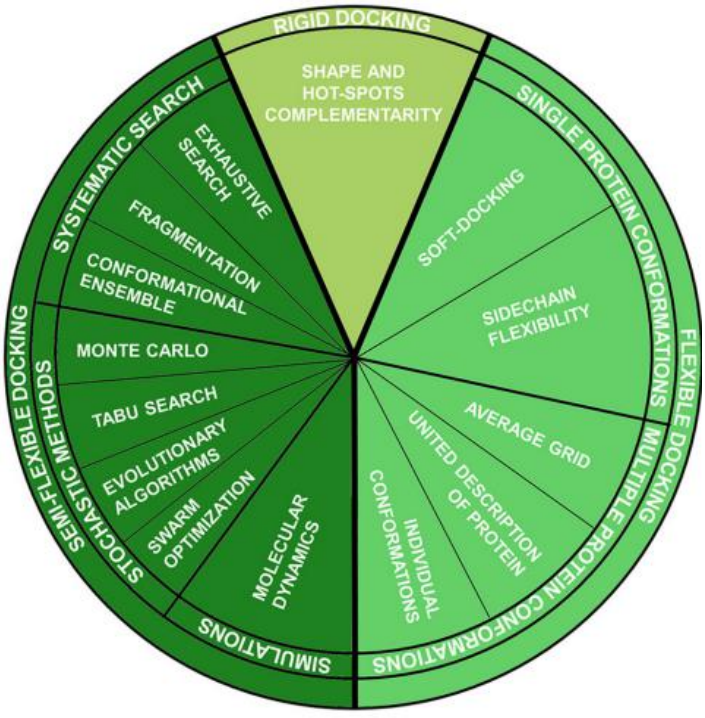


Figure 7: Molecular docking approaches classified by Engines for ligand-protein flexibility and conformational searches

2.6 Molecular dynamics simulation

MD is the most extensively utilized computational tool and an effective way for studying atoms and molecules' physical dynamics. The atoms and molecules are programed to interact for a set amount of time, revealing the dynamic character of the system. The trajectory of atoms and molecules in the simulated environment is numerically determined using Newton's second law of motion (F = ma). The potential energy of the system can be determined using techniques of molecular mechanics with various force fields. MDS can detect not just structural variations caused by environmental changes such as pH and temperature, but also the complex interplay of protein mis-folding and aggregation, which is impossible to detect using experimental methods like as NMR and X-ray crystallography. Various techniques such as x-ray crystallography, NMR, FRET and Cryo-EM can be used in combination with MDS for the prediction of protein structure as these simulation can find the position of atom more accurately. The conformational changes in protein and various biomolecular processes can also be captured using the simulation. In the CADD process, MDS can be a potent tool. For example, Lee and Kollman *et al.* [95] suggest a novel substance as a possible inhibitor of thymidylate synthase, the function of water molecules in HIV-1 protease-substrate complexes were investigated by Okimoto *et al.* [96]. In simulations lasting up to 50 ns, Marrink *et al.* investigated the spontaneous production of dodecylphosphocholine surfactant micelles [97]. Larger and more complicated systems can now be simulated using MD techniques because to advances in computing power [99]. The first simulation took less than 10 ps to complete; it has since grown to last as long (10 ns), and therefore requires less time (nearly 50%) for a system of the same size. Now, it has been feasible to identify a pathway between the closed and open conformations using MDS, which cannot be done using experimental methods. A direct modelling of this system is not yet achievable in the presence or absence of ATP since it is thought that the real shift between the open and closed structures occurs on the millisecond timescale [100]. The underlying dynamics of a protein in a single system can be experimentally determined by running simulations with one component of the system at one temperature and the other component of the system at a different temperature.

The most precise information on the movements of individual particles over time may be obtained via simulations. As a result, they are frequently more effective than tests on the real system in addressing particular concerns concerning the characteristics of a model system. The fact that simulation potentials, although being approximate, are entirely controllable by the user, allowing for the examination of their role in deciding a particular attribute, is another

important feature of simulations. Virtual screening has been employed in the CADD process to improve drug discovery efficiency, however molecular docking fails to offer a more flexible protein structure due to the less dynamic environment, so MDS can improve protein-ligand binding accuracy and thus acquire various target conformations.

In general, MDS have been used to discover potential binding sites on target proteins, as well as to compute the binding free energy between target proteins and drug molecules, determining the mechanism of action of drug molecules, etc. [101]. Based on the models and mathematical formalisms used, there are two basic ways to represent atoms: quantum mechanics and classical mechanics. GROMACS is one of the example of the software which is used for performing simulation.

The following structural parameters can be used to examine the outcomes of simulation:

1. RMSD
2. Secondary structure analysis
3. RMSF
4. Principal Component Analysis (PCA)
5. Radius of Gyration

Working with atomic-level computer simulations of the required biomolecules is a promising alternative to a more traditional experimental approaches because they give detailed insights into the structural information on macromolecules dynamics and the conformational transformations they undergo. As computing power increases, such technologies will become even more important.

CHAPTER 3: MATERIALS AND METHODS

3.1 Compound selection and dataset creation

From the BindingDB (https://www.bindingdb.org/rwd/bind/index.jsp) a database of FAP inhibitor molecules was obtained. There were 636 FAP inhibitor molecules downloaded in total. The chemical structures and activities (IC50 values) of the compounds in this database were varied. Based on their IC50 values, the compounds are selected from the database. The compound with low IC50 value were selected for further docking. The 2D structures of these compounds were created using BIOVIA Draw 2022 and exported to DS to create the matching 3D structures. Structures created with BIOVIA Draw 2022 are stored in MOL format, whereas 3D structures created with DS are saved in PDB format

3.2 Preparation of protein structure

The RCSB Protein Data Bank (PDB) was used to retrieve the experimentally discovered X-ray crystal structure of Human FAP alpha (PDB: 6Y0F), which has a resolution of 2.92 Å and an R-value free of 0.243. The ligand linagliptin forms a complex with the human FAP alpha protein. Due to its homo-tetrameric structure, only chain A and its ligand (linagliptin) is selected for further processing. The *Clean Protein* functionality of the DS software was used to prepare only chain A without the linagliptin for the protein preparation part.

The *clean protein* functionality is utilized for the preparing the protein structure. This is accomplished by going to the *"Macromolecules"* section of DS software. The parameter of the clean protein can be modified by using the "*Edit"* and then the "*Protein Utilities"* section of the DS software as well. In this investigation, all of the default parameters were employed, including the addition of hydrogen atoms and the correction of bond order at pH 7.00. The functionality of *"Prepare Protein"* carries out operations such as "addition of missing atoms, correction of connectivity, correction of names, insertion of missing loops and correction of other issues such as structural disorders (misfolding, disulphide bond formation, etc.). The protein clean report for molecule does not indicate any substantial changes in the protein structure during the preparation of chain A of the 6Y0F protein utilising the clean functionality of the DS program, with the exception of one terminal residue being changed. The UCSF Chimaera was then used to minimise energy in the clean protein structure. The cleaned protein structure file was opened in the chimera, and in the next steps the energy of the structure is minimized by utilizing the *"Structure Editing"* option present in the "*Tools"* section of UCSF

chimera. The force field AMBER ff14SB was utilised in this investigation along with all of the default parameters, such as steepest descent steps (100), conjugate gradient steps (10), etc. Following energy minimization, the RMSD value of the energy minimised protein structure is calculated by utilising the "Structural analysis" option present in the Tools section of UCSF Chimera

Using the Superimpose Proteins capabilities of the DS software, the clean protein structure and the energy-minimized structure were superimposed, and the RMSD value was produced.

3.3 Preparation of ligands

There are two databases; the first is a collection of well-known, highly effective inhibitor molecules, the preparation of which is covered in the section on compound selection and dataset compilation. The second one is the unknown database, of Medicinal Fungi Secondary Metabolite and Therapeutics (MeFSAT) which is obtained from Vivek-Ananth *et al.,* [103]. In this study, the known chemical database obtained from the BindingDB is represented by database 1, whereas the unknown database is represented by database 2.

DS's *Prepare Ligands* feature is used to prepare Database 1. *Prepare Ligand's* functionality performs tasks including eliminating duplication, counting isomers and tautomer, fix bad valencies, set standard formal charges on common functional groups and creating 3D conformations. In this protocol, all default parameters are utilised. Energy minimization of this database is done after utilising the *Prepare Ligand* feature. This study makes use of the DS's *dreiding* capabilities for minimizing the prepared ligand database. The same process is utilised to prepare Database 2, except Charmm force field is employed for the energy minimization step.

3.4 Pre-docking studies

The Re-Docking approach is used to evaluate the precision of the DS LibDock program. Downloading the crystal structure (PDB ID: 6YOF) is followed by the preparation of the protein and ligand, the details of which are provided in the sections on the preparation of protein structures and preparation of ligands, respectively.

Docking is performed using the LibDock program of the DS. The parameter which are utilized for this Re-Docking approach are – *CAESAR* (Conformation Method), 300 (Hotspot), 0.50

(Docking Tolerance), High Quality (Docking Preferences), TRUE (Separate Conformation), 255 (Maximum Conformations). The RMSD is determined because it provides a reliable indicator of how well their atomic locations match, atom by atom. It is generally agreed upon that a docking method reproduces a crystallographic conformation below 2.0 Å can be utilized for further docking method.

3.5 Docking based virtual screening

Because of increase in the cost and time of the drug development process, virtual screening methods, are being more extensively used and continue to be refined. Virtual screening approaches are being employed increasingly frequently and are still being improved as a result of the lengthening of the drug development process and its associated costs. Database screening was conducted for database 2 (which contain 1919 compounds) by the LibDock program of DS to identify the novel compounds. The parameter used for performing this process are -- FAST (Conformation Method), 100 (Hotspot), 0.25 (Docking Tolerance), Fast Search (Docking Preferences), TRUE (Separate Conformation), 255 (Maximum Conformations).

Molecular docking operations were carried out 4 times in this investigation. First, to evaluate how accurate the LibDock program is. Second, for database1. Third for database 2 and fourth with the top 10 compounds obtained from the database 2.

The active site residues are Glu203, Glu204, Phe350, Tyr541, Trp623, Tyr624, Tyr625, Val650, Tyr656, Val705 and His734. By choosing the active site residues, a group referred to as "Active Site" is created and by choosing this Active site group the binding site is defined. A binding sphere gets generated, whose coordinates are adjusted to the following points; X = 41.991857, Y = -72.891943 and Z = 85.148 with radius 15 Å.

After specifying the binding site, docking is carried out using the same protocols as in the sections on Pre-docking studies. From the database 1, compounds with low IC50 value are selected and docked with the target protein. From this the compound with high LibDock score are selected and used as a reference compound. After VS of database 2, top 10 compounds with high LibDock score are chosen for further docking experiment. The results are analysed based on the orientation, molecular interactions and the LibDock score.

CHAPTER 4: RESULTS

4.1 Validation and molecular docking of reference compound

To find promising leads for FAP inhibition, integrated computational techniques such as structure-based virtual screening and molecular docking were used. The strategy of which is depicted in Figure 8.

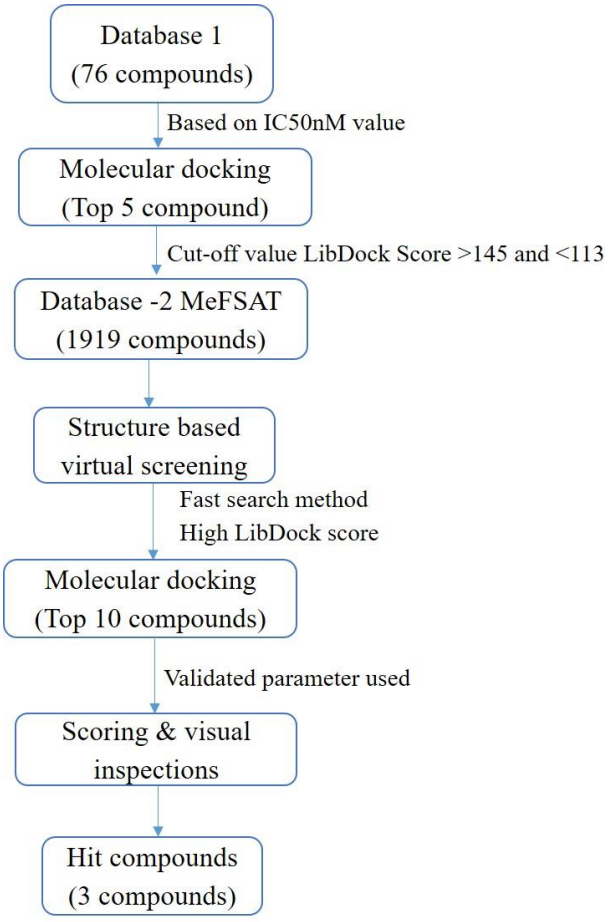
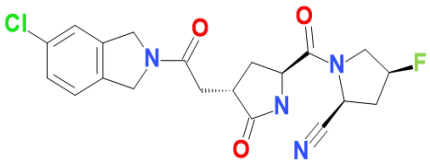
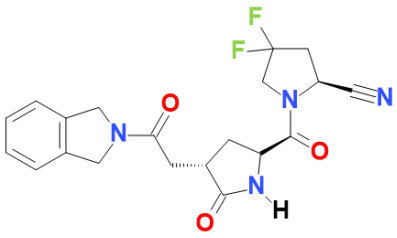
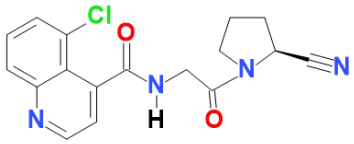
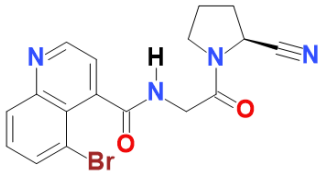
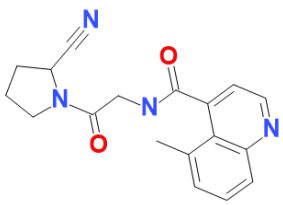


Figure 8: Schematic workflow of VS for identifying FAP inhibitors

The suitability of the DS LibDock program was evaluated prior to its use. The FAP co-crystal ligand (PDB ID - 6Y0F) was docked into the X-ray crystal structure, and the docking pose was obtained with a low RMSD value of 0.8549 Å, indicating the reliability of the LibDock program for this investigation. Following that, using the evaluated parameters, all the compounds mention in Table 1 are docked into the active site of FAP and the compounds with the high LibDock Score are used as a reference which are linagliptin and compound 1 & 2 which are shown in figure 9, 10 & 11 respectively. Each ligand's optimal pose was chosen based on a high LibDock score, chemical interactions, and orientation. The LibDock scores for reference 1, reference 2 and reference 3 were 145.53, 137.96, and 113.24, respectively, and

these scores are utilised as threshold values for additional binding poses. The molecular interaction and orientation of each reference pose were carefully evaluated manually with the active site residue of FAP receptor and are mentioned in table 3.

Table 1: Compounds with their IC50 value and LibDock score

Compound no.	Structure	IC50(μM) of FAP	LibDock Score
1.		0.063	137.962
2.		0.017	113.24
3.		0.0099	93.5742
4.		0.011	92.6014
5.		0.0043	91.9049

4.1.1 Interaction analysis of reference compounds

The reference 1 compound has the LibDock score of 145.53. It interacts with Glu203, Glu204, Tyr625 and Tyr656 through hydrogen bonding, with Trp623 via Pi-sigma interactions, with Phe350, Tyr541, Val650, Tyr656 and Val705 via Pi-alkyl interaction and forming the Pi-Pi interaction with Tyr541 as shown in figure 9. The reference 2 compound has the LibDock score of 137.96. It forms hydrogen bond with Arg123, Tyr625, Tyr656 and Asn704. The Pi-alkyl interaction are observed with Phe350 and Tyr660 residues and the Pi-sigma interaction are seen with Tyr541 and Trp623. The reference 3 compound has the LibDock score of 113.24. It forms hydrogen bond with Arg123, Trp623, Tyr625, Tyr660 and Asn704 residues and the Pi-Pi and Pi-alkyl interactions with Tyr745 and Trp621 respectively.

The active site of FAP contain residue such as Glu203, Glu204, Phe541, Trp623, Ser624, Tyr625, Val650, Tyr656, Val705 and His734, which are important for FAP inhibition. Therefore, the hit compounds, forming the stable hydrogen bond with these residue were given priority as potential hit.

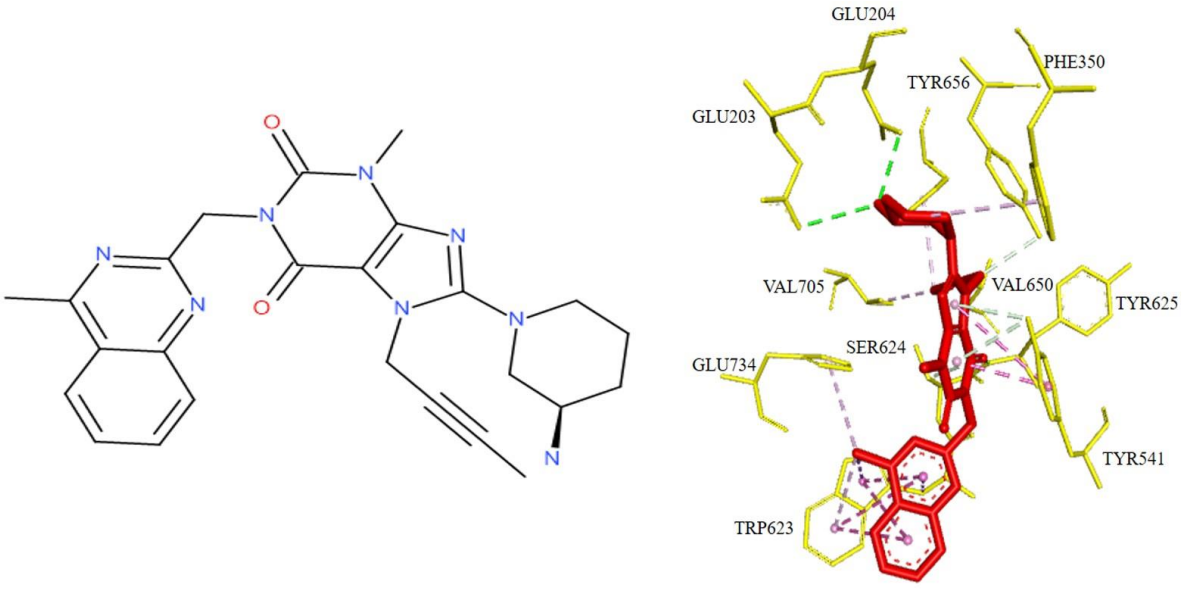
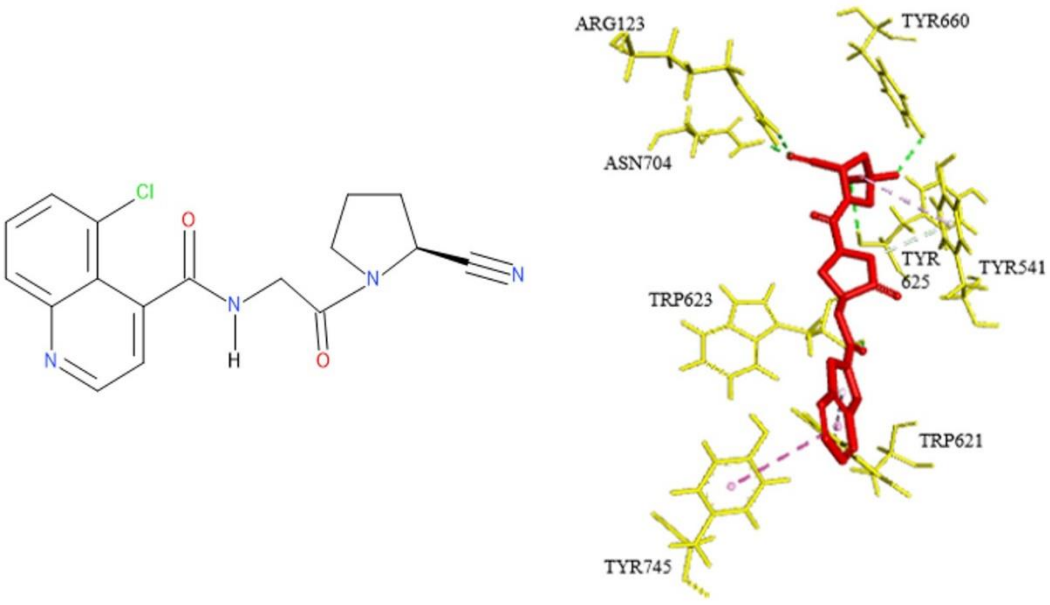
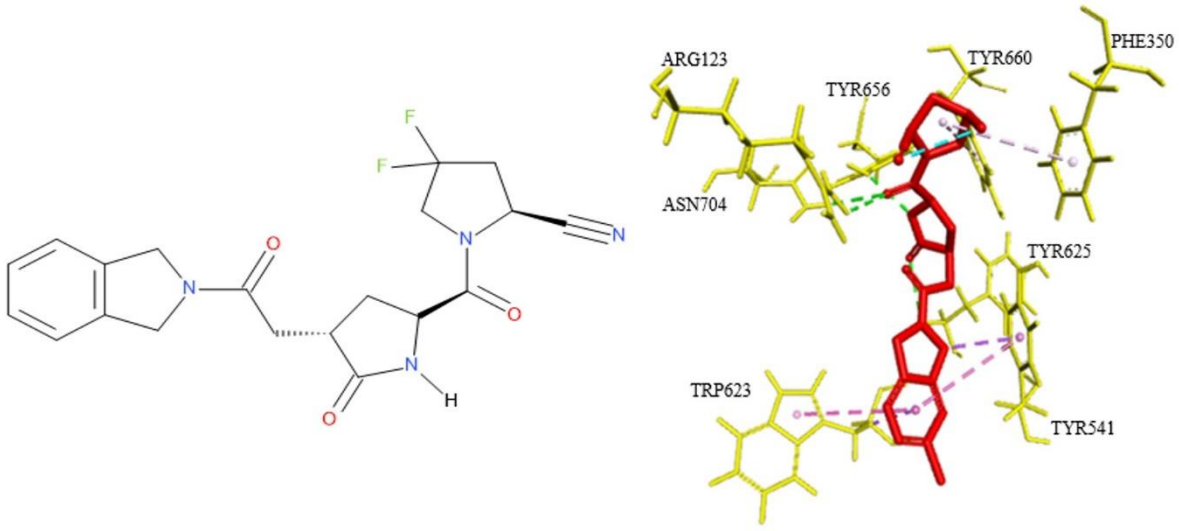


Figure 9: Reference 1 with (A) 2D structure and (B) 3D structure with H-Bond in green dotted line



(A) (B)

Figure 11: Reference 2 with (A) 2D structure and (B) 3D structure with H-Bond in green dotted line



(A) (B)

Figure 10: Reference 3 with (A) 2D structure and (B) 3D structure with H-Bond in green dotted line

4.2 Virtual screening and docking

The database 2 is virtually screened against the FAP receptor using the parameter described in the virtual screening section. The top 10 compounds based on their LibDock score are selected and shown in table 2 and these compounds are further docked using the parameter that are used assessing the LibDock program.

Table 2: Virtually screened compounds with their LibDock score

S. No.	Compound ID	LibDock Score
1.	MSID000698	203.088
2.	MSID000696	202.96
3.	MSID000253	196.396
4.	MSID000693	190.921
5.	MSID001017	190.892
6.	MSID000697	186.553
7.	MSID001261	186.468
8.	MSID001315	186.464
9.	MSID001010	181.51
10.	MSID000253	179.387

After docking these top 10 compounds, only 3 compounds are selected based on the pose, orientation and the molecular interaction. The selected three compounds possess all of the critical features required for FAP inhibition.

4.2.1 Interaction analysis and selection of hit compounds

The first hit molecule (MSID001315) with the LibDock score of 172.04, forms all and also some additional hydrogen bond (shown in Figure 12) which are critical for FAP inhibition and its stability, with the residues Tyr54, Arg123, Glu203, Glu204, Phe553, Val555, Trp623, Tyr625, Gly626 and Tyr656. It also shows the Pi-alkyl interaction with Trp621 residue.

The second hit molecule (MSID001017) with the LibDock score of 149.81, forms hydrogen bond (shown in figure 13) with Tyr625 and Tyr656 residues only. It forms the Pi-alkyl interaction with Phe351, Tyr541, Trp621 and Trp623 and the Pi-Pi interaction is observed with His734 residue.

The third hit molecule (MSID001261) with the LibDock score of 179.05, forms hydrogen bond (shown in figure 14) with Arg123, Glu204, Trp557, Tyr656 and Tyr745. The residues Ile47, Phe350, Ile558, Tyr660 and His734 shows the Pi-alkyl type interaction with the ligand. Furthermore, Tyr541 shows the Pi-sigma interaction with the ligand.

Table 3: Hits and reference compound containing their LibDock Score and H-bond interaction

S.No.	Molecule	LibDock Score	H-Bond Interaction
1.	Linagliptin Crystal structure (PDB ID – 6Y0F) Reference -1	145.531	GLU203, GLU204, TYR625, TYR656
2.	Reference 2	137.962	ARG123, TYR625, TYR656, ASN704
3.	Reference 3	113.24	ARG123, TRP623, TYR625, TYR660, ASN704
4.	MSID001315	172.044	TYR54, ARG123, GLU203, GLU204, PHE553, VAL555, TRP623, TYR625, GLY626, TYR656
5.	MSID001261	179.05	ARG123, GLU204, TRP557, TYR656, TYR745
6.	MSID001017	144.755	TYR625, TYR656

Overall, the identified hit molecules, MSID001315, MSID001261, and MSID001017, show excellent interactions with active site residues and hence may play a role as an inhibitor of FAP protein. To summarise, the work used an integrated strategy of molecular docking and a virtual screening technique to find possible lead compounds. As a result, the detected hit compounds form hydrogen bonds, Pi-alkyl interactions, and Pi-Pi interactions comparable to those seen in the reference compounds. Our findings gives a structural explanation that is needed and could serve as a starting point for further optimisation and development of new medicines targeting the FAP receptor.

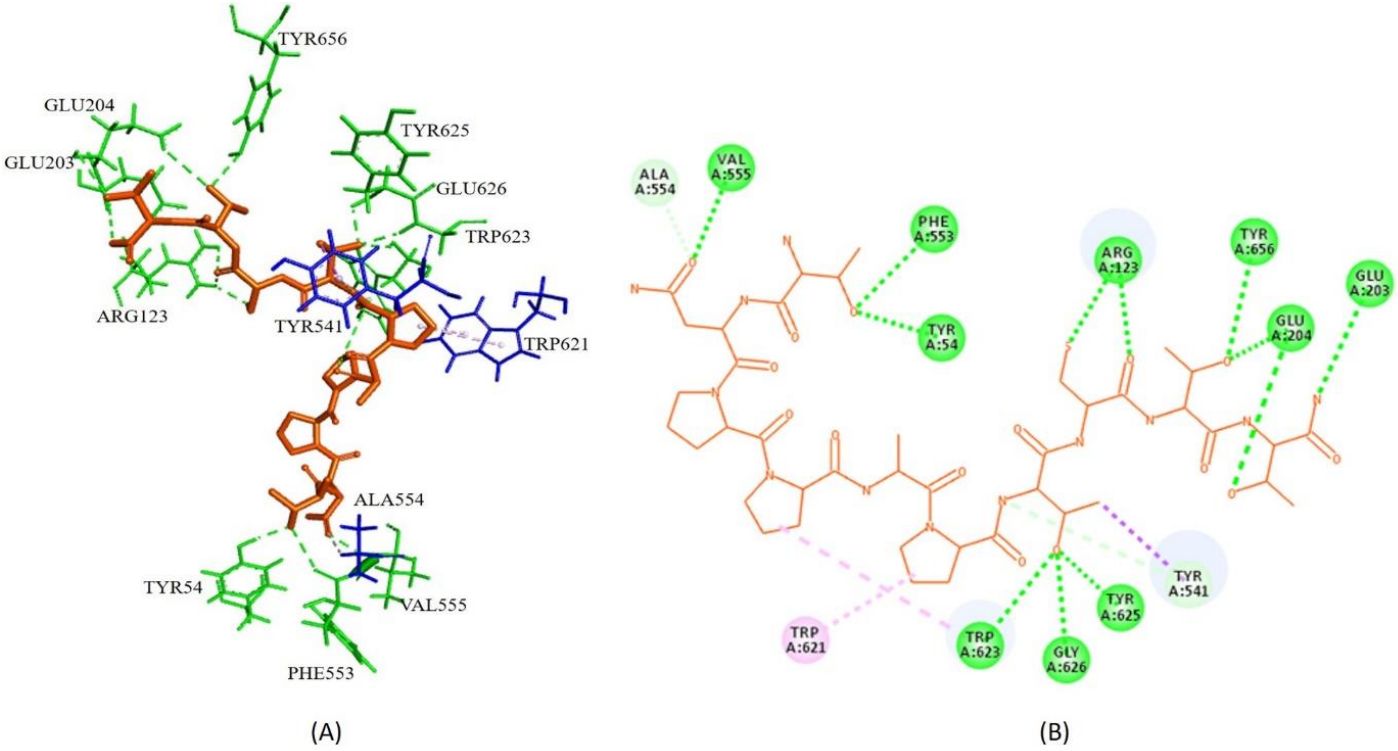


Figure 12: 3D (A) and 2D (B) structure of hit compound (MSID001315) with hydrogen bond forming residue in green colour

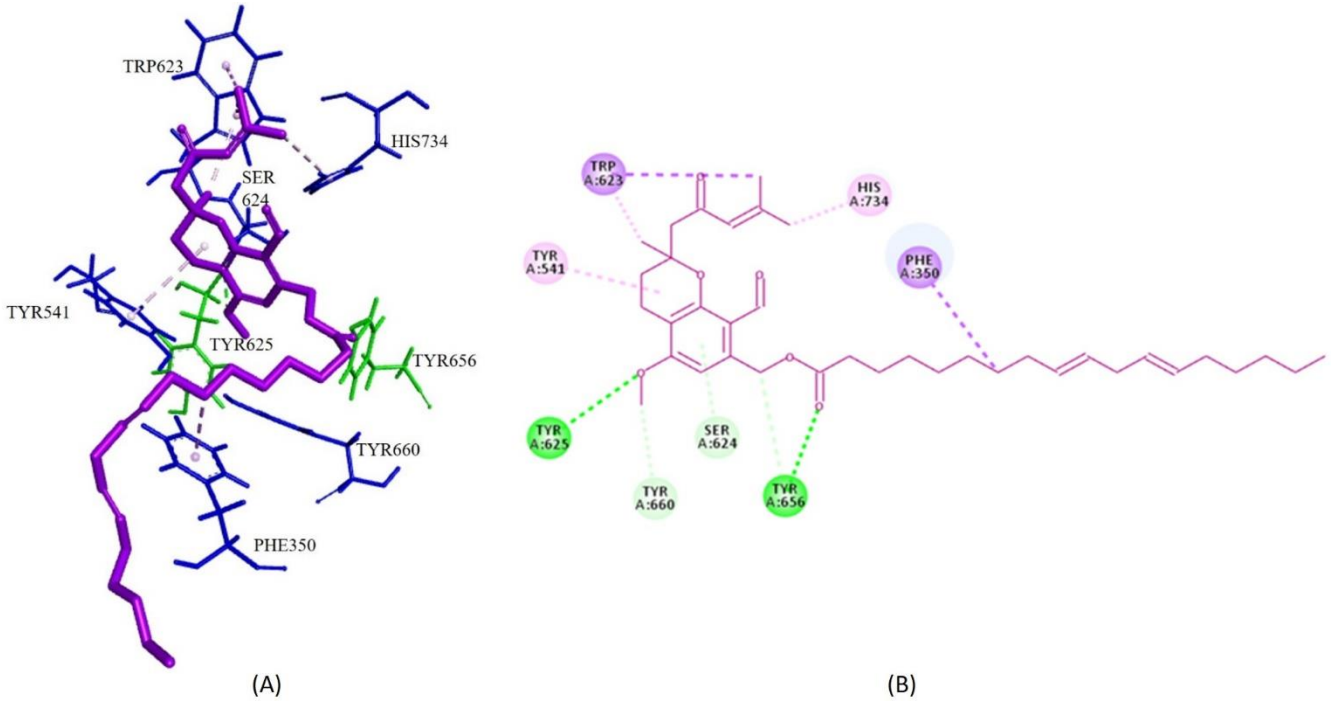


Figure 13: 3D (A) and 2D (B) structure of hit compound (MSID001017) with hydrogen bond forming residue in green colour

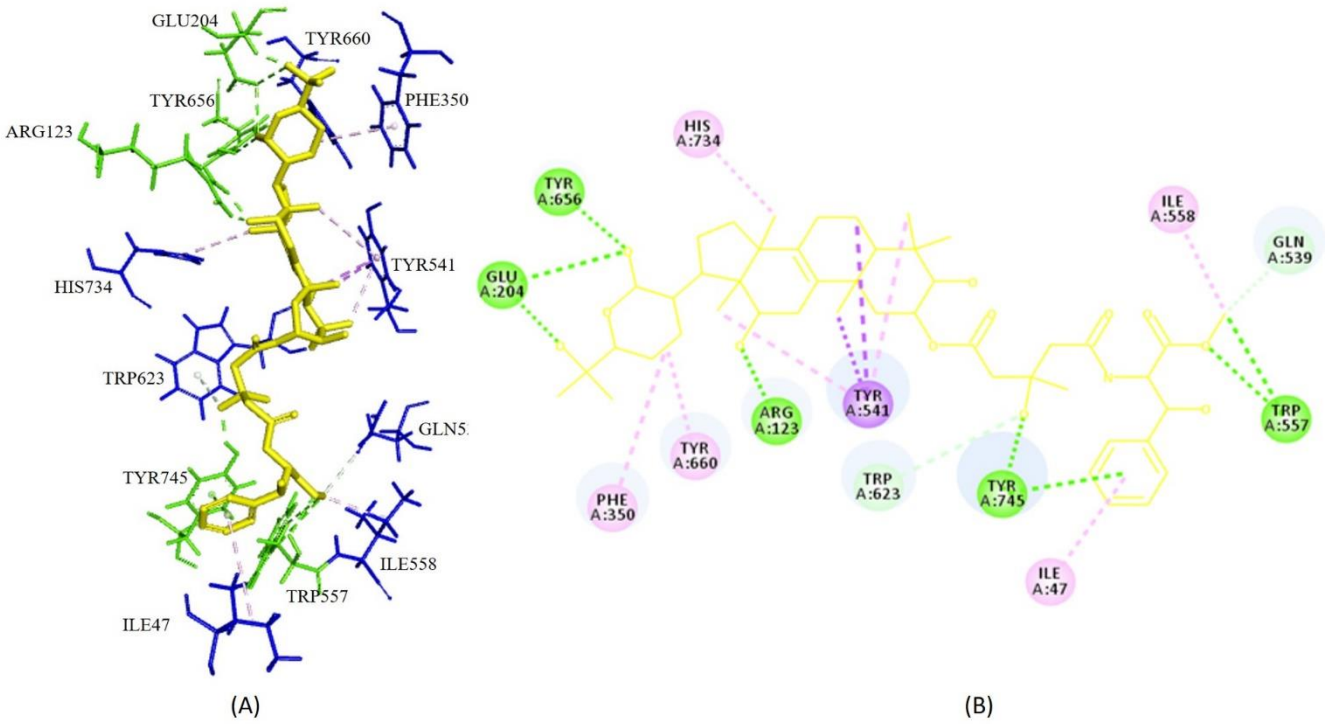


Figure 12: 3D (A) and 2D (B) structure of hit compound (MSID001216) with hydrogen bond forming residue in green colour

CHAPTER -5: CONCLUSION

Finding a drug that inhibits a certain target is always a difficult undertaking. Many therapeutic compounds make it to clinical trials but are ineffective, resulting in a waste of limited resources. In this case, the CADD technique comes in handy. As a result, in this investigation, molecular docking and structure-based virtual screening were used to find compounds capable of specifically inhibiting FAP. Molecular docking was performed to identify the amino acids to which the ligand binds as well as the optimum binding position. Virtual screening is also carried out, as a consequence of which 10 compounds were obtained, of which three were chosen: MSID001315, MSID001261, and MSID001017. These compounds have a chance to become the lead compound.

Furthermore, MDS can be done on these selected compounds to determine whether the interactions that exist can persist in the dynamic behaviour as well. Following that, these compounds can be evaluated *in-vitro* also.

References

[1] W. Cao, H.-D. Chen, Y.-W. Yu, N. Li, and W.-Q. Chen, "Changing profiles of cancer burden worldwide and in China: A secondary analysis of the global cancer statistics 2020," *Chinese Medical Journal*, vol. 134, no. 7, pp. 783–791, 2021. doi:10.1097/cm9.0000000000001474

[2] M. Maule and F. Merletti, "Cancer transition and priorities for cancer control," *The Lancet Oncology*, vol. 13, no. 8, pp. 745–746, 2012. doi:10.1016/s1470-2045(12)70268-1

[3] *Ecancermedicalscience*. doi:10.3332/ecancer.2012.ed16

[4] J. Li *et al.*, "Hulc and LINC00152 Act as novel biomarkers in predicting diagnosis of hepatocellular carcinoma," *Cellular Physiology and Biochemistry*, vol. 37, no. 2, pp. 687–696, 2015. doi:10.1159/000430387

[5] D. V. Makarov, S. Loeb, R. H. Getzenberg, and A. W. Partin, "Biomarkers for prostate cancer," *Annual Review of Medicine*, vol. 60, no. 1, pp. 139–151, 2009. doi:10.1146/annurev.med.60.042307.110714

[6] J. Jia, T. A. Martin, L. Ye, and W. G. Jiang, "FAP-α (fibroblast activation protein-α) is involved in the control of human breast cancer cell line growth and motility via the fak pathway," *BMC Cell Biology*, vol. 15, no. 1, 2014. doi:10.1186/1471-2121-15-16

[7] P. Garin-Chesa, L. J. Old, and W. J. Rettig, "Cell surface glycoprotein of reactive stromal fibroblasts as a potential antibody target in human epithelial cancers.," *Proceedings of the National Academy of Sciences*, vol. 87, no. 18, pp. 7235–7239, 1990. doi:10.1073/pnas.87.18.7235

[8] D. J. Wessels *et al.*, "Reciprocal signaling and direct physical interactions between fibroblasts and breast cancer cells in a 3D environment," *PLOS ONE*, vol. 14, no. 6, 2019. doi:10.1371/journal.pone.0218854

[9] Z. Wang *et al.*, "Role and mechanism of fibroblast-activated protein-α expression on the surface of fibroblast-like synoviocytes in rheumatoid arthritis," *Frontiers in Immunology*, vol. 14, 2023. doi:10.3389/fimmu.2023.1135384

[10] A. Šimková, P. Bušek, A. Šedo, and J. Konvalinka, "Molecular recognition of fibroblast activation protein for diagnostic and therapeutic applications," *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, vol. 1868, no. 7, p. 140409, 2020. doi:10.1016/j.bbapap.2020.140409

[11] T. Kelly, Y. Huang, A. E. Simms, and A. Mazur, "Fibroblast activation protein-α," *International Review of Cell and Molecular Biology*, pp. 83–116, 2012. doi:10.1016/b978-0-12-394308-8.00003-0

[12] Sun YS, Zhao Z, Yang ZN, Xu F, Lu HJ, Zhu ZY, Shi W, Jiang J, Yao PP, Zhu HP. Risk Factors and Preventions of Breast Cancer. Int J Biol Sci 2017; 13(11):1387-1397. doi:10.7150/ijbs.21635

[13] J. Dine and C.-X. Deng, "Mouse models of BRCA1 and their application to Breast Cancer Research," *Cancer and Metastasis Reviews*, vol. 32, no. 1–2, pp. 25–37, 2012. doi:10.1007/s10555-012-9403-7

[14] R. Ali and M. K. Wendt, "The paradoxical functions of EGFR during breast cancer progression," *Signal Transduction and Targeted Therapy*, vol. 2, no. 1, 2017. doi:10.1038/sigtrans.2016.42

[15] A. R. Green *et al.*, "Myc functions are specific in biological subtypes of breast cancer and confers resistance to endocrine therapy in luminal tumours," *British Journal of Cancer*, vol. 114, no. 8, pp. 917–928, 2016. doi:10.1038/bjc.2016.46

[16] Y. Pylayeva-Gupta, E. Grabocka, and D. Bar-Sagi, "Ras oncogenes: Weaving a tumorigenic web," *Nature Reviews Cancer*, vol. 11, no. 11, pp. 761–774, 2011. doi:10.1038/nrc3106

[17] C. R. Boland and M. B. Yurgelun, "Historical perspective on Familial gastric cancer," *Cellular and Molecular Gastroenterology and Hepatology*, vol. 3, no. 2, pp. 192–200, 2017. doi:10.1016/j.jcmgh.2016.12.003

[18] N. Ahmed, "23 years of the discovery of helicobacter pylori: Is the debate over?," *Annals of Clinical Microbiology and Antimicrobials*, vol. 4, no. 1, 2005. doi:10.1186/1476-0711-4-17

[19] R. Ladeiras-Lopes *et al.*, "Smoking and gastric cancer: Systematic review and meta-analysis of Cohort studies," *Cancer Causes & Control*, vol. 19, no. 7, pp. 689–701, 2008. doi:10.1007/s10552-008-9132-y

[20] K. Okada *et al.*, "Seprase, a membrane-type serine protease, has different expression patterns in intestinal- and diffuse-type gastric cancer," *Oncology*, vol. 65, no. 4, pp. 363–370, 2003. doi:10.1159/000074650

[21] M. Huang *et al.*, "TGF-β1-activated cancer-associated fibroblasts promote breast cancer invasion, metastasis and epithelial-mesenchymal transition by autophagy or overexpression of FAP-α," *Biochemical Pharmacology*, vol. 188, p. 114527, 2021. doi:10.1016/j.bcp.2021.114527

[22] A. Malki *et al.*, "Molecular mechanisms of colon cancer progression and metastasis: Recent insights and advancements," *International Journal of Molecular Sciences*, vol. 22, no. 1, p. 130, 2020. doi:10.3390/ijms22010130

[23] H. Yamagishi, H. Kuroda, Y. Imai, and H. Hiraishi, "Molecular pathogenesis of sporadic colorectal cancers," *Chinese Journal of Cancer*, vol. 35, no. 1, 2016. doi:10.1186/s40880-015-0066-y

[24] L. R. Henry *et al.*, "Clinical implications of fibroblast activation protein in patients with colon cancer," *Clinical Cancer Research*, vol. 13, no. 6, pp. 1736–1741, 2007. doi:10.1158/1078-0432.ccr-06-1746

[25] I. Aier, R. Semwal, A. Sharma, and P. K. Varadwaj, "A systematic assessment of statistics, risk factors, and underlying features involved in pancreatic cancer," *Cancer Epidemiology*, vol. 58, pp. 104–110, 2019. doi:10.1016/j.canep.2018.12.001

[26] J. D. Mizrahi, R. Surana, J. W. Valle, and R. T. Shroff, "Pancreatic cancer," *The Lancet*, vol. 395, no. 10242, pp. 2008–2020, 2020. doi:10.1016/s0140-6736(20)30974-0

[27] P. A. Muller and K. H. Vousden, “P53 mutations in cancer,” *Nature Cell Biology*, vol. 15, no. 1, pp. 2–8, 2013. doi:10.1038/ncb2641

[28] A. Blackford *et al.*, “*smad4* gene mutations are associated with poor prognosis in pancreatic cancer,” *Clinical Cancer Research*, vol. 15, no. 14, pp. 4674–4679, 2009. doi:10.1158/1078-0432.ccr-09-0227

[29] E. S. Calhoun *et al.*, “BRAF and FBXW7 (CDC4, FBW7, AGO, SEL10) mutations in distinct subsets of pancreatic cancer,” *The American Journal of Pathology*, vol. 163, no. 4, pp. 1255–1260, 2003. doi:10.1016/s0002-9440(10)63485-2

[30] I. Aier, R. Semwal, A. Sharma, and P. K. Varadwaj, “A systematic assessment of statistics, risk factors, and underlying features involved in pancreatic cancer,” *Cancer Epidemiology*, vol. 58, pp. 104–110, 2019. doi:10.1016/j.canep.2018.12.001

[31] Y. Liu *et al.*, “Fibroblast activation protein targeted therapy using [177lu]fapi-46 compared with [225AC]FAPI-46 in a pancreatic cancer model,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 49, no. 3, pp. 871–880, 2021. doi:10.1007/s00259-021-05554-2

[32] S. J. Cohen *et al.*, “Fibroblast activation protein and its relationship to clinical outcome in pancreatic adenocarcinoma,” *Pancreas*, vol. 37, no. 2, pp. 154–158, 2008. doi:10.1097/mpa.0b013e31816618ce

[33] D. PATSOURAS *et al.*, “Fibroblast activation protein and its prognostic significance in correlation with vascular endothelial growth factor in pancreatic adenocarcinoma,” *Molecular Medicine Reports*, vol. 11, no. 6, pp. 4585–4590, 2015. doi:10.3892/mmr.2015.3259

[34] M. Shi, “Expression of fibroblast activation protein in human pancreatic adenocarcinoma and its clinicopathological significance,” *World Journal of Gastroenterology*, vol. 18, no. 8, p. 840, 2012. doi:10.3748/wjg.v18.i8.840

[35] A. Lo *et al.*, “Fibroblast activation protein augments progression and metastasis of pancreatic ductal adenocarcinoma,” *JCI Insight*, vol. 2, no. 19, 2017. doi:10.1172/jci.insight.92232

[36] T. Kawase *et al.*, “Fibroblast activation protein-α-expressing fibroblasts promote the progression of pancreatic ductal adenocarcinoma,” *BMC Gastroenterology*, vol. 15, no. 1, 2015. doi:10.1186/s12876-015-0340-0

[37] Rivera C. Essentials of oral cancer. Int J Clin Exp Pathol. 2015 Sep 1;8(9):11884-94. PMID: 26617944; PMCID: PMC4637760

[38] S. Gandini *et al.*, “Tobacco smoking and cancer: A meta-analysis,” *International Journal of Cancer*, vol. 122, no. 1, pp. 155–164, 2007. doi:10.1002/ijc.23033

[39] M. Marron *et al.*, “Cessation of alcohol drinking, tobacco smoking and the reversal of head and neck cancer risk,” *International Journal of Epidemiology*, vol. 39, no. 1, pp. 182–196, 2009. doi:10.1093/ije/dyp291

[40] J. Reidy, E. McHugh, and L. F. A. Stassen, “A review of the relationship between alcohol and oral cancer,” *The Surgeon*, vol. 9, no. 5, pp. 278–283, 2011. doi:10.1016/j.surge.2011.01.010

[41] N. L. Rhodus, V. Ho, C. S. Miller, S. Myers, and F. Ondrey, “NF-KB dependent cytokine levels in saliva of patients with oral preneoplastic lesions and oral squamous cell carcinoma,” *Cancer Detection and Prevention*, vol. 29, no. 1, pp. 42–45, 2005. doi:10.1016/j.cdp.2004.10.003

[42] M. A. St. John *et al.*, “Interleukin 6 and interleukin 8 as potential biomarkers for oral cavity and oropharyngeal squamous cell carcinoma,” *Archives of Otolaryngology–Head & Neck Surgery*, vol. 130, no. 8, p. 929, 2004. doi:10.1001/archotol.130.8.929

[43] R. Nagler, G. Bahar, T. Shpitzer, and R. Feinmesser, “Concomitant analysis of salivary tumor markers—a new diagnostic tool for oral cancer,” *Clinical Cancer Research*, vol. 12, no. 13, pp. 3979–3984, 2006. doi:10.1158/1078-0432.ccr-05-2412

[44] S. Hu *et al.*, “Salivary proteomics for Oral Cancer Biomarker Discovery,” *Clinical Cancer Research*, vol. 14, no. 19, pp. 6246–6252, 2008. doi:10.1158/1078-0432.ccr-07-5037

[45] K. Jessie *et al.*, "Aberrant proteins in the saliva of patients with oral squamous cell carcinoma," *ELECTROPHORESIS*, vol. 34, no. 17, pp. 2495–2502, 2013. doi:10.1002/elps.201300107

[46] H. Wang *et al.*, "Downregulation of FAP suppresses cell proliferation and metastasis through PTEN/PI3K/Akt and Ras-Erk signaling in oral squamous cell carcinoma," *Cell Death & Disease*, vol. 5, no. 4, 2014. doi:10.1038/cddis.2014.122

[47] K. Räsänen, I. Virtanen, P. Salmenperä, R. Grenman, and A. Vaheri, "Differences in the nemosis response of normal and cancer-associated fibroblasts from patients with oral squamous cell carcinoma," *PLoS ONE*, vol. 4, no. 9, 2009. doi:10.1371/journal.pone.0006879

[48] P. A. Cohen, A. Jhingran, A. Oaknin, and L. Denny, "Cervical cancer," *The Lancet*, vol. 393, no. 10167, pp. 169–182, 2019. doi:10.1016/s0140-6736(18)32470-x

[49] [1] S. D. Balasubramaniam, V. Balakrishnan, C. E. Oon, and G. Kaur, "Key molecular events in cervical cancer development," *Medicina*, vol. 55, no. 7, p. 384, 2019. doi:10.3390/medicina55070384

[50] Jin X, Iwasa S, Okada K, Mitsumata M, Ooi A. Expression patterns of seprase, a membrane serine protease, in cervical carcinoma and cervical intraepithelial neoplasm. Anticancer Res. 2003 Jul-Aug;23(4):3195-8. Erratum in: Anticancer Res. 2003;23:5371-2. PMID: 12926053

[51] C. Stewart, C. Ralyea, and S. Lockwood, "Ovarian cancer: An integrated review," *Seminars in Oncology Nursing*, vol. 35, no. 2, pp. 151–156, 2019. doi:10.1016/j.soncn.2019.02.001

[52] S. Lheureux, C. Gourley, I. Vergote, and A. M. Oza, "Epithelial ovarian cancer," *The Lancet*, vol. 393, no. 10177, pp. 1240–1253, 2019. doi:10.1016/s0140-6736(18)32552-2

[53] D. Huber, S. Seitz, K. Kast, G. Emons, and O. Ortmann, "Hormone replacement therapy in BRCA mutation carriers and risk of ovarian, endometrial, and breast cancer: A systematic review," *Journal of Cancer Research and Clinical Oncology*, vol. 147, no. 7, pp. 2035–2045, 2021. doi:10.1007/s00432-021-03629-z

[54] C. Stewart, C. Ralyea, and S. Lockwood, "Ovarian cancer: An integrated review," *Seminars in Oncology Nursing*, vol. 35, no. 2, pp. 151–156, 2019. doi:10.1016/j.soncn.2019.02.001

[55] K. Lagerstedt Robinson *et al.*, "Lynch syndrome (hereditary nonpolyposis colorectal cancer) diagnostics," *JNCI Journal of the National Cancer Institute*, vol. 99, no. 4, pp. 291–299, 2007. doi:10.1093/jnci/djk051

[56] J. L. Walker *et al.*, "Society of Gynecologic Oncology Recommendations for the prevention of ovarian cancer," *Cancer*, vol. 121, no. 13, pp. 2108–2120, 2015. doi:10.1002/cncr.29321

[57] X. Huang *et al.*, "Association between dietary fiber intake and risk of ovarian cancer: A meta-analysis of observational studies," *Journal of International Medical Research*, vol. 46, no. 10, pp. 3995–4005, 2018. doi:10.1177/0300060518792801

[58] P. Mhawech-Fauceglia *et al.*, "Stromal expression of fibroblast activation protein alpha (FAP) predicts platinum resistance and shorter recurrence in patients with epithelial ovarian cancer," *Cancer Microenvironment*, vol. 8, no. 1, pp. 23–31, 2014. doi:10.1007/s12307-014-0153-7

[59] M. ZHANG, L. XU, X. WANG, B. SUN, and J. DING, "Expression levels of seprase/fapα and DPPIV/CD26 in epithelial ovarian carcinoma," *Oncology Letters*, vol. 10, no. 1, pp. 34–42, 2015. doi:10.3892/ol.2015.3151

[60] A. Kennedy, H. Dong, D. Chen, and W.-T. Chen, "Elevation of Seprase expression and promotion of an invasive phenotype by collagenous matrices in ovarian tumor cells," *International Journal of Cancer*, vol. 124, no. 1, pp. 27–35, 2009. doi:10.1002/ijc.23871

[61] M. ZHANG, Y. QIAO, J. M. Nesland, and Z. SUO, "Expression of seprase in effusions from patients with epithelial ovarian carcinoma," *Chinese Medical Journal*, vol. 120, no. 8, pp. 663–668, 2007. doi:10.1097/00029330-200704020-00010

[62] "Astrocytoma tumors," AANS, https://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Astrocytoma-Tumors (accessed May 15, 2023).

[63] J. L. Sowers, K. M. Johnson, C. Conrad, J. T. Patterson, and L. C. Sowers, "The role of inflammation in brain cancer," *Advances in Experimental Medicine and Biology*, pp. 75–105, 2014. doi:10.1007/978-3-0348-0837-8_4

[64] P. Busek *et al.*, "Fibroblast activation protein alpha is expressed by transformed and stromal cells and is associated with mesenchymal features in glioblastoma," *Tumor Biology*, vol. 37, no. 10, pp. 13961–13971, 2016. doi:10.1007/s13277-016-5274-9

[65] S. A. Mikheeva *et al.*, "TWIST1 promotes invasion through mesenchymal change in human glioblastoma," *Molecular Cancer*, vol. 9, no. 1, 2010. doi:10.1186/1476-4598-9-194

[66] Yang P, Luo Q, Wang X, Fang Q, Fu Z, Li J, Lai Y, Chen X, Xu X, Peng X, Hu K, Nie X, Liu S, Zhang J, Li J, Shen C, Gu Y, Liu J, Chen J, Zhong N, Su J. Comprehensive Analysis of Fibroblast Activation Protein Expression in Interstitial Lung Diseases. Am J Respir Crit Care Med. 2023 Jan 15;207(2):160-172. doi: 10.1164/rccm.202110-2414OC. PMID: 35984444; PMCID: PMC9893314.

[67] [1] Z. Wang *et al.*, "Role and mechanism of fibroblast-activated protein-α expression on the surface of fibroblast-like synoviocytes in rheumatoid arthritis," *Frontiers in Immunology*, vol. 14, 2023. doi:10.3389/fimmu.2023.1135384

[68] J. M. Milner *et al.*, *Arthritis Research & Therapy*, vol. 8, no. 1, 2006. doi:10.1186/ar1877

[69] R. Micheletti and M. Alexanian, "Transcriptional plasticity of fibroblasts in heart disease," *Biochemical Society Transactions*, vol. 50, no. 5, pp. 1247–1255, 2022. doi:10.1042/bst20210864

[70] C. E. Brokopp *et al.*, "Fibroblast activation protein is induced by inflammation and degrades type I collagen in thin-cap fibroatheromata," *European Heart Journal*, vol. 32, no. 21, pp. 2713–2722, 2011. doi:10.1093/eurheartj/ehq519

[71] E. Puré and R. Blomberg, "Pro-tumorigenic roles of fibroblast activation protein in cancer: Back to the basics," *Oncogene*, vol. 37, no. 32, pp. 4343–4357, 2018. doi:10.1038/s41388-018-0275-3

[72] M. D. GORRELL, “Dipeptidyl peptidase IV and related enzymes in cell biology and liver disorders,” *Clinical Science*, vol. 108, no. 4, pp. 277–292, 2005. doi:10.1042/cs20040302

[73] D. M. Yu *et al.*, “The dipeptidyl peptidase IV family in cancer and Cell Biology,” *FEBS Journal*, vol. 277, no. 5, pp. 1126–1144, 2010. doi:10.1111/j.1742-4658.2009.07526.x

[74] E. Puré and R. Blomberg, “Pro-tumorigenic roles of fibroblast activation protein in cancer: Back to the basics,” *Oncogene*, vol. 37, no. 32, pp. 4343–4357, 2018. doi:10.1038/s41388-018-0275-3

[75] K. Aertgeerts *et al.*, “Structural and kinetic analysis of the substrate specificity of human fibroblast activation protein α,” *Journal of Biological Chemistry*, vol. 280, no. 20, pp. 19441–19444, 2005. doi:10.1074/jbc.c500092200

[76] Z. Wang *et al.*, “Role and mechanism of fibroblast-activated protein-α expression on the surface of fibroblast-like synoviocytes in rheumatoid arthritis,” *Frontiers in Immunology*, vol. 14, 2023. doi:10.3389/fimmu.2023.1135384

[77] E. J. Hamson, F. M. Keane, S. Tholen, O. Schilling, and M. D. Gorrell, “Understanding fibroblast activation protein (FAP): Substrates, activities, expression and targeting for cancer therapy,” *PROTEOMICS - Clinical Applications*, vol. 8, no. 5–6, pp. 454–463, 2014. doi:10.1002/prca.201300095

[78] J. A. DiMasi, R. W. Hansen, and H. G. Grabowski, “The price of innovation: New estimates of drug development costs,” *Journal of Health Economics*, vol. 22, no. 2, pp. 151–185, 2003. doi:10.1016/s0167-6296(02)00126-1

[79] C. M. Song, S. J. Lim, and J. C. Tong, “Recent advances in computer-aided drug design,” *Briefings in Bioinformatics*, vol. 10, no. 5, pp. 579–591, 2009. doi:10.1093/bib/bbp023

[80] S. M. Paul *et al.*, “How to improve R&D productivity: The pharmaceutical industry’s Grand Challenge,” *Nature Reviews Drug Discovery*, vol. 9, no. 3, pp. 203–214, 2010. doi:10.1038/nrd3078

[81] S. Grinter and X. Zou, “Challenges, applications, and recent advances of protein-ligand docking in structure-based drug design,” *Molecules*, vol. 19, no. 7, pp. 10150–10176, 2014. doi:10.3390/molecules190710150

[82] M. Batool, B. Ahmad, and S. Choi, “A structure-based drug discovery paradigm,” *International Journal of Molecular Sciences*, vol. 20, no. 11, p. 2783, 2019. doi:10.3390/ijms20112783

[83] S. Kalyaanamoorthy and Y.-P. P. Chen, “Structure-based drug design to augment hit Discovery,” *Drug Discovery Today*, vol. 16, no. 17–18, pp. 831–839, 2011. doi:10.1016/j.drudis.2011.07.006

[84] S. J. Macalino, V. Gosu, S. Hong, and S. Choi, “Role of computer-aided drug design in modern drug discovery,” *Archives of Pharmacal Research*, vol. 38, no. 9, pp. 1686–1701, 2015. doi:10.1007/s12272-015-0640-5

[85] G. M. Morris and M. Lim-Wilby, “Molecular docking,” *Methods in Molecular Biology*, pp. 365–382, 2008. doi:10.1007/978-1-59745-177-2_19

[86] R. Jakhar, M. Dangi, A. Khichi, and A. K. Chhillar, “Relevance of Molecular Docking Studies in drug designing,” *Current Bioinformatics*, vol. 15, no. 4, pp. 270–278, 2020. doi:10.2174/1574893615666191219094216

[87] B. K. Shoichet, S. L. McGovern, B. Wei, and J. J. Irwin, “Lead discovery using molecular docking,” *Current Opinion in Chemical Biology*, vol. 6, no. 4, pp. 439–446, 2002. doi:10.1016/s1367-5931(02)00339-3

[88] A. S. Abdelsattar, A. Dawoud, and M. A. Helal, “Interaction of nanoparticles with biological macromolecules: A review of molecular docking studies,” *Nanotoxicology*, vol. 15, no. 1, pp. 66–95, 2020. doi:10.1080/17435390.2020.1842537

[89] I. A. Guedes, C. S. de Magalhães, and L. E. Dardenne, “Receptor–ligand molecular docking,” *Biophysical Reviews*, vol. 6, no. 1, pp. 75–87, 2013. doi:10.1007/s12551-013-0130-2

[90] E. Yuriev and P. A. Ramsland, “Latest developments in Molecular Docking: 2010-2011 in Review,” *Journal of Molecular Recognition*, vol. 26, no. 5, pp. 215–239, 2013. doi:10.1002/jmr.2266

[91] K. L. Damm-Ganamet, R. D. Smith, J. B. Dunbar, J. A. Stuckey, and H. A. Carlson, “Csar benchmark exercise 2011–2012: Evaluation of results from docking and relative ranking of blinded Congeneric series,” *Journal of Chemical Information and Modeling*, vol. 53, no. 8, pp. 1853–1870, 2013. doi:10.1021/ci400025f

[92] S.-Y. Huang, S. Z. Grinter, and X. Zou, “Scoring functions and their evaluation methods for protein–ligand docking: Recent advances and Future Directions,” *Physical Chemistry Chemical Physics*, vol. 12, no. 40, p. 12899, 2010. doi:10.1039/c0cp00151a

[93] W. de Azevedo Jr. and R. Dias, “Computational methods for calculation of ligand-binding affinity,” *Current Drug Targets*, vol. 9, no. 12, pp. 1031–1039, 2008. doi:10.2174/138945008786949405

[94] H. F. 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,” *Journal of Medicinal Chemistry*, vol. 48, no. 20, pp. 6296–6303, 2005. doi:10.1021/jm050436v

[95] T.-S. Lee and P. A. Kollman, “Theoretical studies suggest a new antifolate as a more potent inhibitor of thymidylate synthase,” *Journal of the American Chemical Society*, vol. 122, no. 18, pp. 4385–4393, 2000. doi:10.1021/ja9925554

[96] N. Okimoto *et al.*, “Molecular dynamics study of HIV-1 protease−substrate complex: roles of the water molecules at the loop structures of the active site,” *Journal of the American Chemical Society*, vol. 122, no. 23, pp. 5613–5622, 2000. doi:10.1021/ja9929178

[97] S. J. Marrink, D. P. Tieleman, and A. E. Mark, “Molecular dynamics simulation of the kinetics of spontaneous Micelle Formation,” *The Journal of Physical Chemistry B*, vol. 104, no. 51, pp. 12165–12173, 2000. doi:10.1021/jp001898h

[98] T. Hansson, C. Oostenbrink, and W. van Gunsteren, “Molecular dynamics simulations,” *Current Opinion in Structural Biology*, vol. 12, no. 2, pp. 190–196, 2002. doi:10.1016/s0959-440x(02)00308-1

[99] T. Hansson, C. Oostenbrink, and W. van Gunsteren, “Molecular dynamics simulations,” *Current Opinion in Structural Biology*, vol. 12, no. 2, pp. 190–196, 2002. doi:10.1016/s0959-440x(02)00308-1

[100] M. Karplus and J. A. McCammon, “Molecular dynamics simulations of biomolecules,” *Nature Structural Biology*, vol. 9, no. 9, pp. 646–652, 2002. doi:10.1038/nsb0902-646

[101] X. Lin, X. Li, and X. Lin, “A review on applications of computational methods in drug screening and Design,” *Molecules*, vol. 25, no. 6, p. 1375, 2020. doi:10.3390/molecules25061375

[102] Ballehaninna UK, Chamberlain RS. The clinical utility of serum CA 19-9 in the diagnosis, prognosis and management of pancreatic adenocarcinoma: An evidence-based appraisal. J Gastrointest Oncol. 2012 Jun;3(2):105-19. doi: 10.3978/j.issn.2078-6891.2011.021. PMID: 22811878; PMCID: PMC3397644.

[103] R. P. Vivek-Ananth, A. K. Sahoo, K. Kumaravel, K. Mohanraj, and A. Samal, “MeFSAT: A curated natural product database specific to secondary metabolites of medicinal fungi,” *RSC Advances*, vol. 11, no. 5, pp. 2596–2607, 2021. doi:10.1039/d0ra10322e