Politecnico di Torino

Master's Degree in Biomedical Engineering

A.y. 2024/2025

Graduation Session December 2025

Design of novel inhibitors of the FABP12 protein for drug development against advanced prostate cancer

Supervisor: Candidate:

Jack Tuszyński   Gabriele De Rosa

Abstract

Prostate cancer is the second most common type of cancer worldwide, and the fifth leading cause of cancer death. New detection strategies have increased incidence rates in the last twenty years, and innovative therapeutic and risk-assessed strategies have drastically improved the 5-year survival expectancy of patients. However, advanced prostate cancer caused by the metastasization of aberrant cancer cells is invariably fatal. This is thanks to the adaptive survival strategies of the tumor itself, which can become resistant to standard androgen-deprivation techniques and chemotherapeutic drugs. Prostate cancer undergoes two radical changes before becoming metastatic: one is the Epithelial-to-Mesenchymal Transition (EMT), which is the main hallmark of the start of the invasion of distant tissues and organs; the other is a change in the metabolism of the cancer cells themselves, which start relying on the beta oxidation of fatty acids instead of consumption of glucose for their metabolic needs.

These two events are intricately tied together and are correlated to aberrant expression of Fatty Acid Transport Proteins (FABPs) in the surrounding tissues. FABPs are integral to the transport of fatty acids and other lipophilic substances into the cell membrane, and as such represent a source of energy for lipid metabolism of all types of cells in the body. FABP12, the most recently discovered member of the FABP protein family, plays a part in the metabolism of prostate cells, among other FABPs, but has been under the spotlight as the possible main drive behind aberrant metabolism and EMT in advanced prostate cancer.

While potential inhibitors are known for the other FABPs, none of them, except for one, have been tested in-vitro for FABP12 inhibition. On top of that, no known FABP inhibitor has passed clinical trials for commercial drug usage. Finding potential inhibitors of FABP12 could be of great use in the research for therapeutic strategies against advanced prostate cancer: if the aberrant FABP12 expression is regulated by chemical compounds, then the uncontrolled expansion of the prostatic tumor could be slowed down.

The research conducted in this thesis analyzes currently known inhibitors for all FABPs and uses in-silico computational methods to evaluate whether any of these could be a potential FABP12 inhibitor, thanks to molecular docking. Out of 816 investigated compounds, 12 of these are commercially available and possess a higher binding affinity for FABP12 than chemical compound BMS309403, a pan-inhibitor of the FABP protein family, which has historically been used as a comparative metric for both in-vitro and in-silico experiments. The compounds are also analyzed under the lens of ADMET properties, which highlight whether they have potential use as commercial drugs.

Acknowledgements

"It's hard to believe that it's over, isn't it? Funny how we get attached to the struggle. Promise me you'll take care of yourself, ok?"

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Epidemiology of Prostate Cancer

1.1.1 Incidence, mortality and survival

Prostate cancer (PCa) is the second most common type of cancer worldwide, and the fifth leading cause of cancer death [1], [2]. According to the International Association of Cancer Registries (IARC)'s GLOBOCAN website, which compiles data on all cancer types and sites in 185 countries, PCa accounted for 14.2% of male cancers in 2022 [3].

PCa is the most common cancer in 118 out of 185 countries, with its age-standardized incidence rate being highest in Northern and Central Europe, New Zealand, Australia, North and South America, and South Africa, while it is lowest in South-East Asia and Northern Africa (**Figure 1.1**) [3].

The disparity found in the incidence rate between these regions of the world can be explained by analyzing the epidemiology and risk factors associated with PCa, but also partly attributed to how each country deals with disease screening and recommendations given out to their population [4]. Since many cancer cases are discovered following a prostate-specific antigen (PSA) test, countries with developed healthcare systems and where PSA screening is more widely adopted register a higher incidence rate.

GLOBOCAN predicts that in 2045 the annual cases of PCa will increase by 64.9% compared to an increase of 60.3% for all cancers; the highest trend will be registered in Africa (+131.6%) and Asia (+81.2%) as opposed to Europe (+29.7%) [5]. This can be attributed to the introduction of PSA screening in routine testing, better access to health care, and lifestyle choices becoming more homogeneous with Western countries, thus increasing the prevalence of PCa risk factors such as diet, lack of physical activity, obesity, and more [6].

Age-Standardized Rate (World) per 100 000, Incidence, Males, in 2022
Prostate
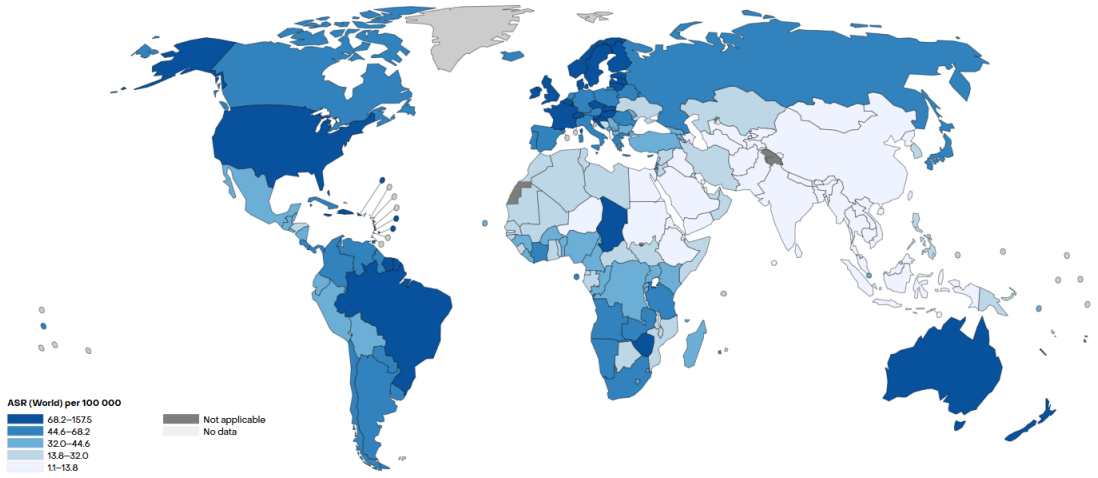
Figure 1.1: Age-Standardized Rate (ASR) for 185 countries, by GLOBOCAN's CANCER TODAY showing the highest incidence in Europe, Australia, and the Americas [3]

In 2045, PCa-associated deaths will more than double (+103.5%) as opposed to an increase of only 70.1% for cancer deaths. Once again, the highest trend will be noticed in developing countries where healthcare resources for early detection of PCa are not accessible to the entire population (Africa, +142.5%; Asia, +124.9%; Europe, +60.6%) [5].

Survival rates for PCa 5 years since detection vary based on the stage of PCa at diagnosis; according to the Surveillance, Epidemiology, and End Results Program (SEER), the survival rate of cases detected in the USA during 2015-2021 for localized and regional stages of PCa is virtually 100%, while distant PCa rates are 37.9%. While comparatively low, the distant PCa rates have been steadily improving since 2004 (30.0%), due to recent improvements in therapeutic drugs and therapies [7]. A similar improvement can be seen across agglomerated 5-year survival cases in Europe, comparing rates for cases discovered in 2000-2004 to cases in 2010-2014 (**Figure 1.2**) [8].

5-year survival rates in developing countries and countries where PSA testing is not as common range from 83.8% in Puerto Rico to 36.4% in China, denoting the importance of early detection and active surveillance even in low-risk cases [9].
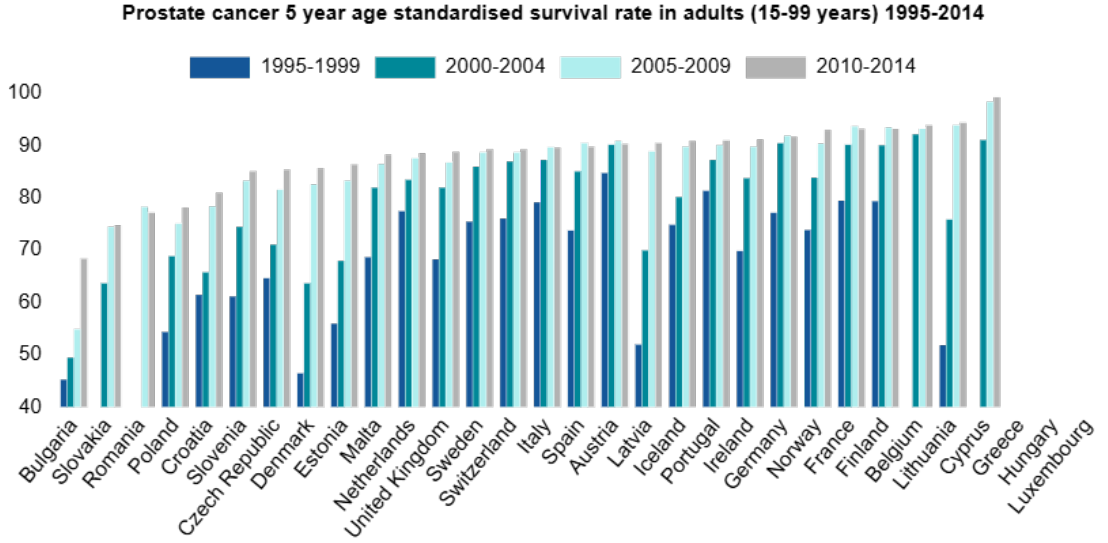
Figure 1.2: Prostate cancer 5-year survival rates in 4 different time periods ranging from 1995 to 2014. In the 2010-2014 timeslot, survival rates ranged from 68.3% in Bulgaria and 94.3% in Lithuania. No data available for Greece, Hungary, and Luxembourg. [8]

1.1.2 Risk factors

USA cancer statistics highlight age as one of the most relevant risk factors for PCa: only 2% of men are at risk of developing invasive PCa from birth to age 59, while the percentage climbs to 5.1% for men aged 60-69 and a 12.5% cumulative chance of being diagnosed. From age 60 onward, the risk of developing invasive PCa is higher than that of any other type of metastatic cancer [10].

Genetics, family history, and race also play an important role. The presence of a first-degree family member who has developed PCa increases the risk of PCa 2.5-fold, and can be further influenced by the grade of the disease, the age of the relative upon diagnosis, and the presence of other cancers in relatives who did not necessarily develop PCa (e.g., breast cancer) [11].

Black African American men are more likely to develop the pathology, twice as much as White Americans, and thrice as much compared to Asian Americans. They are also more likely to be diagnosed with more aggressive PCa; this can be in part explained by the socio-economic differences between ethnicities, since Black Americans are more likely to have reduced access to healthcare [12]. Moreover, studies have proposed that genetic predisposition might also be involved (e.g., African American men).

Several DNA repair pathway genes are associated with PCa risk, with prevalence in metastatic cancer compared to localized [13]. Over 250 single-nucleotide polymorphisms have also been associated with an increased risk of developing PCa [14]. This results in PCa being more genetically heritable than other cancers; more than 50% of an individual's prostate cancer risk is inherited from their parents [13].

Dietary habits also have an effect on PCa risk, demonstrating how the westernization of lifestyle habits can negatively impact the incidence rate of developing countries: consumption of red meat with high-calorie intake of animal fat and of omega-6 fatty acids like linoleic and arachidonic acid can promote lipid metabolism of PCa, favoring growth and cell proliferation [6].

1.2 Prostate Cancer Pathology and Staging

1.2.1 Prostate function

The prostate is an oval gland located at the base of the bladder and surrounding the urethra. It plays a part in both the reproductive and urinary functions of the human body. The glandular tissue of the prostate is made up of acini composed of epithelial cells from which prostatic fluid is secreted.

Most of the glandular tissue of the prostate is located in the peripheral zone, which is the most common site for PCa development. Prostate malignancy primarily involves the epithelial cells found therein and is therefore classified as carcinoma.

There are three main prostatic epithelium cell types: luminal cells, basal cells, and neuroendocrine cells. Laminal cells are the most predominant in the epithelium and are found in the internal surface of the prostatic acini. They possess the capability of secreting proteins such as Prostate Specific Antigen (PSA), which is relevant for predicting PCa development. These cells, unlike the basal and neuroendocrine cells, are androgen-dependent and express a high quantity of Androgen Receptor (AR) [15].

1.2.2 Prostate cancer initiation

Prostate enlargement is commonly associated with increased PCa risk. Prostate enlargement, however, can also happen due to Benign Prostatic Hyperplasia (BPH), which is not considered a precursor of PCa. BPH and PCa both require androgen stimulation for their advancement [16].

Proliferative inflammatory atrophy (PIA) is a pre-cancerous lesion in the prostate gland that involves cell proliferation and inflammation, and can be a precursor to prostatic intraepithelial neoplasia (PIN). PIN is accepted as the earliest stage in cancer onset, especially in its high-grade version. PIN involves the transformation of

epithelial cells in the prostatic acini into malignant cells that proliferate abnormally without invasion of the stroma surrounding the prostate.

Diagnosis of high-grade PIN is included as part of the evaluation of prostate biopsies, as it is not identifiable with transrectal ultrasound (TRUS) and does not elevate PSA serum levels in the blood, which are the two most common diagnostic tools for PCa. High-grade PIN is associated with high predictive values for progression into prostate carcinoma: for example, PIN found during transurethral resection (TURP) done to remove part of the enlarged prostate due to BPH is an important predictive factor for PCa [17].

Though it requires a biopsy to diagnose, PIN can be treated with androgen deprivation techniques or with radiation therapy, similarly to non-metastatic PCa. Given the adverse effects of the therapeutic options, however, it is advised to perform follow-ups every 6 months to one year to evaluate PCa development [18].
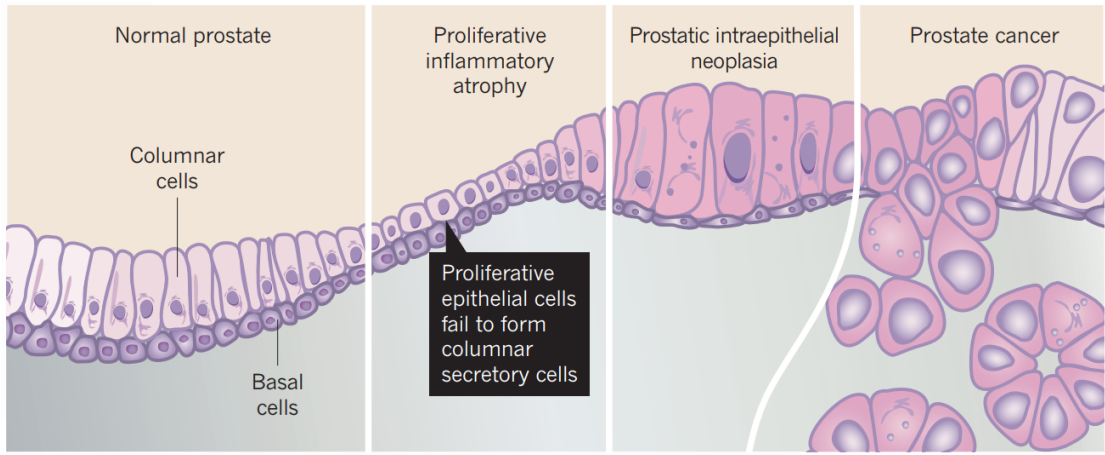
Figure 1.3: Evolution of cell histopathology from normal prostate to prostate cancer. PIA and PIN are intermediary stages that can eventually lead to prostate cancer and have high predictive values for carcinoma progression [19].

1.2.3 Gleason score

The Gleason scoring system is a histopathological grading system that is universally used in PCa staging, prognosis, and treatment planning. Since its inception in 1966 by Donald Gleason, this system has been reviewed and updated a handful of times in order to simplify PCa classification and improve life expectancy, and plan a risk-assessed strategy for cancer treatment.

The most current version of the Gleason score uses a scoring system out of 10, which is further divided into 5 different grades based on the Gleason score (**Figure 1.4**) [20].

The score is calculated by adding the grades of the two most common patterns of cancer cells found in a sample usually obtained from a TRUS-guided biopsy. Metastatic PCa is denoted by a Gleason score >= 8 and a grade of 4 or 5.

The updated Gleason score groups still present some limitations; however, both patient and clinician can have a better understanding of the tumor characteristics and 5-year recurrence progression probability following prostatectomy, where higher grades indicate drastically higher chances of prostate recurrence [21].

Grading Group	Gleason Score	Risk group	Definition	Gleason pattern
1	3 + 3 = 6	Low	Only individual, discrete well-formed glands	
2	3 + 4 = 7	Intermediate favourable	Predominantly well-formed glands with lesser component of poorly formed, fused and/or cribriform glands	
3	4 + 3 = 7	Intermediate unfavourable	Predominantly poorly formed, fused and/or cribriform glands with lesser component of well-formed glands	
4	4 + 4 3 + 5 = 8 5 + 3	High	Only poorly formed, fused and/or cribriform glands Predominantly well-formed glands with lesser component lacking glands Predominantly lacking glands with lesser component of well-formed glands	
5	4 + 5 5 + 4 = 9 5 + 5 = 10	Very High	Lack gland formation (or with necrosis), with or without poorly formed, fused and/or cribriform glands	

Figure 1.4: Gleason grade groups, Gleason scores and Gleason patterns currently used in the diagnostic workflow of prostate cancer [22].

1.2.4 Androgen receptors and prostate carcinogenesis

Androgen receptors are transcription factors involved in the development of the prostate, its growth, and the maintenance of physiological function. ARs can bind to dihydrotestosterone (DHT), which is converted from testosterone produced by Leydig cells in the testis by 5α-reductase. The complex can then translocate into the prostatic cells' nuclei, where binding to Androgen Response Elements (AREs) regulates the expression of target genes.

In the prostate, ARs promote the differentiation of luminal epithelial cells and promote PSA transcription, necessary for prostate function. Just like normal prostatic cells, PCa cells require ARs to proliferate. More specifically, PCa progresses when there is an imbalance in the ratio of cell reproduction rate opposed to the cell death rate, caused by an overabundance of AR [16].

Androgen deprivation therapy has been shown to help in cases of advanced PCa, where it can be obtained via surgical castration of the testis, chemical castration

targeting the synthesis of Luteinizing hormone responsible for testosterone production or, finally, nonsteroidal antiandrogen blockers that reduce androgen signalling in the cell by having a higher affinity for ARs. ADT is not recommended for the early stages of PCa, due to its significant psychological effects and the inevitability of developing castration-resistant prostate cancer (CRPC).

CRPC is invariably fatal, with a life expectancy of 18-24 months at a maximum after the pathology has progressed to this stage. At this point, ADT becomes ineffective for cancer treatment, and PCa is treated with chemotherapeutic drugs like docetaxel for longer survival [23].

The transition of PCa to CRPC is triggered by ADT; however, CRPC can be developed due to different proposed mechanisms: lower threshold of AR androgen activation, AR mutations that render the receptors able to bind non-androgens, ligand-independent AR activation, and AR-independent pathways. The mechanisms surrounding CRPC, though, have not been completely understood; the complexity of cancer progression has limited efforts into finding an effective cure, and will require a deeper understanding through further studies [24].

1.2.5 Metastatic development of prostate cancer

Non-metastatic PCa is contained within the prostatic ducts, where the epithelium separates the stroma from the internal organ. Once cell division becomes unconstrained due to mutations in the DNA, the prostate glands are subject to inflammation, and eventually the cancer cells spread throughout the lumen of the duct and start to infiltrate the stroma [25].

Other than the complications related to CRPC, PCa treatment is limited by cancer microenvironment heterogeneity, lack of molecular signatures to identify tumor subtypes, and reliance of PCa cells on fatty acid oxidation for their energy supply rather than standard glucose consumption. When PCa metastasizes, treatment becomes even more complex, most often being reduced to symptomatic treatment to prolong the life expectancy of the patient with no chance of survival [26].

While the exact mechanisms that can lead PCa to become metastatic are unknown, there are two important processes related to PCa development that are worth analyzing: the epithelial to mesenchymal transition (EMT) and the changes in cell metabolism driven by the tumor.

Epithelial to mesenchymal transition EMT has been identified as an important phenomenon required for the prostatic cells to invade the stroma and obtain a mesenchymal phenotype, therefore being able to spread to distant organs via blood and lymph node transport (**Figure 1.5**). The changes in the tumor microenvironment caused by the uncontrolled proliferation of cancer cells can induce EMT, and EMT can, in turn, further modify the microenvironment via stroma invasion [27].
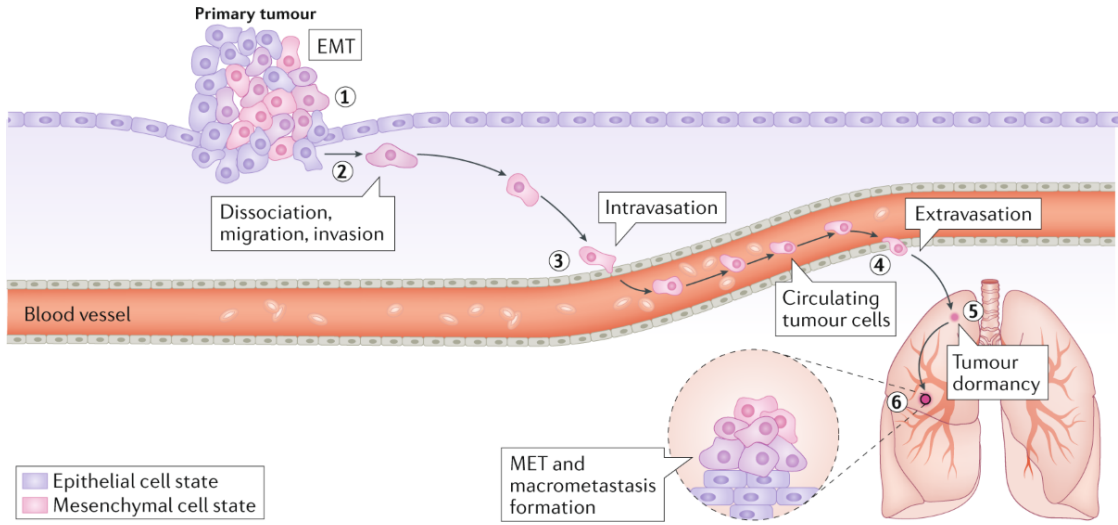
Figure 1.5: Steps of the metastatic process following epithelial to mesenchymal transition. 1) Initial escape from the primary site of motile epithelial cells. 2) Migration into the extracellular matrix. 3) Intravasation in nearby blood vessels. 4) Extravasation from the blood vessels into the parenchyma of distant organs. 5) Proliferation into a secondary tumor. 6) Secondary tumor establishment into a clinically detectable growth. [28]

In PCa cells EMT is first determined by a loss of the expression of E-cadherin in the epithelial cells, which can be suppressed via AR pathways. It is unclear whether the EMT induction is due to an overexpression of AR or due to low AR content due to deprivation therapy, which could indicate ADT as partly responsible for both CRPC and PCa metastases [25]. In fact, AR maintenance is necessary to avoid EMT induction in androgen-independent PCa cells, which could also generate CRPC via alternative pathways.

Estrogen also plays an important role in the development of PCa; its activity is carried out by two types of estrogen receptors (ER), ERα and ERβ [29]. The first one is found in the prostatic stroma and is responsible for the adverse effects of the estrogen response, while ERβ is found within the epithelium and regulates cell proliferation and differentiation and possesses antitumoral properties.

ERβ is also able to promote the degradation of Hypoxia Inducible Factor (HIF), which is characteristic of tumoral microenvironments where cells have limited access to oxygen. HIF induces Vascular Endothelial Growth Factor (VEGF) production, which is a common precursor of angiogenesis. During EMT, and especially in CRPC and high Gleason grade PCa, expression of ERβ is reduced, therefore driving cancer expansion via VEGF and angiogenesis [30].

Fatty acid oxidation   Aberrant lipid metabolism is one of the hallmarks of PCa progression and ADT resistance. PCa cells start to answer their energetic needs by consuming lipids instead of relying on glucose like other cancers. The lipids themselves, given their hydrophobic nature, are transported inside the cell via fatty acid binding proteins (FABPs) [31].

Providing energy to the cell is not the only effect that derives from internalization of fatty acids (FAs); lipids also serve as cell membrane components that exert influence on signal transduction and cell growth properties, for example, by activating peroxisome proliferator-activated receptors (PPARs), which can regulate cancer progression. Not only that, but fatty acid metabolism is correlated to EMT in PCa via the overexpression of FABP12 [32].

The heightened presence of FABP12 in PC3 metastatic cells causes lipid uptake and lipid droplet formation inside the cell, on top of increased energy production from FAs themselves. The aberrant metabolism activates PPARγ receptors, which in turn results in induced expression of EMT transcription factor Slug, loss of E-cadherin, and translocation of β-catenin into the nucleus, promoting EMT via nuclear signaling.

Survivin, which is an inhibitor of cell apoptosis and is known to regulate chemotherapeutic sensitivity of cancer cells, was also found to be a downstream effector of FABP12 [33]. Survivin seems to reside downstream of Slug: it was found that Slug depletion is able to reverse the upregulation of Survivin caused by FABP12. Survivin can cause docetaxel resistance in PCa cells, which is often used as a chemotherapeutic drug for high-risk and metastatic PCa. The development of resistance to docetaxel has proved a challenge, requiring the development of second-generation drugs that are more powerful, albeit more toxic.

The crosstalk between the FABP-derived EMT induction of metastatic cancer and the androgen-AR induction of castration resistance in PCa still needs to be fully understood. While mechanisms are not yet clear, it has been reported that activation of AR is inhibited upon treatment of PPARγ in LNCaP cells [32].

1.3 Prostate Cancer Diagnosis and Treatment

1.3.1 Early detection and biomarkers

PCa has historically been diagnosed by Digital Rectal Examination (DRE) and PSA blood tests, followed by a confirmation exam in the form of TRUS-guided biopsy [34].

PSA is produced by the epithelial cells of the prostate to improve sperm functionality. Elevated levels of PSA in the blood are correlated with an increased risk of PCa; unfortunately, PSA can also be elevated due to benign diseases of the prostate, such as BPH, since it is organ-specific and not disease-specific [35].

PSA was originally used to monitor the progression of PCa in the 1980s, but quickly became a screening tool and a tumor marker to identify the presence of the disease. At the end of the century, in countries that had implemented PSA screening trials, the incidence of PCa drastically increased while mortality decreased [34].

However, the extended use of PSA among the male population was part of controversial opinions regarding the effective prognostic prediction of PCa during the first two decades of the 2000s. PSA testing is, as a matter of fact, associated with a high rate of overdiagnosis and overtreatment [36].

Due to its limitations, other types of prognostic biomarker tests based on PSA have been developed since its first use: age-adjusted PSA accounting for age-related prostate volume changes, free PSA in the blood, PSA density based on TRUS volume obtained of the prostate biopsy, and others. Other tests, like the Prostate Health Index (PHI) score test and the Four–Kallikrein (4k) score test, make use of mathematical formulas combining the other PSA values for more accurate detection [35]. Both types of tests were involved in prospective studies, which indicate that they are more likely to predict clinically significant stages of the disease while reducing the number of biopsies performed for indolent cancers.

These biomarkers, taken by themselves, are still not enough to completely supplant the need for further testing and can still lead to overtreatment, where a more cautious approach would be preferable to avoid the psychological impact of PCa treatment, especially in benign tumors.

1.3.2 Diagnostic tools

Like previously mentioned, TRUS-guided 12-core needle biopsy is the most commonly used technique to diagnose PCa after an elevated PSA is registered or when PCa is suspected. This method, however, carries an intrinsic sampling error, which translates into a high risk of overdiagnosis of indolent cancers and imprecise risk stratification of the therapeutic strategies that should be applied to cure the

pathology [35].

While it is possible to increase the volume obtained by extending the biopsy to up to 50 cores, the accuracy of the diagnosis does not significantly improve, and the procedure becomes more invasive with no real benefit.

Staging of the tumor according to the Gleason scale becomes possible after the biopsy has taken place. By employing fusion biopsy techniques, staging can be improved by ensuring that the biopsy sample does, in fact, accurately reflect the presence of the tumor and its relative Gleason score [37], [38].

Fusion biopsy techniques make use of multiparametric magnetic resonance imaging (mpMRI) to improve the in vivo localization of the tumor while the biopsy is being performed. mpMRI uses multiple sequences to achieve better results, such as diffusion-weighted imaging (DWI), dynamic contrast-enhanced imaging, T1 and T2 weighted imaging, and/or magnetic resonance spectroscopy imaging (MRSI). MRI scans conducted for the detection of PCa are scored based on a scale called Prostate Imaging–Reporting and Data System (PI-RADS), which, similarly to the Gleason Score, indicates the likelihood of PCa [39].

mpMRI is also indicated for local staging after 6 weeks from the confirmation of diagnosis via TRUS-biopsy. MRI is also still recommended in cases where the cancer is suspected to have spread to lymph nodes in medium to high-risk PCa [38]. Both computer tomography (CT) and MRI present limited capability in detecting nodal metastasis; therefore, an additional option is presented by positron emission (PET) combined with CT, which improves staging accuracy and evaluation of PSA recurrence after treatment.

When bone metastases are suspected instead, the classic technique used is technetium bone scintigraphy. Bone biopsy can then be performed to assess and confirm previous suspicions.

1.3.3 Modern diagnostic strategy

With the limitations of PSA screening evidenced by overdiagnosis and overtreatment during the early 2000s in countries where regular PSA testing was recommended, regulatory bodies started advising against routine testing. The claims against screening were due to the reduced quality of life determined by undergoing treatment, especially in healthy men.

The United States Preventive Services Task Force (USPST) initially retracted their recommendation for men above 75 years old in 2008, then expanded the decision to all men in 2012. Whole population screening was recommended against in Germany in 2019 and the UK in 2020, where systematic analyses were conducted. In the USA, it was subsequently found that, between 2008 and 2016, the percentage average of men screened decreased from 61.8% to 50.5% while metastatic PCa

diagnosis increased. In 2018, the USPSTF started to support testing in well-informed men aged 55-69, while the European Association of Urology (EAU) recommends testing if the man is fully informed about the pros and cons of PSA testing [34].

While PSA screening does in fact carry its risks, the increase in tumor aggressiveness recorded during diagnosis in countries that reduced PSA testing indicates that some sort of middle ground is necessary.

The introduction of more advanced biomarkers based on PSA and novel diagnostic techniques like mpMRI has recently given way to risk-adapted strategies based on a combination of early-detection testing, active monitoring, nomograms that take into account patient-related risk factors, and shared decision-making between informed patients and trained physicians (**Figure 1.6**) [4], [34].

These improvements to the modern detection and diagnostic strategies should, with time, prove successful in reducing societal burdens associated with non-opportunistic testing while improving the accuracy of the diagnosis and predicting the correct type of treatment that the patient should undergo.

1.3.4   Treatment

After PCa is diagnosed and staged, clinicians will place the patient in a risk group such that a therapeutic strategy can be formulated based on the risk assessment of the pathology [34], [37], [40].

For low-risk patients, active surveillance is strongly recommended, resulting in PSA testing every 6 months, a DRE examination every year, and the possibility of performing additional biopsies every 1 to 3 years after an mpMRI test to confirm its need.

For medium-risk patients, active surveillance can still be used in the case of patients with unrelated health complications that might affect their life expectancy with a drastic treatment option. Radical prostatectomy, radiation therapy (brachytherapy or MRI-guided external beam radiotherapy), and ADT are the most commonly used strategies at this point. There are also newer techniques, such as High-Intensity Focused Ultrasound and Cryotherapy, which can be used if the tissue has not responded to radiation therapy, though they are not yet recommended as the first treatment option.

For high-risk patients, all therapies previously mentioned can still be used, though the dose will be increased and the size of the surgically removed tissue will be larger. Additionally, combination therapy can be used to obtain synergic effects from the single techniques. Hormone therapy with abiraterone, a second-generation drug, targets adrenal and tumoral androgen production and can also target cancerous cells that have metastasized and/or CRPC. It is also recommended to perform surgical castration since abiraterone does not stop the testis from producing testosterone.
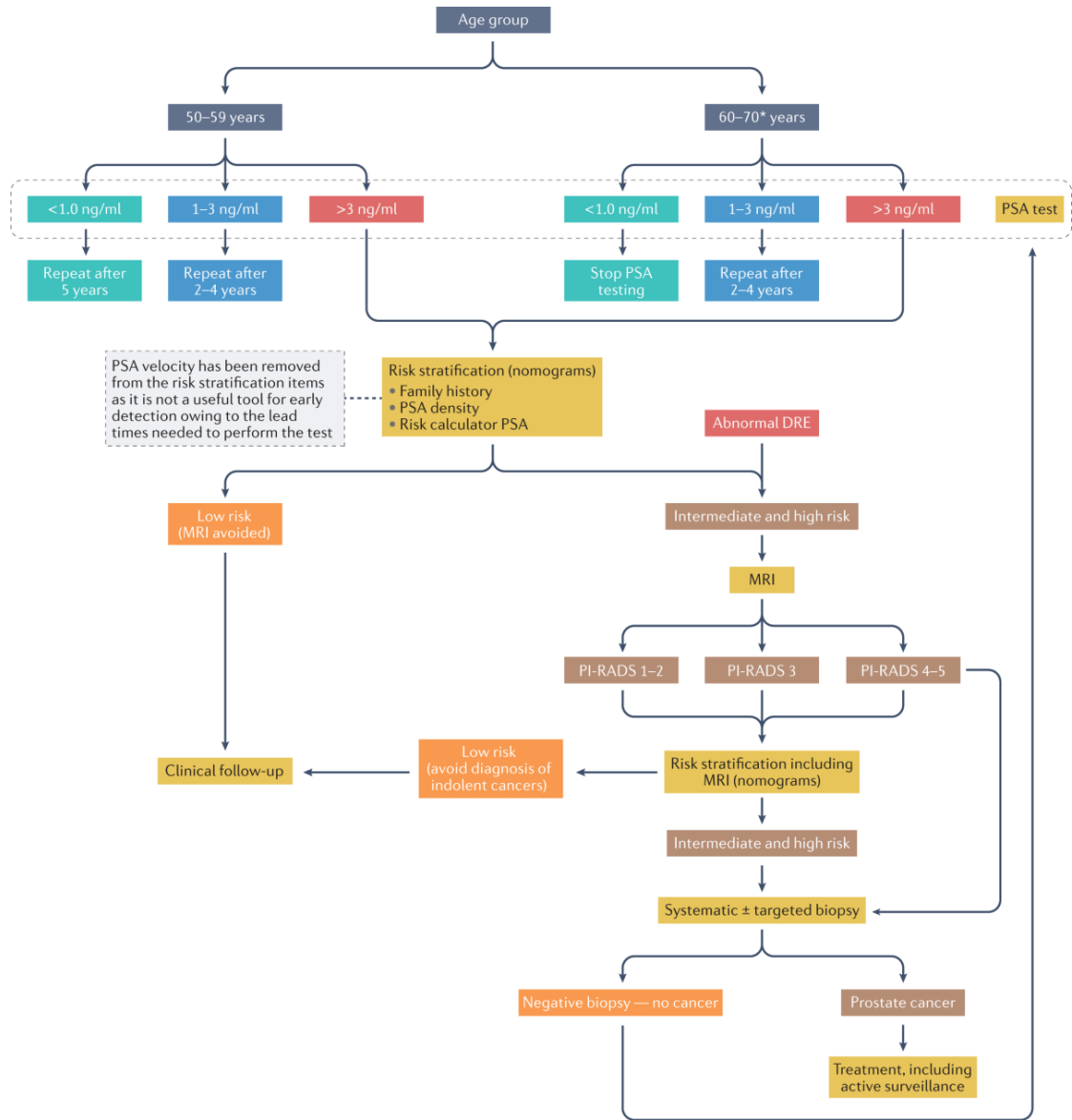
Figure 1.6: Risk-adapted strategy for the early detection of prostate cancer from the EAU. The strategy combines the confirmed effect of PSA testing in patients older than 50 years old and integrates nomograms to improve the successful diagnosis. Patients are then categorized in low and medium-high-risk categories; the former undergo active surveillance while the latter are offered MRI scans of the prostate before proceeding with biopsies. The PI-RADS scores of the MRI determine if the biopsy is needed or if active surveillance is preferable [34].

If cancer has metastasized to distant organs such as the bone, then combination therapy of ADT with radiation therapy, immunotherapy, and chemotherapy (using docetaxel, or second-generation drugs like cabazitaxel) has been shown to not only have an effect on CRPC but also significantly prolong patient survival. Since CRPC and most metastatic cancers cannot be fully cured, therapies are aimed at improving the life expectancy of younger patients and relieving symptoms for those who are older and/or have other health conditions.

1.4 Fatty Acid Binding Proteins (FABPs)

1.4.1 FABPs characteristics and function

Fatty acid binding proteins (FABPs) are a family of small 14-15 kDa, water-soluble, cytosolic FA transporters. 10 isoforms have been identified in humans (FABP1-9, FABP12), with FABP12 being a more recent discovery [41].

The FABP protein family exhibits fairly high amino acid sequence identity among its members, ranging from 30 to 68% [42]. All FABPs share a common tertiary structure made of a beta-barrel with a hydrophobic cavity inside of it that can be filled with water, which is displaced after binding with FAs. The tertiary structure is composed of 10 antiparallel β-sheets and an N-terminal helix-turn-helix moiety, as shown in **Figure 1.7**.

FABPs mediate the translocation of hydrophobic lipids within the cell, where FAs can exert their function on the various organelles inside the cytoplasm, ranging from energy consumption, energy storage (through lipid droplets), external to internal cell signaling, regulation of cell growth, proliferation, and gene expression [43].

FABPs are present in tissues that possess active lipid metabolism, such as the liver, ileum, intestine, heart, skin, brain, testes, and prostate. FABPs are not tissue-specific, though they have an alternative naming convention that is based on the first tissue in which they were discovered. The tissues themselves can express multiple types of FABPs, and cell lines from the same tissue can express FABPs in different amounts [41].

Enhanced FABPs expression has been reported to be associated with cancer development, in adipose tissue, bladder, renal cells, glioblastoma, liver, and prostate cancer [44]. Therefore, depending on the cancer, they might be able to be used as biomarkers for cancer detection.

One of the major risk factors for PCa is fat consumption in dietary habits: high-fat meat, with significant amounts of unsaturated and omega-6 fatty acids, can promote more aggressive PCa. Though the link between dysregulation of lipid metabolism and tumorigenesis is not completely understood, PCa has the characteristic of transitioning to aberrant metabolism based on β-oxidation instead
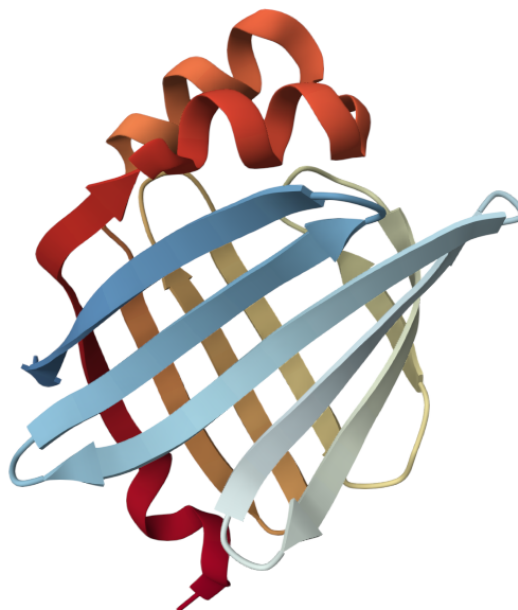
Figure 1.7: FABP4 beta barrel structure, the red end of the molecule is the N-terminus of the protein, while the blue end is the C-terminus. The red-orange section after the first arrow shows the two alpha helices, while the remaining tertiary structure is composed of ten anti-parallel beta sheets forming the barrel. 4NNT PDB structure.

of glucose consumption, unlike other cancers. This causes a cascade of phenomena in epithelial cancerous cells, which ultimately results in their transition to a mesenchymal state and invasion of the prostate gland stroma and blood, thus metastasizing [31], [32].

Targeting PCa cancer cells by using FABPs present in prostatic tissue as both a therapeutic target and a biomarker could provide a measure of prevention and treatment for those cells that have started to rely on fatty acid oxidation and are at risk of undergoing EMT and developing further DNA modifications.

FABP1   FABP1, or Liver-FABP, is the first discovered FABP; it was originally located in hepatocytes but is also highly expressed in the kidney, lung, intestine, and pancreas. Unlike the other members of the FABP family, human FABP1 can bind with two long-chain FAs (LCFAs) [45]. The portal region of the protein, located between the two alpha helices and the two loops connecting β-sheets C-D and β-sheets E-F, is where the protein will initially accept the ligand such that it moves inside of the protein and binds to the main binding pocket by displacing water molecules inside. Once the FA is bound to the canonical site, the portal

region will become able to accommodate another FA.
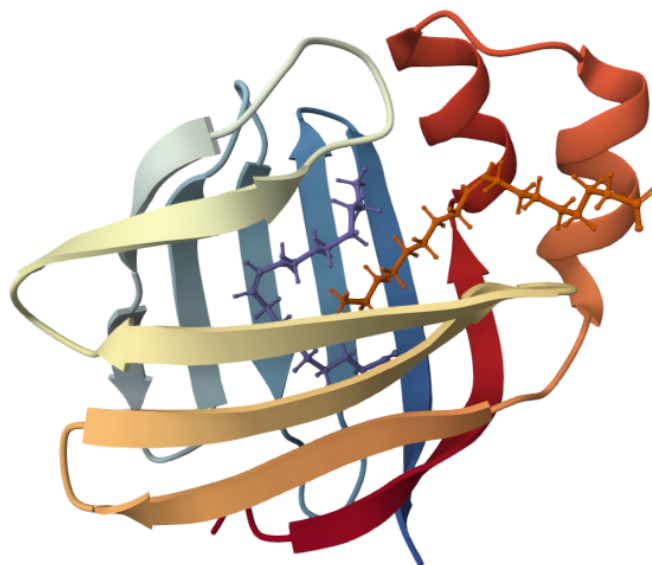
Figure 1.8: FABP1 PDB structure (2LKK) with two molecules of oleic acid bound to it. The purple oleic acid (ball and stick representation) is found in the middle of the beta-barrel structure, while the orange oleic acid is found closer to the portal region near the alpha helices.

FABP 1 is also able to bind to a wider variety of ligands, like cholesterol, bile acids, peroxisome proliferators, endocannabinoids, and more. While FABP3, FABP5, and FABP7 possess cannabinoid transport function in the brain, FABP1 is related to their transport and subsequent metabolism in the liver. The inhibition of FABP1 activity, therefore, prevents the degradation of these ligands, making it more likely for them to accumulate in the brain and exert their antinociceptive functions [46], [47].

FABP1 is also related to non-alcoholic fatty liver disease (NAFLD), characterized by excessive accumulation of fat in the hepatocytes, eventually leading to liver fibrosis. Increased expression of FABP1 has been linked to the development of different types of cancer, like liver, lung, and colorectal cancer [48]. In hepatocellular carcinoma, especially, upregulation of FABP1 is followed by VEGF expression, which helps cell migration and angiogenesis.

FABP2 FABP2, also called Intestinal-FABP, is mostly found in the small intestine, though it is also present in the colon and in the liver. Like FABP1, it has a preference for LCFAs; however it is more expressed in the villi, unlike FABP1, which is more

present in the crypts. Its overexpression is related to abnormal intestinal fat adsorption and transport, leading to atherosclerosis in type-2 diabetes mellitus and villous atrophy in coeliac disease. Polymorphism of FABP2 is, instead, linked to disturbed lipid metabolism, causing hypertriglyceridemia and increased insulin resistance [49].

FABP3 FABP3, or Heart-FABP, is highly expressed in the myocardium and skeletal muscle, but also in a vast variety of other tissues, ranging from the lung, stomach, and adipose tissue to the brain, mammary glands, and ovaries.

FABP3 has been confirmed to be an accurate biomarker for the early diagnosis of acute coronary syndrome, myocardial infarction and is associated with a higher likelihood of developing neurodegenerative diseases like Alzheimer's and Parkinson's due to alterations in signal transduction and neuron membrane structure. FABP3 inhibition is related to endocannabinoid accumulation in the brain, which generates an antinociceptive effect [50].

Upregulation of FABP3 has been correlated with poor prognosis in uveal melanoma tissue, gastric adenocarcinoma, uterine sarcomas, and lung cancer. For breast cancer, however, FABP3 is downregulated; FABP3 is, in fact, a tumor suppressor for breast cancer, where it inhibits the growth of mammary epithelial cells [44].

While the inhibition of FABP3 looks appealing as a therapeutic strategy for many diseases, it has been reported that FABP3-deficient mice can develop cardiac hypertrophy [51]; since FABP4 and FABP5 are the most common therapeutic targets of the FABP family, though for different reasons, developed drugs need to be careful not to possess too much affinity for unwanted FABPs which could cause off-target interactions and risk side effects.

It is important to closely analyze the binding mechanisms of the shared and unshared moieties within the FABPs binding sites, such that drugs can be developed with a rational strategy in mind. For example, by superimposing the 3D structures of FABP3 and FABP4 using molecular visualization tools, it is possible to see how amino acids L115 and L117 in FABP3 have a higher steric hindrance in the binding site as opposed to the respective V115 and C117 amino acids present in FABP4 (**Figure 1.9**). By working around this peculiarity, it is possible to obtain chemical compounds that have a higher affinity for FABP4 compared to FABP3 [52].

FABP4 FABP4 is the most researched protein of the FABP family; because of that reason, the vast majority of chemical compounds that have been developed and studied are based on FABP4 drug design, while the remaining ones are derived from FABP5 compounds. BMS309403 is the most known FABP pan-inhibitor, originally discovered for FABP4 inhibition, and is usually used as a comparative compound for in vitro and in silico testing of new drugs [53].
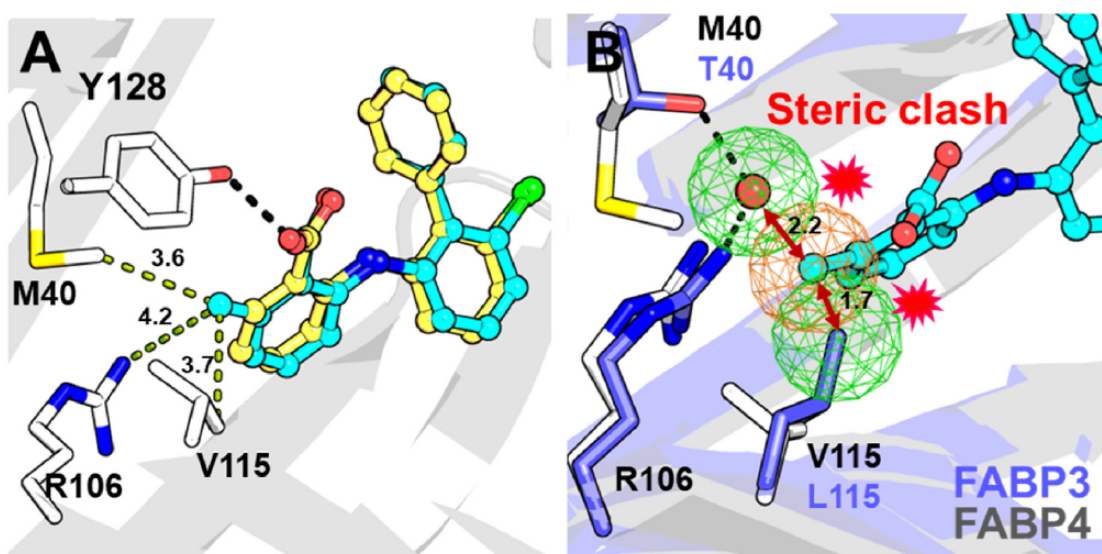
Figure 1.9: A) FABP4 is shown in a standard binding pose with two ligand superposed on top of each other. The one shown in blue can bind residues M40, R106, and V115 thanks to its methyl group protruding towards the FABP amino acids. B) FABP3 and FABP4 residues are superposed on top of each other with the blue compound from (A). The presence of L115 and T40 instead of L115 and M40 in FABP3 generates steric clashes such that the binding with this compound is not favorable [52].

FABP4 is also called Adipose-FABP4, and as the name suggests, it is mostly present in adipose tissue and macrophages. FABP4 can be used as a prognostic biomarker for a variety of pathologies related to metabolism, such as atherosclerosis, NAFLD, and diabetes, but also for acute myocardial infarction. FABP4 possesses a paracrine/endocrine function, meaning it is able to communicate with cells in distant organs after it is released from the adipose cells and influence their metabolism. This effect is particularly evident because FABP4 can influence the progression of cancer in organs where it is normally not found or is present in small quantities [41].

FABP4 is also involved in other cancer development, especially in its exogenous form. In PCa, FABP4 is secreted by periprostatic adipose tissue, and it promotes serum-induced cell invasion in the prostate glands. FABP4 was also found in bone metastases in obese mice, denoting its potential role in disease progression. Other cancers that FABP4 can promote are invasive bladder cancer, ovarian cancer, lung cancer, and breast cancer [54].

FABP5 FABP5 is found in the endothelial cells of many organs, such as skin (from which it derives the name Epidermal-FABP), liver, lung, kidney, brain, mammary, and prostate glands. It is also the most expressed FABP in the human body and is frequently overexpressed in cancer tissues [41].

In colorectal cancer, FABP5 has been linked to a PPARβ/γ-dependent pathway, which enhances cell proliferation and growth following the transportation of signaling factors that enhance the PPAR's transcriptional activity. Moreover, in hepatocarcinoma, FABP5 has been shown to increase tumorigenicity via the induction of EMT.

Chemical inhibition of FABP5 in mouse models with CRPC using compound SBFI26 has been shown to impede the signaling pathway involving PPARγ and VEGF that promotes angiogenesis and cancer progression [55]. SBFI26 was originally designed for analgesic and anti-inflammatory purposes; FABP5, like FABP3 and FABP7, is involved in the transport of endocannabinoids like anandamide to fatty acid amide hydrolase (FAAH), which hydrolyzes them and inhibits their antinociceptive properties. ART26.12, a third-generation drug of the same family as SBFI26, is currently undergoing clinical trials for chemotherapy-induced peripheral neuropathy (CIPN), breast and prostate cancer, anxiety, and psoriasis [56].

FABP6 FABP6, also known as Ileal-FABP, is mostly present in the ileum, where it binds to bile acids and helps digest fatty acids. Similar to FABP4, FABP6 is also present in adipocytes and macrophages, albeit in lower quantities.

FABP6 is involved in colorectal cancer, where its presence can be used as a biomarker for its prognosis and diagnosis. FABP6 overexpression is also related to renal carcinoma, bladder cancer, and ovarian cancer; however, it is not present in PCa and prostate cells [57].

FABP7 FABP7 is called Glial-FABP, due to its almost exclusive presence in glial cells of the brain. The main pathologies in which FABP7 is involved are neurological diseases, Down syndrome, schizophrenia, and cancers such as gliomas and glioblastomas [58]. In the brain, epidermal growth factor receptor (EGFR) activation is necessary for the organ to develop; when the nuclear localization of FABP7 increases, so does the size of the tumor, and its clinical outcome worsens.

FABP8 FABP8 is highly expressed in the central and peripheral nervous system's myelin, from which is takes the name Myelin-FABP [59]. Despite it being decades from its discovery, the role of FABP8 is not yet clear, and so far, there have been no studies that link its expression to tumorigenicity.

FABP9 FABP9 is another poorly understood member of the FABP family; it is called Testis-FABP, where it is mostly expressed and plays a role in germ cell

development and spermatogenesis. Although FABP5 is the most studied FABP in regards to PCa, FABP9 is also present in prostate cells, where its expression is elevated in malignant cell lines [60].

FABP12 FABP12 is the latest member of the FABP protein family to have been discovered [61]; the gene that expresses FABP12 was identified to reside in genomic coordinates 8q21.13, which also includes FABP4, FABP5, FABP8, and FABP9. Not surprisingly, FABP12 shares the same tertiary structure as these proteins and retains high amino acid sequence similarity, which is highest for FABP8 (67%).

1.4.2 FABP12 as a therapeutic target for metastatic PCa

The expression of FABP12 has been confirmed in the retina and in the testis, and it is upregulated in retinoblastoma cancer cells and specifically metastatic PCa cells. In fact, while FABP5 is the FABP with the highest mRNA expression in primary and metastatic PCa, FABP12 is the one that registers the highest amplification when comparing the two populations (24.5x for FABP12, 1.7x for FABP5, analyzed in z-score for mRNA) [31]. For FABP4, FABP8, FABP9, and FABP12, analysis based on mRNA expression measured on disease-free PCa patients indicated that all 4 proteins present a worse prognosis with heightened protein levels. FABP5, instead, shows no prognostic difference between low and high expression.

FABP12 is involved in metastatic PCa progression via the PPARγ-lipid droplet accumulation-EMT induction pathway, which promotes mesenchymal phenotype in reprogrammed prostatic epithelial cells, which are then able to spread to the stroma and invade distant organs via the blood [31]. PPARγ activation is also involved in the EMT-Slug-Survivin pathway, which confers docetaxel resistance to PCa cells, further enhancing their tumorigenic capabilities. Inhibition of FABP12 in PC3 cells transfected with the protein was obtained with the chemical compound BMS309403; results show a significant inhibitor-dependent reduction in cell motility in PC3-FABP12 cells but not PC3 control cells, which significantly express only FABP4 and FABP5 [33].

These findings point to FABP12 being the main FABP responsible for PPARγ activation and subsequent EMT phenomena, therefore driving the metastasization of PCa cells. Information about FABP12 as a therapeutic target is quite lacking, and other than BMS309403, which is a pan-inhibitor for most FABPs, no drug has been tested for FABP12 inhibition.

1.5 Goal of the Thesis

The main goals of this thesis are to gather information regarding possible inhibitors of FABP12, which are yet to be discovered, and, ideally, find possible chemical

compounds that strongly inhibit FABP12 without showing the same preference for the other proteins of the FABP family and other molecular off-targets.

Research for potential ligand candidates started from none other than the other FABPs for which known and tested compounds are available in the literature. It is evident that many of these proteins share common inhibitors, BMS309403 being the best example of them all, being a pan-inhibitor that works exactly the same way in most of the FABPs. As such, the immediate hypothesis is that, among the hundreds of ligands tested for FABPs, some would also inhibit FABP12.

This idea is further validated by the already known inhibitory capability (albeit not too strong) of FABP12 by BMS309403, and by the common tertiary structure shared by the protein family. More specifically, the main binding site of FABPs is always found in the same place and, except for FABP1, is the sole location that is valid for inhibition of the protein and transport of fatty acids in cells.

Furthermore, one peculiarity of the binding site strengthens the bond between FABPs and their transported fatty acids, and is something that researchers in this field have accounted for when trying to discover novel candidates: 7 of the FABPs (FABP12 included) possess an RxY motif (Arginine-X-Tyrosine) in their amino acid sequence that strongly binds with carboxylic groups of fatty acids. All 10 of them retain the arginine residue, while FABP1, FABP2, and FABP6 possess another amino acid in place of the tyrosine and are among the least studied ones as well (**Figure 1.10**).
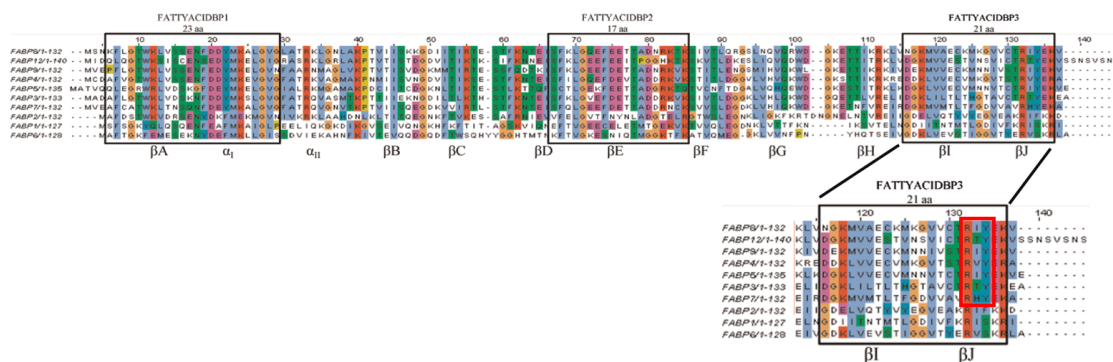
Figure 1.10: Multi Sequence Alignment of FABP protein family, performed on ClustalW2 [42]. FABPs share a three-element fingerprint domain, highlighted by motifs (black boxes). The red outline in BP3 (the third motif) indicates the RxY motif common to most FABPs and important for fatty acid binding.

It is interesting to note that the vast majority of known FABP compounds, BMS309403 and ART26.12 included (**Figure 1.11**), do possess a carboxylic group. The fundamentals of rational drug design teach that keeping functional groups in designed compounds (pharmacophore-based design) while changing other parts of

the molecule to fit in the steric space delimited by the binding site is one of the main ways to develop strong compounds when the target is known and information about the binding mechanisms has been revealed. It follows that researchers would apply this principle and thus generate a sort of inherent bias in the structure of potential ligands.
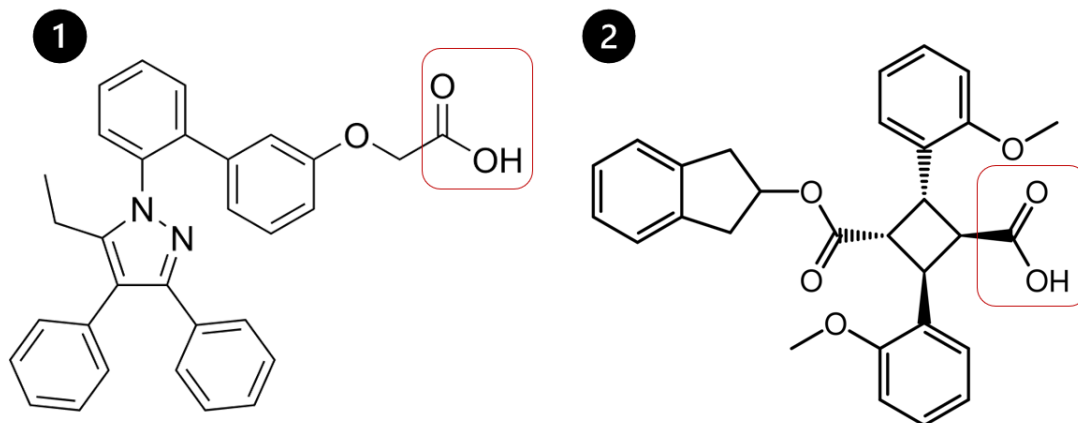
Figure 1.11: 1) BMS309403 and 2) ART26.12 chemical structure. Highlighted with a red box is the carboxylic group of the compounds, where most of the interactions with FABPs happen.

Nonetheless, the exploration of the binding mechanisms of known FABP inhibitors through molecular docking would reveal whether this concept applies to FABP12 as well, and as such, a ligand database was constructed with inhibitors found in the literature to perform computational simulations for this exact purpose. Once the results are validated using in vitro experiments, it will be possible to know which of these chemical compounds work best, and apply rational drug design techniques to further improve their physicochemical characteristics, binding affinity, and selectivity towards FABP12.

Chapter 2

FABPs Structure Preparation

2.1 Materials

2.1.1 Protein structure determination

To perform in silico studies and computational simulations on protein targets, there is one important requirement that must be addressed first: the protein target needs to be represented by a 3D model, which contains various information regarding the protein's tertiary structure, intramolecular bonds, amino acid sequence, and more. The common method used to obtain such a structure is to determine it experimentally by using one of three techniques: X-Ray crystallography, Nuclear Magnetic Resonance (NMR), or Electron Microscopy (EM) [62].

X-Ray crystallography is the most commonly used technique, with upper resolutions of 0.5Å, though under 2.5Åis enough to be considered good. The information obtained with these methods is then stored within a computer file, most often a .pdb file, which takes its name from the Protein Data Bank (PDB), which is a public online database that collects .pdb files of proteins and currently holds more than 240,000 structures [63].

A .pdb file contains 4 main sections [64]: the first one is a header with various annotations which reference back to the internet page where the file was obtained, bibliography regarding who determined the structure and which method was used, and various additional information about the file (if it contains multiple protein sequences or ligands, the organism from which it derives from, mutations, etc.). The second section is a short list that displays the amino acid sequence of the protein (and any additional structures, if present). The third section has coordinate references for all the atoms of the 3D space, which are represented by ATOM if

they belong to the main protein and HETATM if they do not (e.g., water molecules, ligands). The final section, called CONECT, specifies connectivity between atoms for which coordinates are supplied and thus can be used to represent the topology of the structure in 3D space.

For in silico evaluations and simulations, it is best to use a human-derived protein, if it is available. When this is not feasible, either because no one has determined its structure yet or because using the standard techniques is not viable due to their limitations, it is possible to use computational tools to generate the .pdb file instead. There are two main methods to achieve this: computational modelling-based tools (homology modelling) or artificial intelligence-based tools (e.g., AlphaFold).

In both cases, before prediction, it is necessary to obtain a FASTA file of the target protein. This file contains the protein's amino acid sequence represented with a string of single letters based on the residue, and a brief description of the protein's characteristics (name, organism of origin, classification). FASTA files are easily accessible through UniProt, the biggest bioinformatics database for protein sequences of living organisms and viruses [65]. The UniProt website provides a variety of datasets and four main tools.

The most important dataset is UniProt Knowledge Base (UniProtKB), which contains more than 570,000 reviewed entries (proteins) and more than 250 million unreviewed ones. UniProtKB entries include information about the protein's literature citations, taxonomy, subcellular locations, keywords, cross-referenced databases, and related human diseases. The tools that UniProt provides can be used for protein alignment (BLAST, Align), to look up amino acid sequences present in proteins of the database (Peptide search), and finally to reverse search information about a protein for which a FASTA file or an amino acid sequence is available.

The Basic Local Alignment Search Tool (BLAST) suite of tools, in particular, contains a sub-function that is able to align multiple protein amino acid sequences and calculate their sequence identity (BLASTP). This is especially useful to find out which protein regions are shared between a family of proteins and can provide an easy way to select templates for both homology modelling and AlphaFold prediction, based on the highest sequence identity between the target protein and the other members of its family.

2.1.2   Homology Modelling

Homology modelling (HM) is based on the fact that structures of proteins are conserved and are dependent on their amino acid sequence. This means that similar proteins (based on amino acid sequence identity) will fold in similar ways. Various software and web-servers, like Molecular Operating Environment (MOE) [66] and

SWISS-MODEL [67], implement their own version of HM, though the basis of the technique is shared among implementations.

Briefly, HM requires an experimentally determined protein template, which is usually automatically obtained based on the PDB database. For best results, it is recommended to use templates with sequence identity close to or higher than 60%, with which HM can generate models with precision similar to experimental techniques [68]. The template sequence and the target sequence will then need to be aligned, and then the modeler of the chosen computational tool will build the model according to its implementation.

Following model construction, a validation of the model will need to be performed to assess its quality. Various tools are available for this purpose, though the most intuitive one is evaluating the Root Mean Square Deviation (RMSD), which is a measure of the average distance between the atoms of superimposed molecules, in this case template and target. Other useful information can be obtained from the Ramachandran Plot, also known as the Φ-Ψ plot, which checks for the stereochemical quality of a protein by displaying the allowed and disallowed combinations of the dihedral angles of the protein's amino acids. Lastly, the SAVES website [69] contains a series of server-based tools that check for the correct distribution of regions of the protein structure (ERRAT, Verify3D) and other stereochemical properties (WHATCHECK, PROCHECK).

One of the major limitations of HM is the correct modelling of loop regions of proteins, since they are highly variable in terms of composition and run the highest risk of having missing atoms. While there are loop modelers that can be used to repair the structure if the stereochemistry of the protein presents issues, this weakness in the computational modeler has not yet been solved.

2.1.3   AlphaFold

The second computational method that can be used for protein prediction is AlphaFold (AF), which was awarded the Nobel Prize for Chemistry in 2024. AF is an artificial intelligence-based neural network that uses deep learning techniques to predict yet-to-be-determined protein structures. It is based on a hybrid combination of an evolutionary programme, which takes into account the evolutionary history of the proteins themselves, and a physical interaction-based approach, which considers geometric and physical bias based on known structure to place residues and atoms in the 3D space [70].

It first debuted during CASP14 assessment, a biennial competition where recent experimentally determined protein structures, which are not uploaded to the PDB, are used as subjects for the prediction of their structure by various novel computational methods and techniques [71]. AF was able to come in first place by quite a margin: its median backbone accuracy of predicted proteins was

0.96Å(RMSD) against the second-best competitor, which had an accuracy value of 2.80Å.

Its neural network is trained on 170,000 structures from the PDB and can predict protein structures even in the absence of a template; however, it is still recommended that a template is used to improve the prediction, if this is available. AF's performance was benchmarked against a dataset of more than 30000 loop regions, and it was determined to be extremely precise when predicting regions of 10 residues or less (average RMSD of 0.33Å), but still has issues when predicting loops with 20 residues or more (average RMSD of 2.04Å) [72].

The validation pipeline for AF-predicted proteins is the same as the one for HM models. AF itself, however, presents an array of different metrics by which an initial estimation can be performed, which can aid in the evaluation of the quality of the protein. AF associates with every predicted structure a predicted Template Model score (pTM), a predicted Local Distance Difference Test score (pLDDT), and a Predicted Aligned Error graph, which is also available to the user if downloading an AF-predicted .pdb file from PDB.

The pTM score [0, 1] assesses the overall accuracy of the structure of the complex. If multiple chains are present in the complex, then the pTM score will be accompanied by an iPTM score, which evaluates the pairwise modelling accuracy. The pLDDT score [0, 100] is a per-atom estimate of AF's confidence in the prediction, and finally, the PAE graph [Å] shows the confidence of relative positions of two items (atoms, residues, ions) within the predicted structure (**Figure 2.1).**
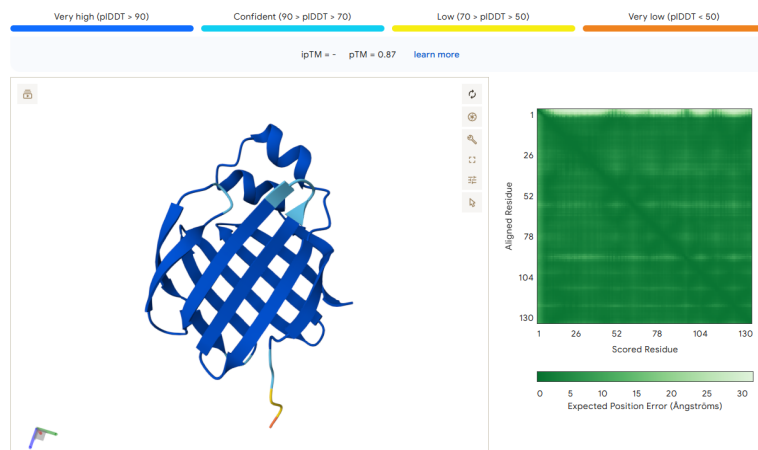
Figure 2.1: FABP12 AlphaFold model generated through the AlphaFold web server. The colored representation of the residues indicates the per-residue plDDT score, while the pTM and Predicted Aligned Error graph are shown at the top and at the right of the structure, respectively.

2.1.4   Molecular Operating Environment

MOE is a state-of-the-art software platform for rational drug design and discovery with many applications built in. MOE allows the user to represent in 3D the .pdf file of one or multiple proteins to perform sequence alignment and HM, superposition of entire or parts of the protein, Molecular Docking simulations with databases of ligands, Molecular Dynamics simulations based on force field applications, structural bioinformatics, structure and ligand-based design, and much more. For this section, MOE was used to perform HM predictions, but also to evaluate the RMSD of models created with both computational methods when compared to their experimentally determined protein, when available (FABPs 1-9).

It is important to note that generating HM and AF-predicted structures of experimentally determined proteins seems counterintuitive. There are two reasons for this not to be the case:

1) While HM possesses a computational ceiling in its accuracy, due to its predictions being based purely on known structures, AF is based on a hybrid programme which also includes evolutionary aspects of the organism's genome. Its accuracy is quite remarkable and has therefore let AF surpass other computational methods in protein prediction, HM included. Not only that, but experimental methods also carry an intrinsic error which rarely falls below 1Å(RMSD), which is around AF's median backbone predicted error (0.96Å). Replacing experimentally determined proteins is not the correct way to perform computational studies (as of yet, at least), but it is also not completely wrong, nor would it produce results that would drastically vary from experimental proteins.

2) The comparison of HM vs AF models for FABPs 1-9 was conducted in a way that could mimic the conditions of current knowledge on FABP12: FABP12 possesses a FASTA sequence and is not experimentally determined; therefore, the best template for its prediction was selected based on sequence identity (FABP8). This same method was used for all other FABPs, making sure to exclude them from the template selection when this was necessary (manual selection of template for HM in MOE, removal of sequence template, if necessary, from Multi Sequence Alignment Json files in the case of AF).

2.1.5   Experimental FABPs and templates

Experimentally determined structures of human FABPs were obtained from the PDB. For all proteins a single structure was selected among those available in the PDB database, according to two parameters: highest structure resolution (if available) and amino acid sequence identity with UniProt's reported data on FABPs (**Table 2.1**).

Table 2.1: List of PDB structures obtained from the PDB for FABP1-9. Where possible, X-ray diffraction with the highest resolution was prioritized, making sure that the protein's amino acid sequence corresponds with the data shown in UniProt.

Protein	PDB ID	Experimental method	Resolution
FABP1	6DO6	solution NMR	N/A
FABP2	1KZX	solution NMR	N/A
FABP3	3WVM	x-ray diffraction	0.88 Å
FABP4	7FWK	x-ray diffraction	0.95 Å
FABP5	7FYO	x-ray diffraction	1.12 Å
FABP6	2MM3	solution NMR	N/A
FABP7	6L90	x-ray diffraction	2.02 Å
FABP8	4BVM	x-ray diffraction	0.93 Å
FABP9	7FY1	x-ray diffraction	2.01 Å

Template selection was based on the highest sequence identity of the proteins in the FABP family. The matrix of %ID was obtained via BLASTP, and each FABP's selected template can be seen in **Table 2.2**.

Table 2.2: Chosen templates for AlphaFold and Homology Model simulations for each FABP, based on the highest amino acid sequence identity within the protein family.

Protein	Most similar FABP	FABP PDB template	Sequence ID%
FABP1	FABP6	2MM3	36.72%
FABP2	FABP7	6L90	36.15%
FABP3	FABP7	6L90	67.18%
FABP4	FABP8	4BVM	67.18%
FABP5	FABP8	4BVM	58.73%
FABP6	FABP1	6DO6	36.72%
FABP7	FABP3	3WVM	67.18%
FABP8	FABP9	7FWK	67.42%
FABP9	FABP8	4BVM	67.42%
FABP12	FABP8	4BVM	65.91%

2.2 Methods

2.2.1 Protein structure preparation

Template preparation was performed before HM simulations. The templates mentioned in the paragraph above were loaded in MOE after their respective .pdb file was downloaded from the PDB. Proteins whose info is contained in .pdb files often contain missing atoms, alternate geometry, or artifacts that should be attended to before moving forward with other computational studies. These issues stem from the inherent limitations of experimental data derived from X-Ray Crystallography and NMR which cannot resolve the atomic data with absolute accuracy.

The Structure Preparation panel in MOE shows the major issues present in the loaded protein, which are categorized into three different types based on their impact: Issues, which are must-fix problems that would otherwise negatively impact other simulations; Protonation, which highlights residues with possible rotamers, protomers, or tautomeric states; Warnings, which are problems with the structure which will not be automatically fixed by MOE and will require user intervention.

The Issue category is further split up into multiple other entries: Alternates are residues that have alternate locations and/or ambiguous sequence identities, typically arising from regions of the crystal structure allowing multiple possible conformations. Termini identifies protein chain C or N-termini, which need to be charged or capped, as well as empty residues preceding a capped N-terminus or following a capped C-terminus. Break highlights interruptions in the amino acid sequence, which can be fixed using the loop modeler built into MOE. Library lists a variety of issues related to the determination of the protein and subsequent transcription in the .pdb file, which can create discrepancies based on the name of the residues and how MOE handles their actual structure. Lastly, Charge identifies molecules in the system with incorrect or no charges.

The Protonate3D function included in the Structure Preparation panel can be used to automatically assign tautomeric and protomeric states, optimize the position of hydrogens present in the protein, add them where necessary, and finally assign partial charges if an energy minimization step is also present in the workflow. This is performed based on user-imputed settings for temperature, pH, and the salt concentration in mol/L for the solvation model.

The Quickprep function available in MOE provides an automatic process that handles the correction, protonation, tethering, and minimization stages of structure preparation of a newly obtained protein. It includes the functionality of Structure Preparation and Protonate3D, but also fixes atoms further than a selectable distance from the binding site (if present) to improve computational times and tethers the atoms that are not fixed with restraints. During minimization, these restraints

enforce geometric constraints such as distances, bond angles, and torsion angles, thus allowing the system to be relaxed in the conformational space and avoiding steric clashes.

2.2.2   Homology Modelling

For each FABP, 12 included, the corresponding FASTA file was loaded into MOE, along with its selected template's .pdb file. Like previously explained, the templates were corrected and minimized using the Quickprep function. Using the SEQ panel, it is possible to start the HM prediction by clicking on the Homology Model button, which lets the user select a template and sequence. It is also possible to set constraints on the generations of the rotamers which make up the side chains of the sequence, limit modelling to select parts of the sequence, include C and N terminal modelling and setting a value for the precision of the refinement stage, which is when the final model is ultimately selected after scoring a selected number of intermediate models (10 by default). The scoring of the intermediates was performed according to the Generalized Born/Volume Integral (GB/VI) method, which minimizes the electrostatic solvation energy of the constructed model. MOE automatically builds the structure of the protein such that its residues are constrained within the Ramachandran Plot, meaning that backbone amino acid dihedral angles are permitted. If there are outliers in the intermediates, MOE uses its Loop Modeler to rebuild the selected residues based on an internal database of PDB-derived loops or by using a *de novo* method. The best model was saved after performing an additional Quickprep routine for later comparison.

2.2.3   AlphaFold

AlphaFold simulations make use of four main parameters to run protein structure simulations, which can improve the Ranking Score (RS) [0, 1] of the obtained structures and, therefore, their quality: 1) MSA depth, which is the number of sequences in the Multi Sequence Alignment (MSA). The quality of MSA depends on depth and diversity of representation, meaning that high MSA will help identify co-evolutionary signals and use them to figure out the protein's 3D structure; 2) Number of cycles, which is the number of times the results of the neural network are recycled through the same steps to get refined; 3) Number of random seeds, which define the random starting conditions for the neural network to perform the simulation; 4) Template, which can either be user selected or automatically implemented by the MSA search algorithm of AF and improves prediction if a similar structure is found. For the first two settings, the maximum allowed number was selected, while 10 seeds were generated. Templates were loaded according to **Table 2.2**.

JSON files were built according to AF documentation, containing the amino acid sequence of the FABP target and all previous parameters. A Python script was written to perform the data stage (which builds the sequences contained in the MSA) and the inference stage (which constructs the models of the target protein). AF's output is a folder containing JSON and .pdb files of the newly generated model, which also include result metrics (such as a general Ranking Score, pTM, pLDDT, and PAE) and can be used for validation and comparison.

2.2.4   Protein structure comparison

Once both HM and AF models of a given FABP were finalized, they were loaded in MOE along with the experimental FABP. Using the SEQ panel, the Align/Superpose function was selected in order to superpose the protein structures in such a way as to maximize the spatial overlap of a common atom subset, typically the gap-free positions in a sequence alignment. Such a superposition highlights both the regions of conserved structure and the areas of divergence or modification in the protein set, which can be useful to evaluate differences in the binding pocket of a family of proteins, such as the FABP family. After superposition, the SEQ panel will highlight the RMSD value between two selected structures.

2.3 Results

Determination of AF and HM models and evaluation of the results were performed by Enrico Astara, a Master's Thesis student of Politecnico di Torino, and explained in more detail in his own thesis. I have reported a summary of Materials, Methods, and Results necessary to obtain a full understanding of his work, such that my own results (which do include all HM models generated) are better placed in the context of the research, since my work is a continuation of Enrico's and relies on his results. Protein and template selection were performed together to have continuity between the two theses.

For the AF models, the RS obtained was between 0.93 and 0.94 for all structures. pLDDT and pTM values for generated AlphaFold structures can be found in **Table 2.3**. iPTM values are not present since FABPs are single-subunit structures; therefore, iPTM cannot be calculated.

AF's effectiveness in predicting structures can be seen after RMSD comparison of model structures (both HM and AF) to experimental ones, as shown in **Table 2.4**. The lower the RMSD, the better the superposition, which translates to a closer prediction of the generated structure. In particular, structures that make use of a template with sequence identity higher than 58% return RMSD scores that are lower than 1 Å. In fact, almost every experimental structure except for FABP6 is predicted to be closer to the models obtained through AF than to the ones from HM. FABP8 is the AlphaFold structure closest to its experimental respective structure, and it's also the protein that was used for FABP12 prediction as its template, which would indicate potentially good results.

As for validation of the quality of the structures, the Ramachandran plot of every single structure was generated, and the number of favorable residue positions was calculated. Results are available in **Table 2.5**. Generally, a structure is considered good if it has >90% favorable residues; this is true for all experimental structures except for FABP2, which comes close at about 89%, is valid for all AF generated structures, and for all HM structures except for FABP8, which once again sits at 89%.

It is interesting to note that HM generally performs worse or about equal to experimental structures (and this can be attributed to precision errors in standard techniques), which is to be expected. AF, however, performs exceptionally well for all FABPs for which multiple structures are available (all except FABP12) and performs above HM for the FABP12 model. This can be explained by how AlphaFold performs the search of structures to be used in its Multi Sequence Alignment algorithm of multiple proteins from which it can source the best possible conformation for segments of the final model itself.

Table 2.3: pLDDT and pTM values for FABPs 1-12 for AlphaFold generated models. Every score being higher than 0.9 indicates very robust results, particularly for proteins other than FABP12, which already possess experimental structures. The FABP12 model, however, does not find itself too much behind and was generated with good results even without an available structure on the PDB.

Protein	pLDDT	pTM
FABP1	94.66	0.9
FABP2	94.66	0.91
FABP3	96.98	0.95
FABP4	96.28	0.94
FABP5	95.4	0.93
FABP6	93.71	0.9
FABP7	96.02	0.93
FABP8	96.4	0.94
FABP9	95.87	0.93
FABP12	92.59	0.9

Table 2.4: RMSD comparison of superpositions of AlphaFold structures on experimental PDBs and superpositions of HM structures on experimental PDBs. Bolded numbers highlight the closest simulated structure to the experimental one.

Protein	Sequence ID% (FABP)	RMSD AF (Å)	RMSD HM (Å)
FABP1	36.72% (FABP6)	**1.03**	2.00
FABP2	36.15% (FABP7)	**1.19**	1.19
FABP3	67.18% (FABP7)	**0.23**	1.11
FABP4	67.18% (FABP8)	**0.25**	1.00
FABP5	58.73% (FABP8)	**0.59**	1.35
FABP6	36.72% (FABP1)	2.33	**2.32**
FABP7	67.18% (FABP3)	**0.39**	1.01
FABP8	67.42% (FABP9)	**0.20**	1.18
FABP9	67.42% (FABP8)	**0.45**	1.26

Table 2.5: Ramachandran evaluation of Experimental, AlphaFold, and Homology Modelling structures. Structures with >90% favorable residues are considered good.

Protein	Experimental	AlphaFold	Homology Modelling
FABP1	95.50%	100%	96.88%
FABP2	89.14%	99.21%	91.54%
FABP3	99.23%	100%	93.84%
FABP4	99.23%	99.21%	95.38%
FABP5	98.49%	97.69%	94.65%
FABP6	92.00%	100%	92.80%
FABP7	99.23%	98.38%	95.38%
FABP8	100%	99.23%	89.23%
FABP9	98.49%	99.21%	93.84%
FABP12	N/A	95.55%	93.84%
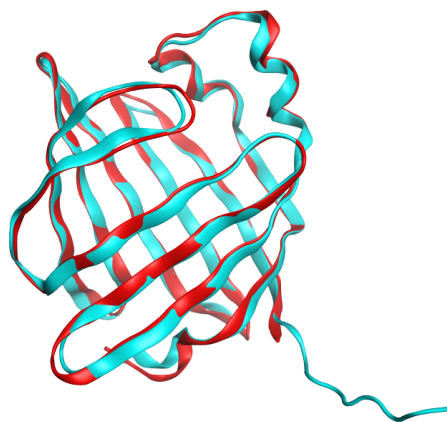
Figure 2.2: Superposition of HM FABP12 (red) and AF FABP12 (blue) models. Most of the difference between the two structures comes from the N-termini, and especially the C-termini, where AF is able to generate the end of the structure, which many experimental structures usually are not able to represent.

Chapter 3

Docking Simulations

3.1 Materials

3.1.1 Molecular Docking introduction and general workflow

With available target structures of the proteins of interest, it is possible to perform computational studies that reveal specific properties or mechanisms concerning the interactions between the amino acidic structures and the ligands that can potentially inhibit them. Molecular Docking, in particular, can be used to predict the strength of the interactions between two biological objects through the determination of their relative binding affinity.

In the context of rational drug design, Molecular Docking becomes evidently appealing when the binding mechanisms of a protein-ligand complex are unclear, and when the known ligands of the protein are few, if any. This is because this type of simulation requires only a 3D structure of the protein and a virtual database containing any number of ligands. These databases can contain a very high number of molecules (from a few dozen to millions), and as such, performing in-vitro experiments on that scale would be impossible due to the resources required, so performing molecular docking is far cheaper and faster.

Experimental tests are essential to ensure the validity of the in-silico predictions and remain the most precise method to determine target inhibition; computational software, however, can be very useful in the early stages of drug discovery to eliminate inactive or inadequate ligands from the databases and thus reduce the number of required in-vitro tests for validation. Molecular Docking, in fact, falls under the vast umbrella of Virtual Screening, which is a collection of computational methods and techniques for the evaluation of libraries of potential drugs with the final aim of identifying a few promising compounds or *hits* [73]. An example of the steps performed in Virtual Screening is shown in **Figure 3.1**.

When performing a drug discovery search for potential ligands, one of the first
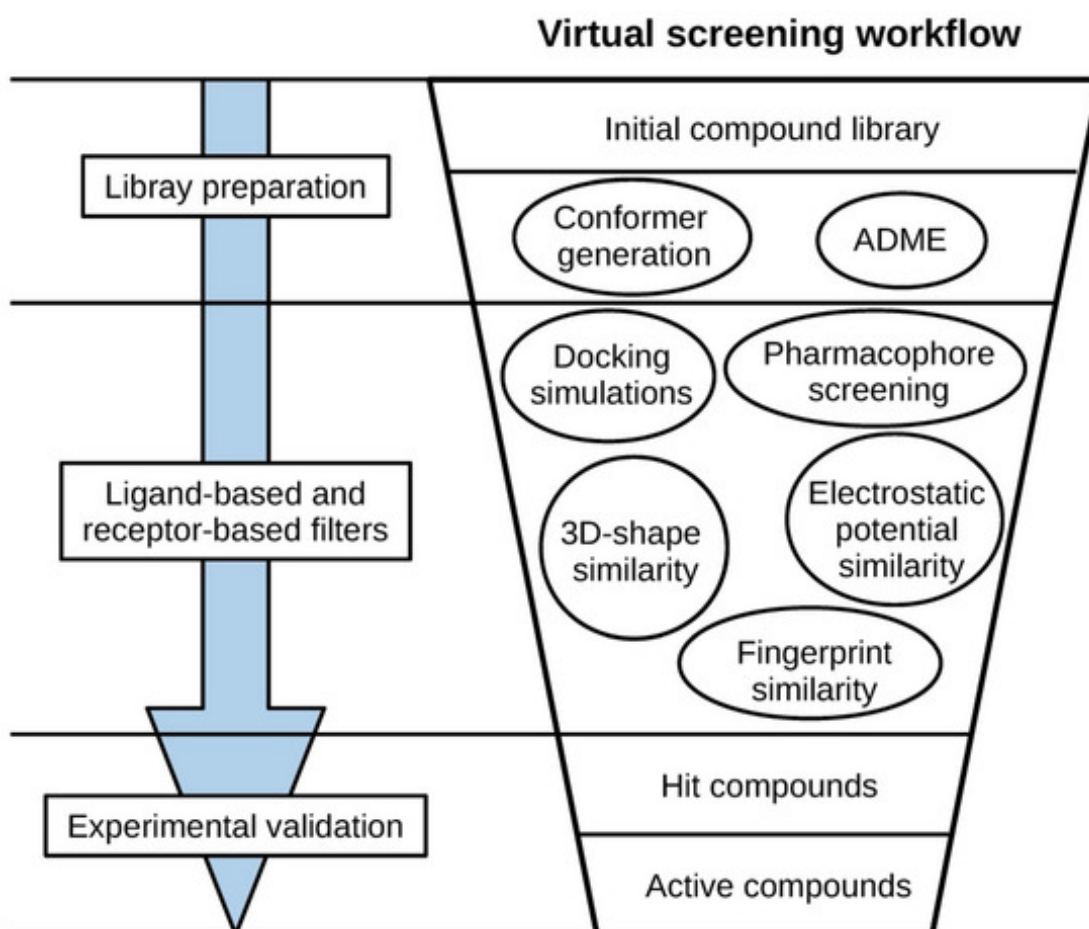
Figure 3.1: General Virtual Screening workflow. It is important to note that some steps can be performed earlier or later depending on the study performed, and some other steps can be ignored entirely based on the type of research conducted [73].

steps includes combing through the literature in order to find known inhibitors of the protein target. If the resulting database is not particularly big, similarity search models can be implemented to screen through other libraries obtained from public repositories (such as ZINC) to find similar compounds. Other Virtual Screening methods can then be employed to further reduce the number of potential inhibitors; for example, through the use of molecular descriptor filters (e.g., Lipinski's Rule of 5) or through the quantification of the binding affinity between protein and ligand obtained through Molecular Docking.

Virtual Screening and drug discovery as a whole are iterative processes that require iterative cycles of simulations, testing, and experimentation to optimize

results and obtain strong lead compounds for in-vitro assays. While the decision to implement most of the Virtual Screening strategies depends on the target, the starting databases, and the phase of the research, the determination of the binding affinity between the target and compound through Molecular Docking is unavoidable.
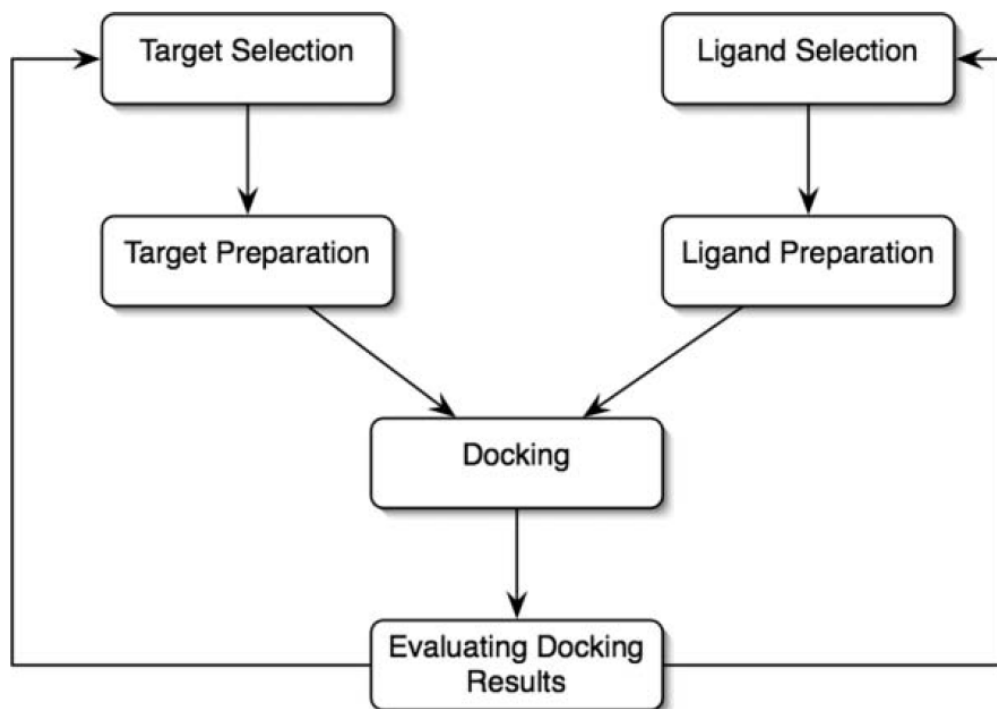
Figure 3.2: General docking workflow. Both target structures and ligands need to be prepared accordingly to the docking software used. After docking, the results of the scoring functions must be analyzed, selecting the binding poses with the best scores [74].

Figure 3.2 highlights the general workflow that is shared by docking protocols in different computational software. In the case of this thesis, target selection and target preparation have been explained in Section 2, while ligand selection, preparation, docking procedures, and evaluation of the results follow in the current Section.

3.1.2 Docking and Molecular Mechanics

Molecular Mechanics and Dynamics Molecular Docking, as previously mentioned, is a computational method that estimates the strength of the interactions

and the preferred binding orientation between a target (usually a macromolecule like a protein, DNA, or RNA) and a second molecule, which can be another protein (in protein-protein complexes) or, more commonly, a small compound. The vast majority of Docking software finds its roots in Molecular Mechanics, which considers the atom as the smallest unit to represent in the schematization of biological structures. This is possible due to the assumption of validity of the Born-Oppenheimer approximation, which allows to treat potential energy contributions of atoms and electrons separately [75].

The mechanics involved in these systems are dependent on Newton's laws and classical dynamics applied to atoms and bonds, which can be represented as mass-spring systems, respectively. This is a simplified alternative to another method, called *ab initio*, used for the representation of molecules, which involves Quantum Mechanics and the solution of Schrödinger's equation to find the most precise approximation.

While Quantum Mechanics-based docking protocols exist, they require expensive computational setups and far more time, whereas Molecular Mechanics-based simulations can efficiently provide results while not sacrificing too much robustness. The other advantage pertaining to Molecular Mechanics models is the ability to study larger structures and systems, whereas Quantum Mechanics generally has a far more limited scope [76].

The macromolecular structures that serve as the inputs for docking and the protein-ligand complexes obtained at the end of the docking process can be further refined with a Molecular Dynamics approach, which is an extension of Molecular Mechanics. As the name suggests, Molecular Dynamics studies the evolution of the system (its trajectory) under specific thermodynamic conditions and over short time frames, which are usually microseconds long at most. The trajectory of the protein or complex explores the possible conformational spaces available to it; due to Molecular Dynamics being deterministic and dependent on the initial conditions applied to the system, multiple sets of starting parameters can and should be used to adequately sample the possible states [77].

Energetic state of the system and force fields   Molecular Mechanics involves the determination of the potential energy states of the system using classical mechanics and Newton's law, which ends up being translated into the best binding pose in a protein-ligand system when performing docking studies. By extension, Molecular Dynamics simulations create coordinate snapshots of the system and its atoms' positions and velocities under the effect of the thermodynamic evolution of the parameters applied.

The main advantage of Molecular Dynamics comes from the analysis of the target protein or the complex placed in an explicit solvation system, which adds water molecules to the surrounding space. This improves the accuracy of the results
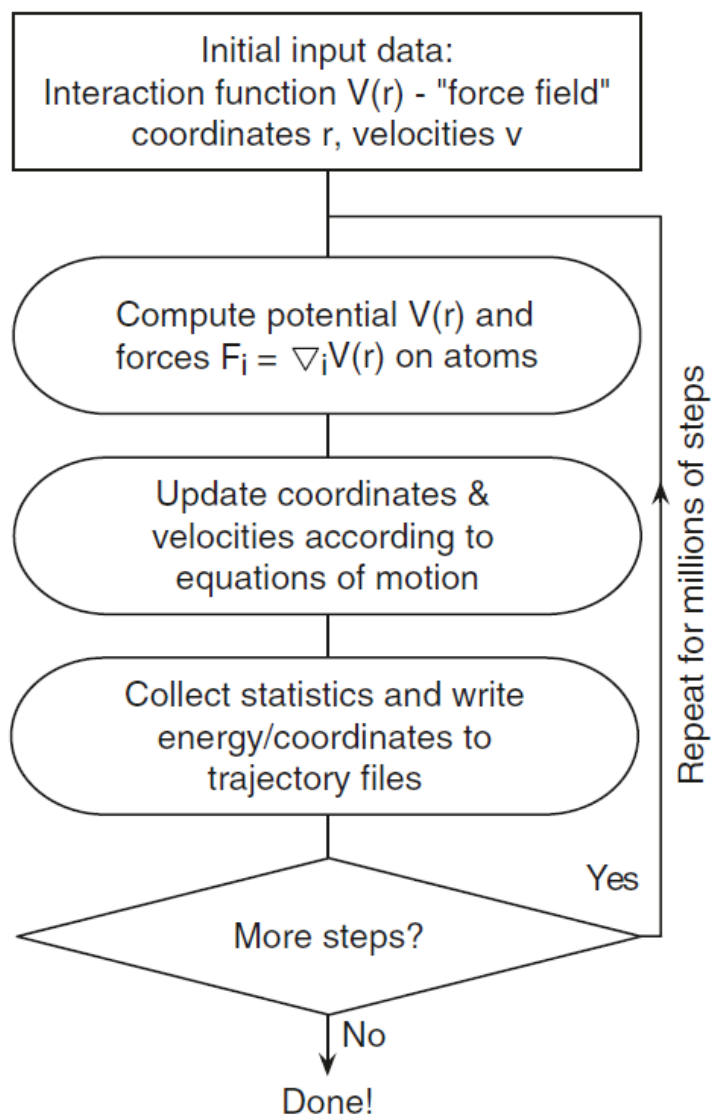
Figure 3.3: Simplified flowchart of a Molecular Dynamics simulation. The step time of the simulation is usually 1 fs, and this repeats for 1 μs for standard studies [74].

due to the higher similarity to a physiological state (surrounded by water in the body), but also due to the ability to model hydrogen bonds formed by the protein's residues, which can be important to elucidate its binding mechanisms [78].

Calculating the energetic state of the protein-ligand complex requires the use of a set of empirical functions, which is called a force field. Every computational software has its own type and implementation of a force field, but all of them consider the

same general energetic contributions: the atoms in a force field are represented as a combination of physical and electrostatic interactions, and their potential energy is calculated as a combination of bonded and non-bonded interactions with other surrounding atoms [79]:

$$E_{\text{MM}} = E_{\text{bond}} + E_{\text{non-bond}} \tag{3.1}$$

Every term in the force field's equation can be expressed using different mathematical expressions (also called functional forms), which become the components of the force field. Every functional form is defined by its parameters and the variable of the equation, and different force fields can possess the same forms but different parameters. An important note is that the parameters within a force field can be transferred to different systems (e.g., substances or functional groups); this is done because defining a unique equation for each different atom in a macromolecule would translate into intensive computational needs [80].

The parametrization of the functional forms has the added benefit of defining formulas that can be solved both experimentally and in silico, for result reproducibility. Every force field can have higher performance in certain situations, and depending on the scope of the system, a more complex one can be implemented alongside cross-terms and non-standard functional forms.
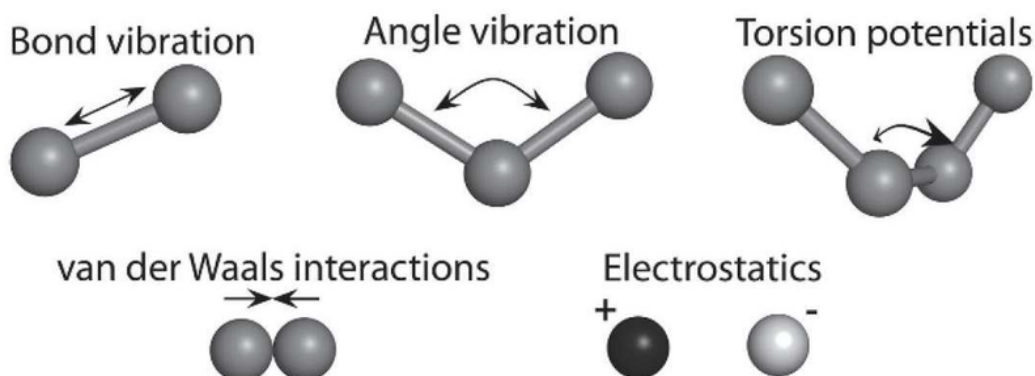
Figure 3.4: Examples of bonded and non-bonded interactions in force fields [81].

3.1.3 Bonded Interactions

As previously mentioned, the potential energy of a system's atoms in a force field is calculated considering bonded and non-bonded interactions. The former involves interatomic covalent bonds, and its contribution is calculated using 4 terms:

$$E_{\text{bond}} = E_{\text{str}} + E_{\text{bend}} + E_{\text{tors}} + E_{\text{cross}} \tag{3.2}$$

The stretching terms account for the modelling of atoms and bonds as mass-spring systems. Two atoms bonded together will be kept at an energetic equilibrium, such as to represent the most stable configuration. This can be expressed in simple terms with Hooke's law:

$$E_{\text{str}} = \sum_i \frac{k_i}{2}(l_i - l_{i,0})^2 \tag{3.3}$$

where k_i is the force constant of different atoms, l_i is the bond length between the two atoms and $l_{i,0}$ is the equilibrium bond length. This takes the form of a harmonic oscillator and can be used when the bond lengths are close to equilibrium. The further from the equilibrium, the less the term possesses significant relevance and thus requires the implementation of cubic or quartic terms in the Taylor expansion of Hooke's law, or should be supplanted by using the Morse potential formula instead.

The bending terms depend on the equilibrium bond angle that is formed between a set of two atoms and a third in common with both, and represent their most stable configuration. The formula describing the bending term can be written as:

$$E_{bend} = \sum_{angles} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 \tag{3.4}$$

where k_i is once again a force constant which varies atom by atom, θ_i is the bond angle between the three atoms, and $\theta_{i,0}$ is the equilibrium bond angle. Just like for the stretching terms, a third-order term can be inserted in the formula to improve the accuracy of the results.

The torsional terms include energetic contributions of dihedral angles, which are defined as the angle between two planes described by sets of three atoms in a four-atom configuration, where the first plane contains atoms A, B and C while the second contains B, C and D. Dihedrals are considered proper dihedrals when they allow 360° rotation of the angle and all the atoms are sequentially bonded, while in improper dihedrals this is not true. Improper dihedrals can be included to improve the quality of the force field, but are not always necessary. For proper dihedrals, the potential energy can be calculated with the following formula:

$$E_{tors} = \sum_{tors} \sum_{n} \frac{V_n}{2}(1 + cos(n\omega - \gamma)) \tag{3.5}$$

where V_n is the energy necessary to perform a rotation around a given bond, ω is the dihedral angle, and γ is a phase offset.

The last term is generally optional and only included in more advanced force fields. The cross terms, like the name suggests, consider the coupling of the different types of motions mentioned previously, which in real molecules are not independent of each other.

3.1.4 Non-Bonded Interactions

Non-bonded interactions can mainly be divided into two categories, the first one being Van der Waals interactions and the other being electrostatic or Coulomb interactions. These terms are dependent on the relationship between the atom and every other particle in the system, with potential energy changing due to the atom's movement in space. Both terms are sums of pairwise combinations of all items in the system; however, some manner of approximation is usually performed to improve performance.

$$E_{non-bond} = E_{vdW} + E_{el} \tag{3.6}$$

Van der Waals forces arise between pairs of atoms that are not chemically bonded and, as such, are relatively weak over long distances, and can occur even in atoms that possess no charge because of dipole charge fluctuations due to electrons shifting their position. Van der Waals interactions can be described using the Lennard-Jones potential, which takes into account a repulsive term that very rapidly decreases to zero and an attractive term that is valid up to 1.5-2 nm, which is also called London dispersion. The Lennard-Jones potential is calculated as:

$$E_{vdW} = \sum_{i=1} \sum_{j=i+1} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{3.7}$$

where σ is the Van der Waals radius (or the distance at which the potential energy is zero) and ϵ is the depth of the potential well, which is a measure of how tightly the particles interact. The two terms contain exponential terms that rapidly make the energetic potential decrease depending on the distance between the pairwise atoms. This means that, for the parametrization in force fields, the Van der Waals contribution can be disregarded at 1 nm or more.

Unlike Van der Waals forces, Electrostatic contributions are considered long-range and can be described using Coulomb's law:

$$E_{el} = \sum_{i} \sum_{j=i+1} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{3.8}$$

The longer range necessitates an approximation or the implementation of alternative equations due to the high computational power that would otherwise be necessary, and represents the major bottleneck in Molecular Mechanics simulations.

3.1.5 Solvation and minimization of potential energy

Protein complexes exist in the human body in an aqueous environment, surrounded by water. Generating a solvation system is, therefore, important to ensure the validity of the simulations and not perform them in a vacuum. There are two types of solvation systems that are used in Molecular Mechanics and Dynamics, which are called implicit and explicit systems [82]. The choice of the model, like for force fields, is dependent on the application and the size of the system, mostly due to the fact that the water molecules in an explicit system account for roughly 80% of the total particles. Given the influence and high performance required for the calculation of non-bonded interactions, the processing time necessary can grow exponentially.

Implicit Solvation   Implicit systems are simplified models that don't generate discrete water molecules surrounding the protein target. The idea behind this approximation is that, ultimately, only the effects of the solvent on the protein's surface are considered essential for the potential energy calculations. It is then possible to represent water as a continuous medium and parametrize its contribution in the force field through the implementation of a Solvent Accessible Surface Area (SASA).

The main issue related to implicit systems is that the correction term implemented is not easily comparable with other force fields (depending on the parametrization) nor with experimental data, which, of course, does not involve a direct comparison. Docking simulations involve a high number of ligands and sometimes even many possible targets, so the solvation models therein involved are mainly implicit-based.

Explicit Solvation   Explicit systems model all water molecules in the environment of the simulation, and while being much closer to reality, require much more computational power. Explicit inclusion of water also allows for the description of enthalpic, entropic, and kinetic effects, which makes this model required for Molecular Dynamics simulations, since the protein is affected by the thermodynamic evolution of the environment. The most common explicit system is TIP3P, which represents water molecules as a rigid three-point system of one oxygen and two hydrogens.

According to the minimum total potential energy principle of structural mechanics, bodies and molecules optimize their spatial disposition such that their potential energy is minimized. This means that, to perform molecular docking, it is important to explore their Potential Energy Surface (PES) and utilize energy minimization algorithms to find local energetic minima. The PES represents the

43

energetic landscape of the protein, and the local minima in the energetic potential generally correspond with the protein's equilibrium state and thus its active state, where the binding site is more exposed to interactions with a ligand.

Energy minimization protocols in computational software are able to employ force fields using solvent models to determine these low-energy configurations and prepare the structures for Molecular Docking or Molecular Dynamics simulations. Because docking simulations are mostly interested only in the area surrounding the binding site, they often employ a constrained energy minimization by tethering and fixing atoms that are too distant from the active site of the protein. This also allows for more easily determining a cutoff for the atomic non-bonded interactions to improve simulation performance.

3.1.6 Docking methodology

For this thesis, the explanation of the docking methodology will be limited to protein-ligand complexes; however, as mentioned previously, it is possible to analyze protein-protein interactions as well.

Molecular docking simulates possible binding poses between the two biological objects with the aim of finding the most energetically favorable configuration. The docking problem, then, becomes an optimization problem where the solution is found when optimizing the location of the ligand within the receptor through the use of a search algorithm and a scoring function [83].

These two aspects of docking software, alongside the protein's representation and flexibility, are the main defining characteristics of these computational tools, of which there exist a few dozen, between commercial ones and freely available ones. Due to recent advancements in artificial intelligence development, a few studies related to AI-accelerated docking and virtual screening workflows have been published [84], [85].

Protein Representation Protein representation in docking software can happen in various ways. The first ever program invented for molecular docking, DOCK [86], would perform what is called an explicit representation by placing every single atom in a set of Cartesian coordinates, and needed to check every possible translation and rotation of each atom. This proved to be quite time-consuming; therefore, modern alternatives make use of grid-based representations or molecular surface representations [87]. The former is based on a fast Fourier transform correlation method, which samples the protein after it is discretized on a grid [88]. The pre-calculated grid can be used to perform faster calculations of the binding affinity in the protein-ligand complex and can also allow for parallel computations on a single processor.

The other method is the one used in MOE and consists of detecting binding sites in the protein through the exploration of the protein's molecular surface by rolling a spherical solvent probe on the Van der Waals surface defined by the radii of all atoms of the structure. This molecular surface, which is none other than the SASA, is defined by all the points of contact between the Van der Waals surface and the probe and can identify gaps and cavities that can act as binding sites of the structure. During docking placement, the ligand will be constrained to the gaps and cavities in order to accelerate the process.

Protein Flexibility Docking software has different ways of handling protein flexibility, due to the interactions between ligand and receptor being limited to the area surrounding the binding site. Because simulating movement and thermodynamic fluctuations in the entire backbone of the protein would be inefficient (fixed docking), some programs like MOE introduce an induced-fit docking method [87]. This lets atoms within a set distance from the binding site move around when interacting with the ligand, therefore allowing for conformational changes and improved accuracy when calculating the binding affinity. MOE, in particular, will first perform a quick docking simulation using rigid docking and place the compound in the binding site, before performing a refinement step following the induced fit strategy.

Search Algorithms Search algorithms try to find the most energetically favorable disposition of the receptor-ligand complex by repeatedly simulating multiple binding poses based on some type of criteria [77], [83], [89].

Matching algorithms are among the simplest to exist and are generally very fast, though reliability is impacted by a priori knowledge related to the binding site of the protein. The general idea behind this type of method consists of placing the ligand into the most probable binding areas defined by the receptor's molecular surface or SASA. MOE implements several types of matching algorithms, and they all work best with different sizes of compounds. A pharmacophore search engine can also be implemented and generate binding poses based on the essential steric and electronic characteristics of an already bound ligand.

Stochastic algorithms are based on the random search of the possible conformation space of the binding poses, performed by generating random variations of orientation for rotatable bonds or translation of the whole ligand. One example is the Monte Carlo algorithm, which is based on random sampling of the search space and will accept/reject new binding poses based on the Metropolis-Hastings criteria, which determines the result based on the probability of the candidate state compared to the current state.

Systematic search algorithms, like the name suggests, try to systematically explore the entire search space of the complex. They can either do this by gradually

testing all the possible variations in translation and orientation of the atoms in the ligand (which works best for small ones) or by performing a fragment-based search: the ligand is divided into 2 parts and docked separately, before reconstructing the entire structure to analyze the all the valid configurations while drastically reducing the degrees of freedom of the ligand.

Scoring Functions   Scoring functions quantify the strength of the interactions between the receptor and the ligand by taking into account various characteristics, like Van der Waals and electrostatic potentials, but also bonds such as π-stacking and hydrogen bonds. Some docking methods also include covalent bonding in their calculations; however, this is generally disregarded due to covalent ligands having toxicity concerns [77], [83], [89].

A large number of docking programs use empirical scoring functions where the ligand-receptor interactions are categorized, counted, and assigned a score each, which are summed up to generate the final binding score. While they are based on actual physics, the empirical attribution is derived from the fact that the scoring is built with parameters such that it agrees with experimental data, which also means they work best with specific sets of ligands.

Something very similar happens with knowledge-based scoring functions, which derive their data from known protein structures and ligand-receptor complexes from databases like the PDB. The binding affinity is then calculated from statistical probability based on the frequency and distribution of the interactions in the complex.

Machine learning scoring functions are a subset of knowledge-based ones; while the latter are based on user-defined rules, the former learn complex patterns based on artificial intelligence algorithms. The datasets they use for training can be customized for specific proteins and ligands to further improve performance.

MOE has several different types of scoring functions; however, they are all based on force fields (also known as physics-based). They involve the same functional forms mentioned in Section 3.1.2 regarding bonded and non-bonded interactions between atoms, but if covalent bonds are not present, only non-bonded interactions are accounted for. Solvent models are also incorporated to keep track of the effect of water molecules on the system, though for docking, this is usually represented by implicit solvent systems with specific correction parameters depending on force field implementation. Quantum Mechanics and *ab initio* calculations can also be employed to further improve the accuracy of the calculations [76].

When performing a docking simulation between a protein target and a database containing a list of ligands, the docking software will output a list of the most probable configuration, or binding pose, for each combination. The best poses are determined by the highest binding affinity found by the search algorithms, which

is measured as the Gibbs Free Energy (ΔG) of the binding. This energy is related to the experimentally determined dissociation constant K_d through the following equation:

$$\Delta G = RT \ln(K_d) \tag{3.9}$$

where R is the universal gas constant and T is the temperature at which the simulation or experiment is performed. The K_d measures how likely a larger biological object (the ligand-receptor complex) will dissociate into separate parts; in other terms, k_d equals the concentration of the free ligand at which half the molecules of the receptor are bound to the ligand itself:

$$K_d = \frac{[\text{ligand}] \cdot [\text{protein}]}{[\text{ligand} - \text{protein complex}]} \tag{3.10}$$

K_d and ΔG share a relationship due to the above equation, which means that the lower K_d is, the more negative ΔG becomes, which translates into the compound having higher affinity for the target. The two measurements, however, indicate separate things: K_d is a pharmacokinetic descriptor of the reaction between the two components of the complex, and ΔG is a thermodynamic descriptor of the interaction. ΔG takes into account enthalpic and entropic contributions of the complex, which are none other than the bonded/non-bonded interactions and the degree of freedom of the system, respectively. The thermodynamic formula describing ΔG is as follows:

$$\Delta G = \Delta H - T\Delta S \tag{3.11}$$

While it is possible to convert the values of K_d and ΔG from one to another, it is important to remember that the binding affinity calculated through docking software is based on empirical formulas and functional terms that have been approximated, and as such, it only represents a general indication of the overall affinity of the ligand.

3.2 Methods

3.2.1 Ligand Database preparation

Before proceeding to docking simulations, it was necessary to generate a database containing the ligands that the user wants to test against the protein target. One of the most common strategies involves searching for existing and active ligands that are known to inhibit the receptor of interest. In the case of FABP12, there is, to the author's knowledge, no experimentally published and public info about known inhibitors except for one case: compound BMS309403 [32], which was already known to be a pan-inhibitor of other proteins in the FABP family, most notably FABP4 [90].
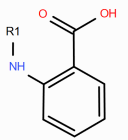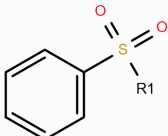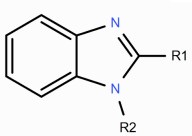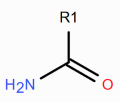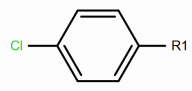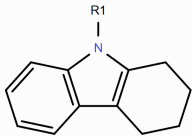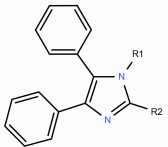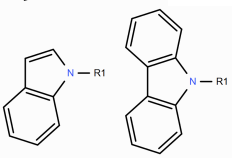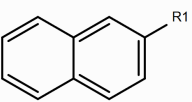
Due to the shared characteristics of the FABP protein family, their high degree of similarity in terms of homology and common aspects of their binding site, the search for known inhibitors was expanded to include all compounds mentioned in both experimental and in silico studies of FABPs. Common web tools like Google Scholar, PubMed, and PubChem were used to perform a search of all scientific publications that were released up to May 2025, which contained compounds tested for FABPs inhibition. These ligands were either tested experimentally in displacement assays of FABPs to find their inhibitory capability (in the form of k_i, k_d, or IC50) or were compounds found through Virtual Screening and possess docking data in the form of a scoring function.

Each scientific article reported the structures of the chemical compounds used for testing either in the form of IUPAC notation, 2D molecular graphs (which are just a visual representation of the compound's structural formula), or using a SMILES representation. The SMILES (Simplified Molecular Input Line Entry System) notation can be used to succinctly and uniquely describe a molecule's structure using only one line of ASCII characters.
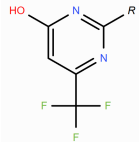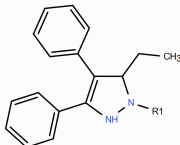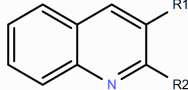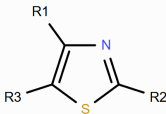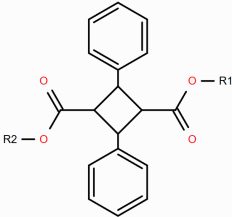
The 2D structural graphs of the potential inhibitors were converted to SMILES notation using a combination of web tools: Decimer [91], for image to 2D structure conversion, Ketcher [92], for 2D structure modification and 2D to SMILES translation, and IUPAC to SMILES converters [93].

The final database, stripped of duplicate ligands and stereoisomers, contains 816 compounds from 57 different publications [45], [46], [48], [50]–[56], [94]–[140]. Every chemical compound was categorized based on its chemical structure, representation, and commercial use (if available) as reported in the original paper [Data not shown]. The main chemical classes and base structures used for derivatization are shown in **Tables 3.1** and **3.2**.

48

Table 3.1: Main classes of chemical compounds.

Chemical Classification	Chemical Structure
Aminobenzoic acid Derivatives	
Benzenesulfonamide Derivatives	
Benzimidazole Derivatives	
Carbamoyl Derivatives	
Chlorobenzene Derivatives	
Hydrocarbazole Derivatives	
Imidazole Derivatives	
Indole Derivatives	
Naphthalene Derivatives	

Table 3.2: Main classes of chemical compounds (cont.).

Chemical Classification	Chemical Structure
Pyrimidine Derivatives	
Pyrazole Derivatives	
Quinoline Derivative	
Thiazole Derivatives	
Truxillic acid Derivatives	

For brevity and relevance, each category will not be described; however, reviews from Floresta et. al. [95]–[97] on FABP4 known inhibitors and scientific articles on truxillic acids for FABP3, FABP5, and FABP7 inhibition [128]–[130] contain most of the relevant information on the categorization. **Figure 3.5** shows the distribution of the chemical compounds based on their classification: there are 33 different classes; however, half can be included in a miscellaneous category ("Other"), as it includes fragments, unclassified compounds, and classes with only a couple of compounds each.

Every ligand is associated with one or multiple FABPs if it meets one or more of these conditions: the ligand has known experimental data based on displacement assays, the ligand was part of computational docking simulations for which a score was obtained, the ligand was found as a result of a high throughput virtual screening or ligand derivatization studies for one or more FABPs even though no experimental or computational results are shown in the papers. Due to this choice,
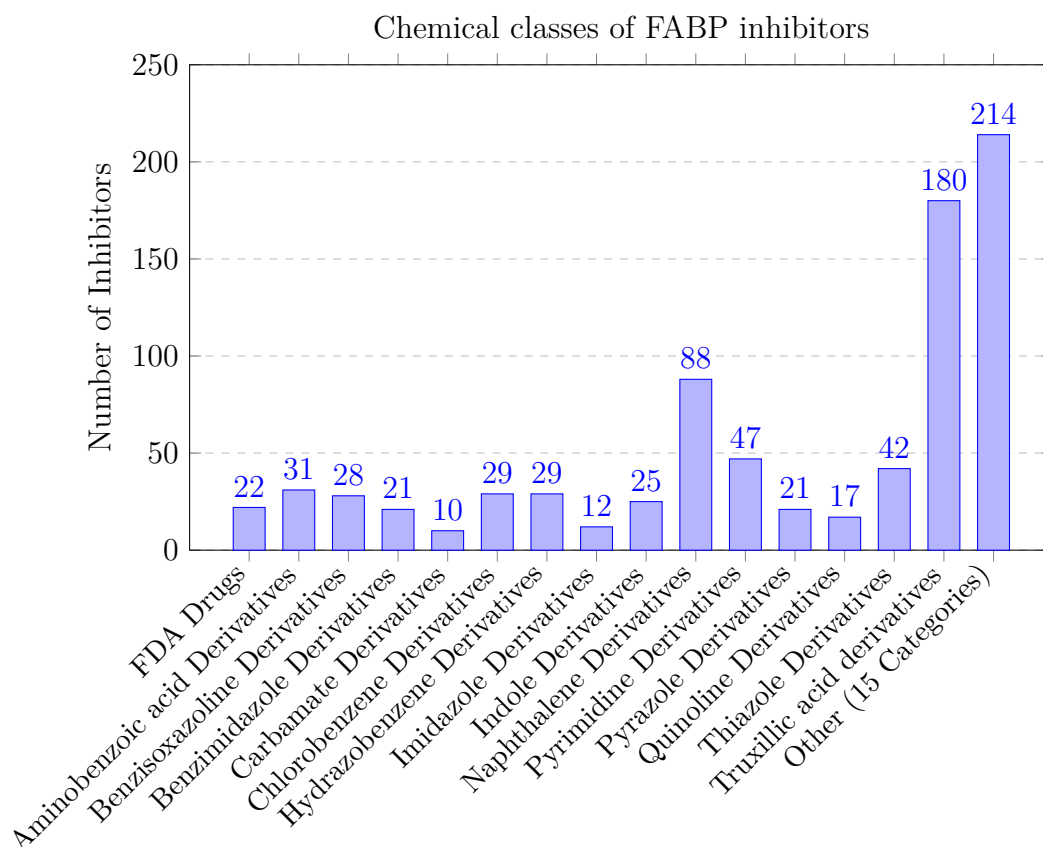
Figure 3.5: Chemical classes of FABP inhibitors. There is a good distribution of different types of chemical compounds within the database, meaning results are not biased towards only one singular type of inhibitor, and provide a good base for the exploration of the binding mechanisms of FABPs. The "Other" category contains fragments, other types of chemical classes, and unclassified chemical compounds.

each ligand is tagged with one or multiple FABPs, which represent at least one instance of the conditions above being met. Duplicate instances were not tracked (e.g., a compound appearing in multiple publications from the same research group without presenting new experimental data on that compound).

Figures 3.6 and **3.7** show the distribution of how each compound was tagged: the first table keeps track of ligands that have been involved with only a singular FABP, while the second one shows instances where the ligand is associated with multiple proteins. These tables can be used to notice an approximate trend in FABP testing, which has mostly consisted of FABP4, FABP5, FABP3, FABP7, and FABP1, in that order.

51
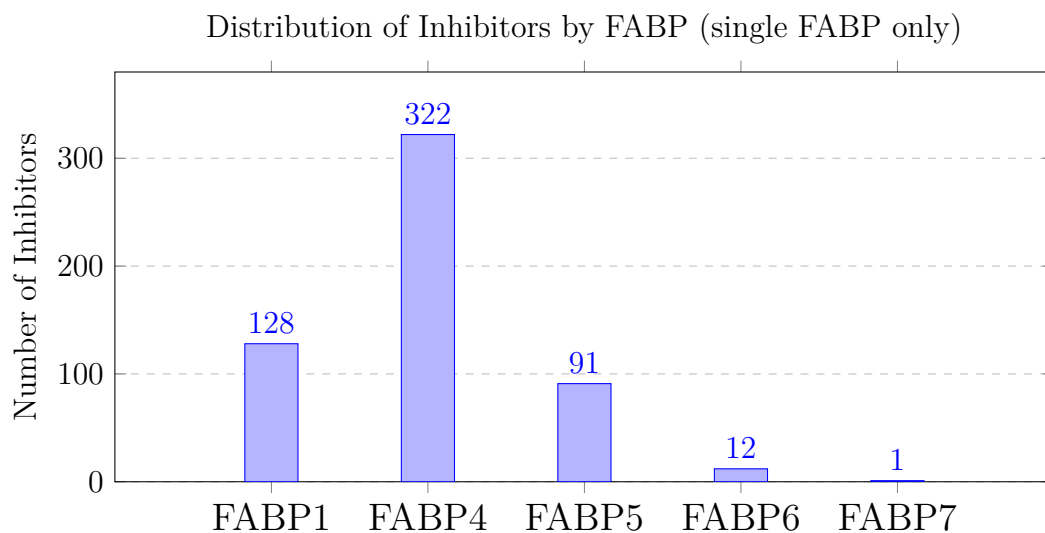
Figure 3.6: Number of inhibitors identified that are related to only one type of protein. FABPs that do not appear in the graph do not possess a specific ligand associated with them.
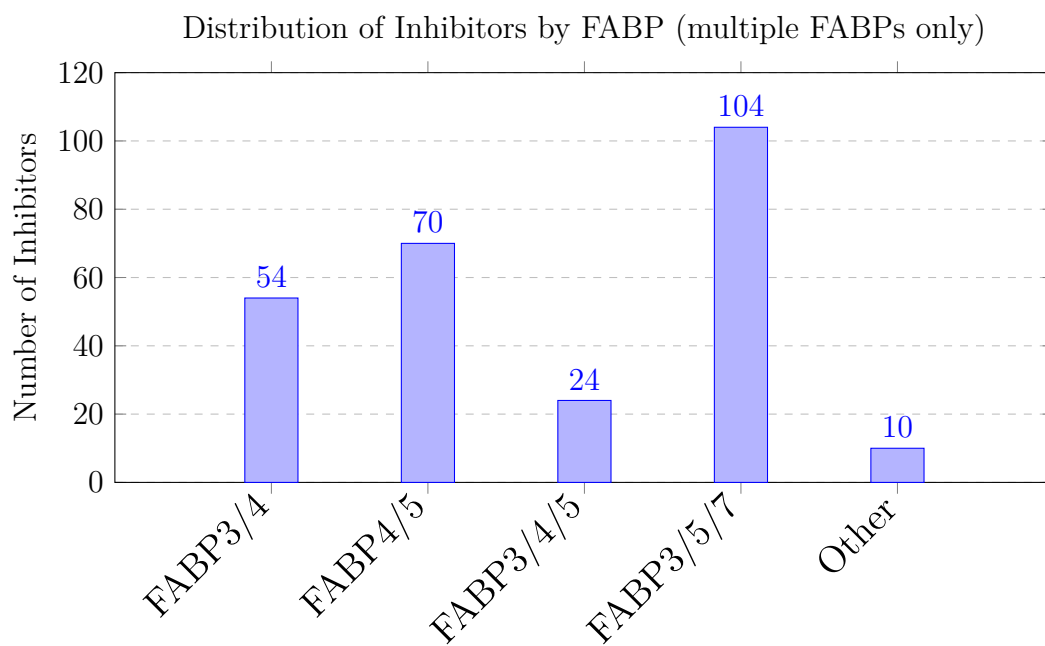
Figure 3.7: Number of inhibitors identified that are related to multiple proteins.

3.2.2 Molecular Docking in MOE

Molecular docking was performed on all FABPs with experimentally determined structures (FABPs 1-9) using .pdb files obtained from the PDB as described in Section 2.1.5. Moreover, docking was also performed on AlphaFold-generated structures for all FABPs, including 12. Due to the testing required to compare results of Homology Modelling and AlphaFold models in Section 2.2.4, all structures had already gone through the Quickprep, Protonation, and Energy Minimization steps required for the preparation for docking.

Binding Site Identification   The Site Finder tool of MOE was used to identify the binding pockets available for ligand interaction in the FABPs. After a protein has been loaded into the software, Site Finder uses a geometric-based algorithm to calculate the possible active sites starting from the 3D atomic coordinates of the structure. MOE will identify sites that are exposed to solvent and initially classify them as hydrophobic and hydrophilic, before placing alpha spheres used for ligand placement in the same spots.

Alpha spheres are pruned when the location in which they are placed is not favorable for the steric fit of the ligand. Finally, collections of alpha spheres are clustered to form binding sites, which are then ranked based on their Propensity for Ligand Binding (PLB) score, which is calculated using the amino acids that make up the pocket. The higher the PLB score, the better the chances of the ligand interacting with the receptor.

Site Finder was used to identify the main binding pocket used by the proteins to interact with fatty acids, which is reported by the literature for all instances of FABPs as mentioned in Section 1.4.2. Both the main site and the portal region's site in FABP1 were identified by Site Finder; however, the latter was disregarded due to the complexity of replicating the actual binding mechanisms related to the portal region's site.

When and if necessary, alpha spheres identified by MOE were manually removed based on the binding sites knowledge as obtained from the literature (e.g., if MOE had placed an alpha sphere outside the protein, necessary due to the binding site being inside the beta barrel structure of FABPs).

Docking Setup   Before loading it into MOE, the SMILES database containing the information about the chemical structure of the compounds was opened in DataWarrior [141], and each ligand was converted into .sdf files, which contain the atomic coordinates for their 2D representation. The database was then opened in MOE and underwent a washing step: this converts the 2D ligands into low-energy 3D conformations with the correct tautomeric and protonation states.

With the structure and the database loaded, the docking simulations were run

by opening the Compute -> Dock -> General menu. After selecting the binding site. MOE requires user input regarding the Placement and Refinement steps of the procedure. When MOE performs a docking simulation between ligand and receptor, it does so in two steps: the Placement step uses one among several selectable matching algorithms to generate possible binding poses, while the Refinement step uses one of several force field or empirical based scoring functions to quantify and score the binding affinity of the pose generated in the previous step, only keeping the best ones.

The Refinement step initially refines the preliminary pose by minimizing the energy in the complex. It can do so using either a Rigid Receptor mode, where only the ligand is allowed to move, or the Induced Fit mode, where the side chains of the receptor become flexible, allowing for more accuracy at the cost of higher computational time.

Each step has some poses generated or saved associated with it, and in the case of the simulations performed with FABPs, the number of poses placed with the Placement step was 50, and the number of poses refined with the Refinement step was 5. The Placement method chosen was the Triangle Matcher algorithm, which generates poses by aligning triplets of atoms of the ligand on triplets of alpha spheres designated within the binding site by the Site Finder function.

The Refinement mode selected was Induced Fit, due to its higher performance, and the scoring function used to determine the S Score (the equivalent to ΔG) was the force field-based, Generalized-Born Volume Integral/Weighted Surface Area (GBVI/WSA) scoring function, which uses an implicit solvation method. The formula that describes the force field is:

$$\Delta G \approx c + \alpha \left[\frac{2}{3}(\Delta E_{\text{Coul}} + \Delta E_{\text{sol}}) + \Delta E_{\text{vdW}} + \beta \Delta SA_{\text{weighted}} \right] \tag{3.12}$$

where c represents the gain or loss in degrees of freedom of the system, α and β are constants which were determined during training of the force field, E_{Coul} is the Coulombic electrostatic term, E_{Coul} is the solvation electrostatic term calculated using the GBVI solvation model, E_{vdW} is the Van der Waals contribution to binding, and SA_{weighted} is the SASA.

3.3 Results

In each docking run, comprising one protein structure and the entire ligand database, every receptor-ligand combination had output 5 refined docking poses. Only the best 3 poses per combination were saved due to the variability of the results obtained, to ensure the correct binding mechanisms in the poses, and to avoid outliers.

The final S Score for every complex combination was then obtained by averaging the 3 best poses. The binding affinity is expressed as a negative number in favorable binding conditions, in accordance with Gibbs Free Energy; this means that the more negative the S Score, the higher the affinity of the ligand.

The S Score is a predictive number that should converge with sufficient computational time and a sufficient number of simulated poses. However, given the relatively large size of the starting ligand database and the 19 FABPs structures, the computational time necessary to reach consistency in multiple poses was deemed too high for this research.

Docking Simulations and Ligands   Due to the sheer number of simulations performed, data reported in this thesis regarding the S score of docked ligands will only include the best 137 ranked compounds for FABP12; however, data and rankings were calculated for all 816 with all 19 protein structures.

This choice is made due to multiple reasons: the 113th-ranked compound for FABP12 is pan-inhibitor BMS309403. 12 of the 113 compounds that rank better than BMS309403 have also been identified as commercially available, which means that they can be used to validate the results of the models and docking simulations by performing displacement assays in vitro.

Additionally, 2 other compounds, ZTY-28 (ranked 115th) and ART26.12 (ranked 137th), are also shown in the results: the first one is another commercially available ligand that ranks just below BMS309403 and, as such, was included due to potential variability of the S Score prediction; the second ligand, instead, is included due to being the only FABP-related compound with a clinical trial currently in progress.

While it doesn't rank as well as other compounds for FABP12 docking, ART26.12 is a potential drug that is seemingly specific for the inhibition of FABP5 in the context of PCa, but also breast and cervical cancer, dermatology, and Chemotherapy-Induced Peripheral Neuropathy [142]. As such, both it and BMS309403 can be used as a good metric of comparison. The chemical structure of the 15 compounds can be seen in **Figures 3.8** and **3.9**. As a note, the names of the compounds reported have been taken from PubChem (if available) or from their respective publication article otherwise.

(1) GW7647

(2) OMDM2

(3) STK-22

(4) Compound 1

(5) AM1172

(6) SBFI-103

(7) OMDM1

(8) STK-15
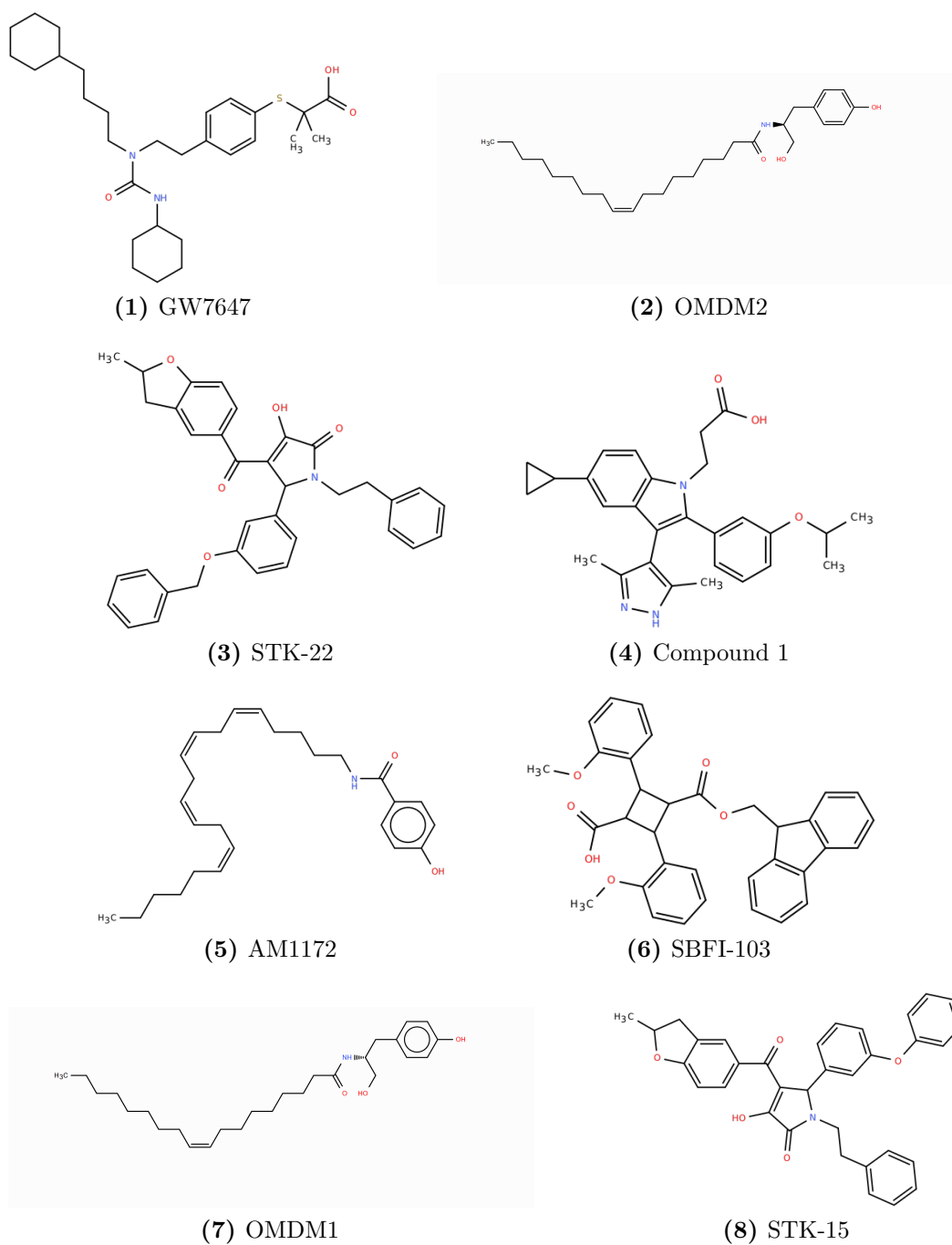
Figure 3.8: Chemical compounds chosen for further analysis (1-8), these compounds perform better than pan-inhibitor BMS309403 and are commercially available, and as such can be used for result validation.

9 Antraferine

10 α-FABP-IN-1

11 Compound 34

12 AM404

13 BMS309403

14 ZTY28
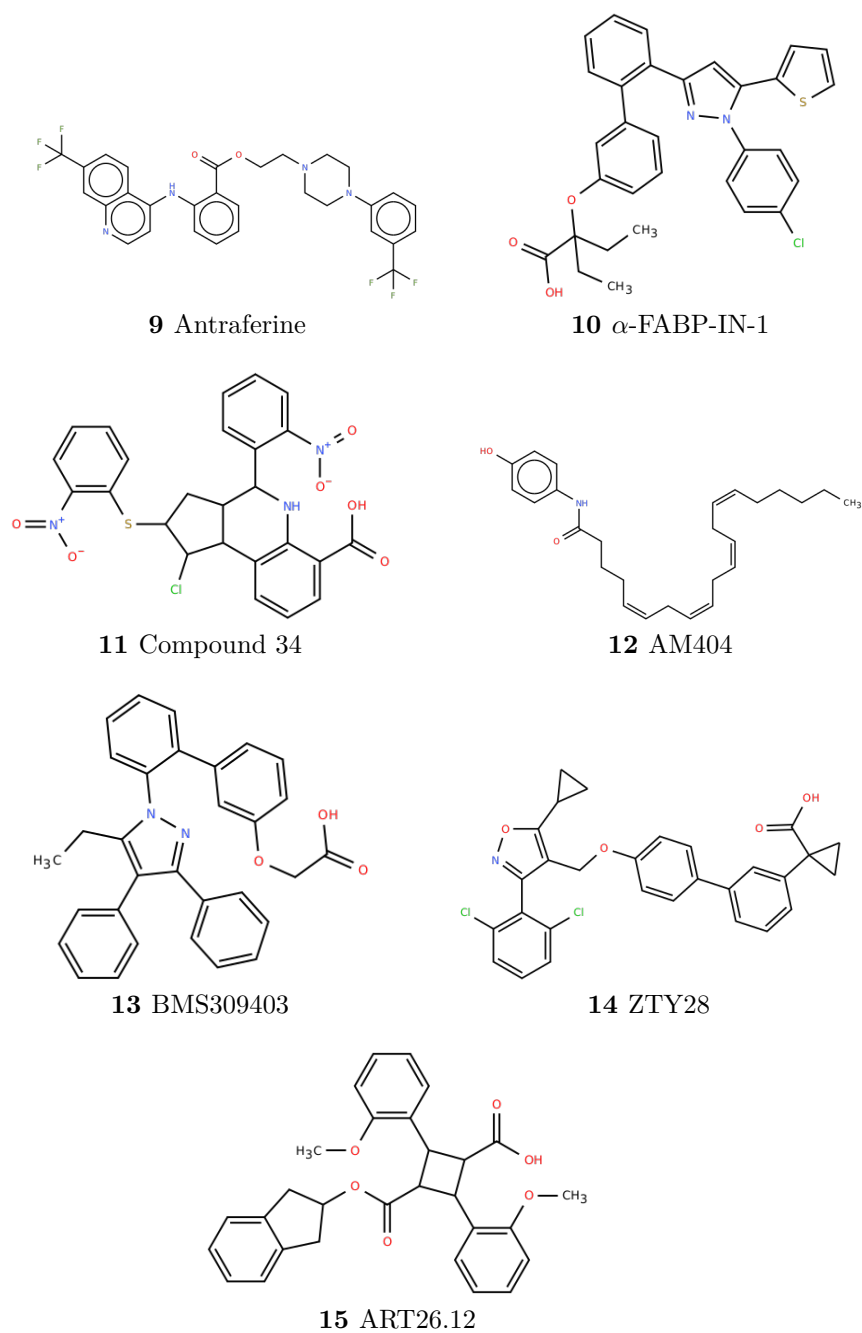
15 ART26.12

Figure 3.9: Chemical compounds chosen for further analysis (9-12). Compound 13 is BMS309403, Compound 14 is ZTY28, and Compound 15 is ART26.12.

AlphaFold vs Homology Modelling For each FABP other than 12, for which an experimental structure is not available, docking simulations between the AlphaFold model and the experimental protein were compared. The average error measured in the AlphaFold docking S Scores is about 3 to 9%, depending on the protein. This error also considers the variability of the S Score of the generated poses not converging, which amounts to around 3-6% depending on the simulation; this would further reduce the real error between AlphaFold and experimental structures simulations.

It is interesting to note that FABP8, FABP12's template in the AlphaFold model, had the least RMSD when comparing the AlphaFold and HM structures obtained in Section 2.2.4. This, combined with the good quality of the generated FABP12 AF model, signifies there is a high chance that the model accurately represents the binding mechanisms of its active site of the protein. The measured average error between AlphaFold models and experimental structures is shown in **Figure 3.10**.

Average percentage error of AlphaFold docking simulations
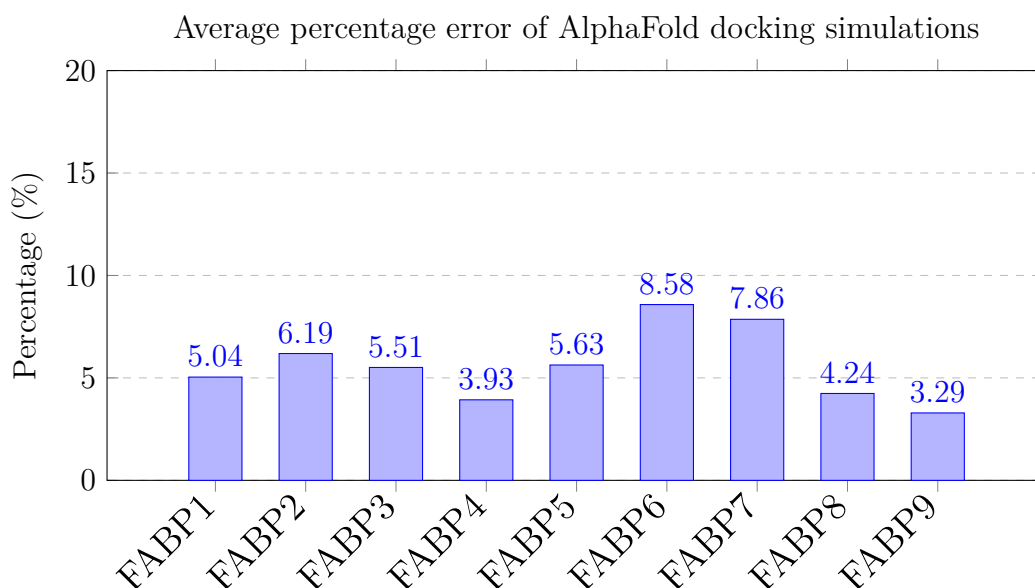
Figure 3.10: Average percentage error of AlphaFold docking simulations. The average error for all FABPs except 12 (for which there is no experimental structure to compare the AF results to) is between 3.29% for FABP9 and 8.56% for FABP7. Given the quality of the FABP12 AF model, the average error for the structure is hypothesized to be less than or around 10%. Longer docking procedures, which would generate more poses and reduce S score variability, would further reduce the expected error.

Ligand Rankings   Using the S Scores, each ligand was ranked using two different metrics. The first one, named Affinity Ranking, was already mentioned as using the highest to lowest binding affinity. **Table 3.3** shows the S Scores for the best 137 ligands docked to the most relevant FABPs in PCa (FABP3, FABP4, FABP5, FABP9), ranked based on the compound's FABP12 performance.

Table 3.3: Binding affinities (S Score, ΔG) of the top 137 compounds.

Compound	FABP3	FABP4	FABP5	FABP9	FABP12
1	-11.51	-10.66	-10.83	-9.79	-11.74
2 (GW7647)	-9.85	-9.62	-8.98	-9.68	-10.24
3	-10.60	-8.68	-9.71	-9.72	-10.10
4	-10	-9.22	-8.78	-10.2	-10.04
5	-9.68	-9.08	-8.91	-9.19	-9.98
6	-9.94	-9.34	-9.12	-8.77	-9.91
7	-10.06	-9.03	-9.08	-8.92	-9.91
8	-9.65	-8.87	-9.12	-8.96	-9.91
9	-9.84	-9.63	-8.48	-9.35	-9.89
10	-9.53	-9.43	-8.58	-9.2	-9.78
11	-9.41	-8.68	-8.46	-8.75	-9.78
12	-9.53	-9.51	-8.81	-8.95	-9.77
13	-9.89	-9.76	-9.45	-9.32	-9.72
14	-9.21	-9.27	-8.4	-9.03	-9.71
15	-9.8	-9.37	-8.97	-8.66	-9.69
16	-9.3	-8.92	-8.63	-8.43	-9.69
17	-9.36	-9.56	-8.62	-9.15	-9.68
18	-9.41	-9.54	-8.99	-8.43	-9.68
19	-9.76	-9.93	-9.05	-10.03	-9.67
20	-9.55	-8.91	-8.99	-9.26	-9.67
21	-9.23	-9.05	-8.59	-9.13	-9.65
22	-10.39	-9.51	-7.81	-9.36	-9.64
23	-9.66	-9.19	-8.88	-9.3	-9.64
24	-9.39	-8.78	-8.07	-8.91	-9.63
25	-9.94	-9.4	-8.62	-8.51	-9.63
26 (OMDM2)	-9.33	-8.73	-9.34	-9.63	-9.6
27 (STK-22)	-9.62	-8.85	-8.4	-9.45	-9.6
28	-10.28	-9.22	-8.9	-9.1	-9.6
29	-9.99	-9.26	-9.41	-9.94	-9.59
30	-10.29	-8.99	-8.85	-8.49	-9.59
31	-9.43	-8.73	-8.72	-9.75	-9.55
32	-9.99	-9.52	-8.75	-9.19	-9.53
33	-9.14	-9.14	-8.23	-8.82	-9.53
34	-10.15	-9.91	-8.79	-9.02	-9.53
35	-10.01	-9.5	-8.8	-9.16	-9.52

Continued on next page

Table 3.3: (continued)

Compound	FABP3	FABP4	FABP5	FABP9	FABP12
36 (Compound 1)	-9.03	-9.19	-9.14	-9.51	-9.52
37	-9.83	-9.06	-8.59	-9.12	-9.51
38 (AM1172)	-9.43	-9.55	-8.88	-9.29	-9.49
39	-9.78	-9.47	-9.58	-9.37	-9.48
40	-9.49	-8.95	-8.44	-9.61	-9.47
41	-9.56	-9.09	-8.41	-8.47	-9.46
42	-9.96	-9.16	-8.9	-8.43	-9.46
43	-9.43	-9.25	-8.52	-9.37	-9.43
44	-9.72	-9.39	-8.36	-9.15	-9.43
45	-9.62	-8.99	-8.92	-9.39	-9.42
46	-9.69	-9.48	-8.17	-8.47	-9.41
47	-9	-7.72	-8.6	-9.23	-9.39
48	-9.07	-8.92	-8.58	-9.06	-9.39
49	-9.87	-7.81	-8.46	-6.61	-9.39
50 (SBFI-103)	-10.01	-9.15	-8.34	-9.18	-9.38
51 (OMDM1)	-9.96	-9.46	-9.23	-9.8	-9.37
52	-9.04	-8.51	-8.4	-8.98	-9.37
53 (STK-15)	-9.93	-9.25	-9.09	-9.63	-9.37
54 (Antrafenine)	-9.81	-9.07	-9.22	-9.23	-9.34
55	-9.8	-8.55	-9.16	-9.04	-9.34
56	-9.72	-8.71	-8.82	-8.88	-9.34
57	-9.03	-9.17	-9.24	-9.23	-9.32
58	-9.4	-8.99	-8.9	-9.04	-9.32
59	-9.92	-9.45	-7.99	-9.37	-9.32
60	-9.52	-9.35	-7.85	-9.02	-9.31
61	-9.03	-8.85	-8.88	-8.88	-9.3
62	-9.26	-8.92	-8.05	-8.6	-9.3
63	-9.5	-9.24	-7.89	-8.64	-9.28
64	-9.98	-8.4	-8.85	-9.21	-9.28
65	-9.2	-8.82	-8.23	-8.44	-9.28
66	-9.47	-9.42	-8.76	-9.27	-9.27
67	-9.2	-9.27	-7.72	-8.45	-9.27
68	-9.49	-8.9	-8.51	-8.62	-9.27
69	-9.34	-10.2	-8.38	-8.92	-9.27
70	-9.78	-9.27	-8.46	-8.98	-9.26
71	-9.06	-8.78	-8.43	-8.81	-9.26
72	-9.08	-8.67	-8.12	-8.81	-9.25
73	-9.03	-8.94	-7.94	-8.98	-9.24
74	-9.31	-8.63	-8.65	-8.46	-9.24
75	-9.16	-9	-8.74	-9	-9.24
76	-9.08	-9.34	-8.27	-9.02	-9.24
77 (a-FABP-IN-1)	-9.58	-9.28	-9.07	-8.79	-9.23
78	-9.52	-9.26	-8.32	-8.57	-9.23

Continued on next page

Table 3.3: (continued)

Compound	FABP3	FABP4	FABP5	FABP9	FABP12
79	-9.55	-8.82	-8.85	-9.22	-9.21
80	-8.96	-9.08	-8.07	-8.96	-9.2
81	-8.93	-9.35	-8.61	-9.05	-9.2
82	-9.05	-8.46	-8.43	-9.02	-9.19
83	-9.85	-8.55	-8.4	-8.83	-9.19
84	-9.2	-8.83	-8.02	-9.45	-9.19
85	-9.61	-8.54	-8	-9.39	-9.19
86	-9.56	-8.94	-8.42	-8.78	-9.19
87	-8.96	-9.47	-8.61	-8.72	-9.18
88	-9.39	-9.27	-8.33	-9.56	-9.18
89	-9.47	-8.88	-8.09	-8.83	-9.18
90	-9.14	-9.47	-8.3	-8.83	-9.17
91	-9.93	-9.16	-9.23	-9	-9.17
92	-9.27	-9.51	-8.16	-8.67	-9.17
93	-9.66	-9.14	-8.24	-8.53	-9.17
94	-8.37	-8.8	-8.4	-8.48	-9.16
95	-9.47	-9.23	-8.94	-8.37	-9.16
96 (Compound 34)	-9.62	-9.1	-8.39	-8.28	-9.16
97	-9.6	-8.68	-7.92	-8.82	-9.16
98 (AM404)	-9.28	-9.01	-9.14	-9.24	-9.15
99	-9.2	-8.45	-8.47	-8.72	-9.15
100	-9.26	-8.38	-8.65	-8.82	-9.15
101	-9.04	-9.34	-8.56	-9.48	-9.15
102	-8.88	-9.01	-7.89	-8.24	-9.15
103	-9.36	-8.85	-8.55	-8.45	-9.15
104	-9.34	-9.09	-8.16	-8.73	-9.13
105	-8.98	-8.82	-8.79	-9.27	-9.13
106	-8.79	-8.55	-8.73	-8.58	-9.13
107	-9.08	-8.73	-7.96	-8.65	-9.13
108	-8.97	-9.33	-8.35	-8.89	-9.13
109	-9.23	-8.83	-8.19	-9.24	-9.12
110	-9.38	-9.15	-7.96	-9.54	-9.11
111	-8.45	-9.16	-8.5	-8.73	-9.11
112	-9.35	-8.17	-8.31	-8.95	-9.11
113 (BMS309403)	-9.11	-9.35	-8.35	-8.75	-9.1
114	-9.52	-9.5	-8.54	-8.82	-9.1
115 (ZTY28)	-8.95	-7.97	-8.39	-8.58	-9.09
116	-9.48	-8.8	-8.24	-8.61	-9.09
117	-9.73	-8.66	-8.26	-9.03	-9.09
118	-8.92	-8.13	-8.03	-8.83	-9.08
119	-9.06	-8.52	-8.32	-8.9	-9.08
120	-8.15	-8.84	-7.98	-8.08	-9.08
121	-9.19	-7.99	-8.44	-8.57	-9.07

Table 3.3: (continued)

Compound	FABP3	FABP4	FABP5	FABP9	FABP12
122	-8.84	-8.7	-7.91	-8.92	-9.07
123	-9.3	-9.39	-8.32	-9.51	-9.06
124	-8.92	-7.69	-7.62	-7.73	-9.06
125	-9.18	-8.54	-7.98	-8.98	-9.06
126	-9.14	-8.91	-8.18	-9.01	-9.05
127	-9.53	-9.08	-9.03	-9.67	-9.05
128	-9.65	-8.85	-8.58	-8.95	-9.04
129	-9.46	-8.05	-8.04	-8.27	-9.03
130	-8.96	-9.03	-7.75	-8.64	-9.03
131	-9.42	-7.29	-8.33	-8.84	-9.02
132	-9.49	-9.07	-8.81	-9.39	-9.02
133	-9.11	-9.17	-8.19	-8.84	-9.02
134	-9.59	-8.98	-8.33	-8.72	-9.02
135	-9.19	-8.77	-8.27	-8.58	-9.02
136	-9.79	-9.08	-8.14	-9.19	-9.02
137 (ART26.12)	-9.48	-8.91	-8.08	-9.30	-9.02

The second metric is the Selectivity Ranking, which is meant to rank the ligands based on the difference between the S Score obtained with FABP12 and the S Scores obtained with other FABPs. This can be displayed as a matrix where each cell is identified by a ligand (rows) and by the difference between the S Scores of FABP12 and a FABP other than 12 (columns). An example is reported in **Table 3.4**.

Compound	FABP3	FABP4	FABP5	FABP9
137 (ART26.12)	0.46	-0.11	-0.94	0.28

Table 3.4: Selectivity Ranking matrix row example. Every cell of the one-row table represents the difference between the S Score obtained with FABP12 docking and that column's FABP's S Score. For FABP4 and FABP5, the score is negative, indicating good selectivity for FABP12; the opposite is true for FABP3 and FABP9, where the difference is positive.

Each cell represents a numerical value that follows the S Score convention: the more negative the number in the cell, the more affinity the ligand has for FABP12 compared to that column's other FABP. If the number is positive, instead, the ligand has more affinity towards the other FABP. Sorting the columns of the matrix can lead to the calculation of a Selectivity Ranking for each ligand.

On top of that, an average Selectivity Ranking was then calculated by using the rankings of the compounds and averaging them. The higher the rank, the more likely the ligand will be unaffected by off-target interactions with other FABPs other than 12 when administered, and will, in other words, be more specific for FABP12. A summary of the Affinity Rankings and the Selectivity Rankings for the 15 highlighted compounds is reported in **Table 3.5**.

Table 3.5: Affinity and selectivity ranks of selected compounds.

Compound	Affinity Rank	Selectivity Rank
GW7647	2	6
OMDM2	26	156
STK-22	27	81
Compound 1	30	119
AM1172	38	160
SBFI-103	50	230
OMDM1	51	732
STK-15	53	609
Antraferine	54	460
α-FABP-IN-1	77	394
Compound 34	96	192
AM404	98	550
BMS309403	113	219
ZTY28	115	31
ART26.12	137	465

Chapter 4

ADMET Properties Prediction

4.1 Materials

4.1.1 ADMET Properties and physicochemical filters

While there are over 100 different types of possible drug administration routes, the most common ones include intravenous, topical, intramuscular, and especially oral administration, which is by far the most used [143]. It is important to understand, through the study of pharmacokinetics and pharmacology, how the drug interacts with the components of the body after administration: how far it can go from the administration site, how it gets excreted, how and if it manages to bypass the innate barriers of the body (like the endothelial and blood brain barriers), if it breaks down into other components after being metabolized and if it causes systemic or local toxicity to the organism [144].

These properties related to the pharmacological properties of drugs can be summarized by 5 descriptors, which in turn can signify many aspects of the drug, collected under the acronym ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) [145]. These properties are directly related to various physicochemical descriptors that can be calculated by computational software and used by algorithms that predict the ADMET risk associated with the ligands [146].

It is the failings in the developed drugs related to these characteristics that limit their ability to transition from computational studies to preclinical and clinical trials thereafter, which is translated into their high attrition rates [147]. Rational drug design development makes use of physicochemical and ADMET filters to ensure that the compounds in the ligand database that will be used for docking or the ones that have been designated as lead compounds after the simulations are, in

fact, similar to most other available drugs on the market. One of the oldest filters that can be used to evaluate the drug-likeness of a compound is Lipinski's Rule of 5, which identifies whether the potential drug would be orally active [148]. This set of empirical rules is called so due to the heavy presence of the number 5 in the rules themselves:

- Molecular Weight < 500 g/mol;

- Number of rotatable bonds < 10;

- Octanol partition coefficient (log P) < 5;

- Number of hydrogen bond acceptors < 10 and hydrogen bond donors < 5;

- (Optional) Total Polar Surface Area < 140Å.

These rules, however, are generally only indicative of a good candidate. In fact, Lipinski's Rule of 5 itself is only valid for about 50% of orally administered compounds. They also cannot be used by themselves to evaluate whether the drug is pharmacologically active towards any given target, and as such, they need to be accompanied by docking simulations.

4.1.2   Absorption

Drugs travel from the administration site to the site of interest, and can do so under different forms (tablets, solutions, etc.). Drug absorption is impacted not only by the dosage and the form in which drugs are delivered, but also by their characteristics and route of administration. Unless an intravenous route has been used, virtually every possible drug will have to cross several different cell membranes before it is able to reach systemic circulation.

Cell membranes consist of a phospholipid bilayer, made by amphiphilic phospholipids that orient their hydrophilic head towards the exterior and interior of the cell, while the hydrophobic tails lie at the center of the membrane. This characteristic allows cells to block the passive transport of hydrophilic and polar large molecules, while small molecules and ions can diffuse through the bilayer.

Passive transport and Octanol partition coefficient   Passive transport consists of diffusion powered by a gradient of concentration, which nudges molecules in the direction of low concentration without requiring additional expenditure of energy or secondary actors (e.g., ion pumps, transmembrane proteins). This phenomenon can be described by the log P descriptor, also known as the partition coefficient, which is defined as the ratio of the concentrations of a solute between

65

aqueous and lipophilic phases (in logarithmic terms), which are water and octanol, as indicated by the formula:

$$\log P = \log [\text{drug}]_{\text{octanol}} - \log [\text{drug}]_{\text{water}} \tag{4.1}$$

Log P is one of the most common physicochemical descriptors for chemical compounds, and most commercially available drugs possess a log P value ranging from 1 to 6. The higher the log P, the higher the affinity for the lipidic phase and the easier it is for a molecule to bypass the cell membrane without requiring active transport.

It is also possible to account for the effect of pH when trying to predict the hydrophobicity of a drug: the log D descriptor, called the distribution coefficient, depends on the pH of the aqueous phase, which could matter in situations where the drug has to exert its effect in areas of the body where pH is more acidic, as happens with cancer.

Total Polar Surface Area Another descriptor that is often used to predict the hydrophobicity of chemical compounds and the ability to permeate the cell membrane is the Polar Surface Area (PSA), which is defined as the sum of the surface area of all polar atoms alongside hydrogens in the molecule. Calculation of the PSA can be time-consuming since it requires the generation of the 3D model of the molecule and subsequent minimization of the energy of its surface.

As such, computational predictors often employ the use of Topological PSA (TPSA), which uses a fragment-based calculation method where the molecule is split up into many parts, each tied to a TPSA contribution based on a fragment library. When the TPSA is under 140Åthe drug will likely be able to pass through cell membranes, while to cross the blood-brain barrier, a TPSA of under 90Åis needed.

Other parameters related to Absorption are skin permeability, gastrointestinal absorption, water solubility, and interactions with active transport transmembrane proteins. The first two, like their name suggests, can be useful to evaluate the efficacy of topical and oral drugs. The water solubility (log S) of a compound is calculated as the base-10 logarithm of its molar solubility in water (mol/L); highly water-soluble drugs are usually better absorbed.

4.1.3 Distribution

Once the drugs have been absorbed and have entered the systemic circulation, they are transported around the body by the circulatory system. The amount of drug that has entered systemic circulation post-administration is indicated as the bioavailable portion of the dose. Depending on the original administration route,

the bioavailability of a drug can be penalized (in the case of oral administration) or ensured if the drug is introduced via intravenous injection.

Oral drugs also need to go through the liver, which can further reduce the bioavailable portion due to its metabolism. The most known indicator for a highly bioavailable oral drug is the previously mentioned Lipinski's Rule of 5. Due to Lipinski's Rule of 5 only being valid for 50% of accepted drugs, sets of less stringent rules and alternative variants have been designed. Among these are the Veber Filter, the Ghose Filter, the Egan Filter, and the Muegge Filter.

After absorption, the drugs will be carried around the body thanks to the vasculature and the plasma contained therein, before reaching the site of action. When found in the plasma, drugs can either assume two forms: one is their standard free form, and the other is in conjunction with plasma proteins. This second form is reversible; however, drugs in this state possess no pharmacological activity, thus reducing the bioavailable amount.

The amount of unbound drug in the circulatory system is called Fraction Unbound and can be predicted using ADMET property predictors. Drugs can also accumulate in interstitial spaces in tissues; if the accumulation is too high, the drug could lead to toxicity in the body. However, it could be that the compound is designed to persist in the body for a longer time. In fact, it is important that the drug, during its therapeutic window, has an equal rate of clearance to its steady-state concentration, which in turn is the amount of the drug that needs to be uniformly distributed in the tissues to equal the amount of the drug in circulation.

4.1.4 Metabolism

Drugs that are absorbed by the body are often metabolized by a variety of enzymes into secondary products or metabolites. This event can happen at any point after administration, both before and after reaching the site of action. This process creates substances that are generally more polar and soluble in water, which means that it's easier for the body to eliminate them in the urine. The effectiveness of drugs can also be modified post-metabolism: they can either gain or lose their pharmacological properties and become active or inactive compounds. Metabolic reactions are usually split into 2 categories, called Phase I and Phase II.

The former includes changes in the functional groups of substances thanks to oxidation, reduction, carboxylation, and hydrolysis, and is also called the functionalization phase of metabolism. The majority of Phase I reactions happen in the liver and disable an already active drug. Phase II reactions, instead, mostly consist of conjugation reactions (e. g. sulfate, glutathione, glucuronide) which produce highly soluble metabolites. Methylation and acetylation, however, have the opposite effect.

Phase I oxidation reactions are catalyzed by cytochrome 450 (CYP). CYPs are a superfamily of enzymes that contains over 100 isoforms, 12 of which are of particular interest in ADMET prediction due to their importance in the metabolism of xenobiotics. Chemical compounds that have high lipophilicity also possess a high chance of being metabolized by CYPs. It is important to try and elucidate through ADMET predictors whether the drug is a substrate of CYPs and whether it can act as their inhibitor, to avoid systemic toxicity due to accumulation.

4.1.5 Excretion

Excretion is the elimination of metabolites and waste from the body. The excretory system can exert its function through different ways and organs, though the urinary system is the most involved one. The kidneys eliminate waste collected in the form of urine, formed from the bloodstream, and regulate homeostasis of extracellular fluids. Most drugs are eliminated this way, either in their unchanged form or as a structurally different metabolite; however, some drugs can be excreted through the bowel in the form of feces or during biliary excretion thanks to the liver. Clearance is necessary to avoid systemic toxicity of the drug, but it also shouldn't be too high, as the therapeutic effect of the drug requires a build-up of concentration in the bloodstream to function properly.

4.1.6 Toxicity

The potential toxicity of a drug is tied to all four previous macro categories of pharmacological descriptors, as toxicity can be derived due to the accumulation of the compound in the tissues and in the organs or due to metabolite-induced toxicity caused by toxic byproducts. The term toxicity refers to damage caused to cells and their components, but also to their phenotype and disruptions in the regulatory pathways of the body. There are many different types of ADMET-related toxicity tests and descriptors, derived both from animal and human assays [149]. Among these are:

- **AMES** and **T. Pyriformis toxicity**, which predict whether a given compound is toxic based on bacteria-related inhibitory and mutagenic tests;

- **Rat acute toxicity**, which measures the Lethal Dose 50 (LD50) of the drug, which is the concentration of the drug that needs to be administered to a group of animals to cause the death of half of that group;

- **Maximum Tolerated Dose** and **Maximum Therapeutic Dose**, both are human descriptors that define acceptable concentrations of the drug to be used. The former indicates the highest amount of the drug that can be administered

without incurring severe side effects due to toxicity, while the latter indicates the theoretical limit within which an increase in the concentration of the drug is correlated to an increase in its therapeutic effect.

- **Hepatotoxicity** and **Cardiotoxicity**, still human descriptors, but related to the specific functionality of organs. Hepatotoxicity occurs when side effects due to the drug's administration cause damage in the liver and in hepatocytes, resulting in dysfunction of the normal function of the organ. Cardiotoxicity, instead, is probably one of the most common points of failure of a novel compound and a potential marketable drug. If a drug is cardiotoxic, it can inhibit the potassium ion channels present in cardiomyocytes, which can result in cardiac dysfunction. These channels are also called hERG channels due to the human Ether-à-go-go gene, which encodes the protein, and hERG inhibition is considered a major health risk for ventricular arrhythmia.

4.1.7   ADMET Predictors

As the prediction of the ADMET properties and descriptors of a potential drug is nowadays essential in ensuring the pharmacological activity and lack of toxicity in the human body, a wide array of software has been developed to calculate the necessary descriptors to evaluate potential risks in any given compound.

These predictors are purely based on in silico computational tools that employ machine learning and artificial intelligence algorithms to perform the job without the need to synthesize or buy the compound, and rely on in vitro experiments [150], [151]. In fact, the only requirement necessary to run the predictions is a database that contains the SMILES of the compounds that are to be tested. These methods do not necessitate protein structures either, so they can be used at any point in drug discovery, either to weed out compounds if the initial database is fairly sizable or to analyze potential lead compounds more in depth.

Fully integrated ADMET prediction in virtual screening studies can vastly reduce the economic strain placed on the synthesis of novel compounds for testing, especially avoiding late-stage failures in clinical trials.

Machine learning applications make use of various methods (e.g., regression models, artificial neural networks, random forest, k-nearest neighbor) in order to apply pattern recognition to newly introduced elements of a dataset. These elements will be compared to the ones contained in a training set, which, instead, has been used to train the model using chemical compounds for which the ADMET descriptors have been experimentally quantified in some way.

More specifically, the predictors make use of quantitative structure-relationship activity (QSAR) and quantitative structure-property activity (QSPR) models, which can be built using the training sets through regression and classification

analyses of the compounds' physico-chemical properties [152]. Recently, ensemble methods or multiple classifier systems have been used to predict information related to unbalanced datasets by employing multiple individual models, which overall reduce the potential bias of datasets where a given class of compounds is expressed more than other classes [153].

ADMET predictor tools are available both in web-based form, like SwissADME [154], pkCSM [155], ADMETLab 3.0 [153], CypReact [156], vNN [157], and others, or through commercially available software, such as ADMET Predictor from Simulations Plus [158] and ChemAxon's Calculators and Predictors.

4.2 Methods

For the purpose of analyzing the chemical compounds of the ligand database used to perform molecular docking simulations for FABP inhibition, SwissADME, pkCSM, and ADMET Predictor were used for the prediction of the molecular descriptors and ADMET properties of the compounds themselves. While the prediction was performed on the entire database, in-depth analysis will only be reported for the 12 commercially available compounds mentioned in Section 3.3, alongside BMS309403, ZTY28, and ART26.12.

4.2.1 SwissADME

SwissADME is a web-based tool developed by the Molecular Modelling Group of the Swiss Institute of Bioinformatics and is part of a bigger suite of computational tools called SwissDrugDesign. The other tools can perform docking based on AutoDock Vina (SwissDock), parametrization of small ligands for applications in force fields (SwissParameter), find similar ligands to an input one and virtual screen a database of compounds (SwissSimilarity), and estimate the most probable macromolecular targets of small ligands based on known libraries of actives and proteins (SwissTargetPrediction).

The two most known features of SwissADME are the graphical representation in form of a radar plot of the bioavailability of a compound, and the boiled-egg plot, which, as the name suggests, takes the form of an egg in which the molecules placed in the yellow core are expected to cross the blood brain barrier and the ones in the egg white can be absorbed in the gastrointestinal tract. An example of the boiled-egg plot can be seen in **Figure 4.1**.

The bioavailability plot, shown in **Figure 4.2**, is composed of 6 main parameters, and its central pink area defines the confines within which a ligand possesses good characteristics. The six parameters, alongside their optimal ranges, are:

- **LIPO**: lipophilicity of the compound, measured in terms of XLOGP3 (which is a predicted value of log P), between -0.7 and 5.0;

- **SIZE**: expressed in Molecular Weight, between 150g/mol and 500g/mol;

- **POLAR**: polarity of the compound expressed in terms of TPSA, between 20 and 130 A2;

- **INSOLU**: insolubility, measured with log S, between -6 and 0;

- **INSATU**: insaturation, expressed as the fraction of Cps3, also known as the ratio of sp3-hybridized carbon atoms in a molecule to the total amount of carbon atoms, between 0.25 and 1;

- **FLEX**: flexibility of the compound, if it possesses between 0 and 9 rotatable bonds within its structure.
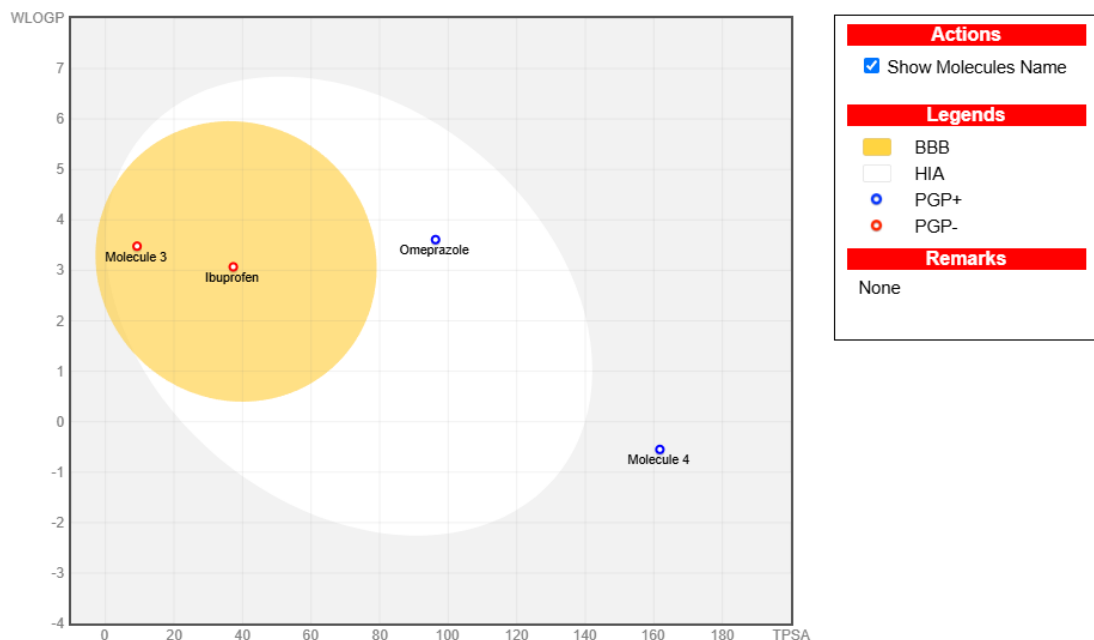
Figure 4.1: Boiled-egg plot. Four different molecules are placed in the plot area, two of which are inside the egg yellow (Molecule 3 and Ibuprofen) and are BBB permeant, one inside the egg white (Omeprazole), which can be absorbed in the gastrointestinal tract, and one (Molecule 4) outside of the egg.

SwissADME, however, can go into a lot more detail when it comes to physico-chemical descriptors and drug-like properties and filters. The tool possesses 6 other categories which detail physicochemical descriptors, lipophilicity, water solubility, pharmacokinetics, druglikeness, and medicinal chemistry.

Physicochemical Descriptors   The physicochemical properties of a compound are purely based on its structure and molecular formula and, as such, require no prediction, only a calculation. Among these, SwissADME highlights the molecular weight (g/mol), the number of heavy atoms and aromatic heavy atoms, the fraction of Csp3, the number of rotatable bonds, hydrogen bond acceptors and donors, the molar refractivity, and the TPSA.
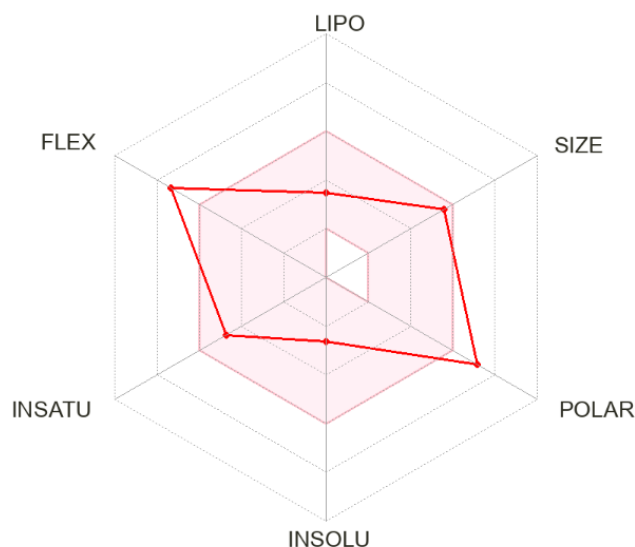
Figure 4.2: Bioavailability plot of a compound. The six descriptor categories are displayed as the spokes of the radar plot. If the molecule's parameters are confined within the pinkish area, it means that it possesses good bioavailability characteristics.

Lipophilicity A consensus value of log P is obtained by averaging five other lipophilic descriptor predictors:

- **iLOGP**: implicit log P calculated using an in-house, physics-based method, which is constructed on the generalized born/surface area (GB/SA) implicit solvent model;

- **XLOGP3**: an atomistic and knowledge-based model that predicts the value of log P of a compound by comparing the structure with other compounds of which the log P is experimentally derived;

- **WLOGP**: an atomistic predictor based on an atom type classification method;

- **MLOGP**: a topological method obtained by multiple regression analyses of datasets of compounds;

- **SILICOS-IT**: a hybrid model using fragmental and topological methods implemented in the FILTER-IT program;

Water Solubility The water solubility of a molecule, expressed in terms of log S, classifies whether the molecule itself is insoluble or highly soluble based

on a range of values between -10 and 0. In particular, SwissADME performs this classification based on the prediction of three different log S values, two of which are topological-based methods (ESOL and Ali) and one is fragment-based (SILICOS-IT). The class is then assigned based on the ESOL method, as reported in the bioavailability radar plot, and is determined as follows:

Insoluble < -10 < Poorly < -6 < Moderately < -4 < Soluble < -2 < Very < 0 < Highly

Pharmacokinetics   The Pharmacokinetics section highlights whether a molecule is permeant to the blood-brain barrier and if it's absorbed by the gastrointestinal tract, just like the boiled-egg plot. In addition, it provides information on the permeability to the skin based on a QSPR model and whether the compound is an inhibitor to the most important CYPs (1A2, 2C19, 2C9, 2D6, and 3A4).

The models implemented for CYP inhibition prediction are support vector machine models, which are none other than machine learning classifiers based on datasets containing several thousand compounds. Lastly, the pharmacokinetics section indicates if the ligand is a substrate for P-glycoprotein, which is an ATP-dependent efflux pump located in the membrane of all cells and especially in the intestinal epithelium, and transports various kinds of molecules across the extracellular and intracellular membranes.

Druglikeness   The druglikeness of a compound describes how close the characteristics of the analyzed ligand are to a series of filters. These filters have been mentioned before and include the Lipinski's rule of 5 and the Ghose, Veber, Egan, and Muegge filters. This section also includes a bioavailability score, which is defined as the probability that a compound will have >10% bioavailability in rats.

Medicinal Chemistry   The last section includes some pseudo-descriptors that are helpful to identify chemical compounds with high synthesizable potential.

Compounds are initially catalogued based on whether they are pan-assay interference compounds or not; these are chemical compounds that often count as false negatives in high-throughput virtual screening procedures because they tend to react with a vast number of molecular targets. The Brenk filter evaluates whether the compound contains one out of 105 different molecular fragments that can interfere with the normal metabolism of the drug and render it more toxic.

A leadlikeness filter is implemented to identify potential candidates for lead compounds, which are good starting points for subsequent virtual screens or ligand derivatization studies. This filter includes 3 descriptors and ranges within values are acceptable for a good lead compound: molecular weight between 250 and 350 g/mol, XLOGP3 less than 3.5, and less than or equal to 7 rotatable bonds. The last score is the synthetic accessibility, which is a number that ranges from 1 (easily

74

synthesized) to 10 (very difficult synthesis) and is based on a linear regression model built with over 1000 fragments and ten million molecules.

4.2.2 pkCSM

pkCSM is a web-based tool that allows the upload of a text-based file containing multiple SMILES strings for the prediction of characteristics related to ADMET of chemical compounds. This tool relies on graph-based signatures, which encode distance patterns between atoms and are used to represent the molecules [155]. These graphs allow for the decoration and denotation of specific characteristics of the compounds (from physicochemical to structural properties) to build predictive models using machine learning and regression classifiers. The models are separated into five classes, which are none other than the five letters that compose the ADMET acronym. These contain (but are not limited to) the following predictive models:

- **Absorption**: Water Solubility, Intestinal Absorption, P-glycoprotein Substrate and Inhibitor, Caco2, and Skin permeability

- **Distribution**: Fraction unbound, Volume of distribution at steady state, BBB, and Central Nervous System permeability;

- **Metabolism**: 7 CYP enzyme inhibitor predictors;

- **Excretion**: Total clearance and Renal OCT2 substrate;

- **Toxicity**: AMES, *T.Pyriformis*, Minnow, Rat (acute and chronic) toxicities, alongside Hepatotoxicity, Cardiotoxicity (hERG I and II inhibitor), and Maximum Tolerated Dose.

4.2.3 ADMET Predictor

ADMET Predictor is a commercially available software that is part of a suite of tools developed by Simulations Plus, encompassing various aspects of drug design. The ADMET Predictor module allows for the examination of ADMET properties related to chemical compounds through the prediction and QSAR analysis of more than 170 descriptors. Just like SwissADME, the only requirement needed to run the program is a SMILES database in Excel format to serve as an input. The software also allows the user to build QSAR and QSPR models for descriptors that are not included in the module.

The most interesting feature of ADMET predictor is the aggregate ADMET risk score, which takes into account the various descriptors calculated regarding absorption, distribution, metabolism, mutagenicity, and toxicity and summarizes

them in a score that ranges from 0 to 22. Compounds with an ADMET risk score of above 7 are generally considered at potential risk of failing during or before clinical trials. This value is not strict, as it is a cutoff value where out of 2260 commercial drugs, 10% of them possess a value of ADMET risk above 7. The program also integrates substrate, metabolism, and kinetic predictions to generate maps of likely metabolites.

The models and descriptors calculated by the software are split into 4 major categories, plus a fifth one, which is the risk summary based on the findings contained in the other ones. These are:

- **Physicochemical and Biopharmaceutical** (PhysChem), among which are models for ionization and dissociation constants (pKa), lipophilicity, solubility in water, octanol, and other intestinal fluids, permeability in water and through the skin, cornea, and blood-brain barrier, and antiviral infection models for inhibition of HIV-1.

- **Metabolism**, which has substrate classification models for nine human CYPs (1A2, 2A6, 2B6, 2C8, 2C9, 2C19, 2D6, 2E1, and 3A4), alongside other descriptors related to the estimated clearance of predicted metabolic sites of the CYPs mentioned above. It also contains a model for the substrate classification of phase II metabolic enzymes of the UGT family.

- **Transporter**, which includes prediction of efflux/inhibition of P-glycoprotein and substrate/inhibition for various other transport proteins (OAT1, OCT1, OAT3, OCT2, BCRP, OATP1B3, BSEP).

- **Toxicity**, wherein are models for prediction of reproductive and genetic toxicity, rodent toxicity, human toxicity (cardiotoxicity and hepatotoxicity), excitotoxicity, Maximum Tolerated Dose, and Maximum Therapeutic Dose.

The risk section and ADMET risk score are comprised further of three subcategories, which are defined by risk rules that can be violated or not, depending on the calculated descriptors. As mentioned, the total score, also called ADMET_risk, can be 22, as it incorporates 20 rules from the three subcategories (8 for absorption risk or ABSN_risk, 6 for metabolism risks or CYP_risk, and 6 for toxicity risk or TOX_risk) plus another 2 miscellaneous rules regarding the fraction unbound and the steady-state volume of distribution of the compound.

Each rule possesses a mnemonic code associated with it so that every violation can easily be seen in the output file, on top of the risk score associated with the 4 risk values. ADMET predictor also offers the possibility of generating custom rules or modifying the existing ones if not relevant to the study at hand. Codes can be accompanied by a + or a -: this is because rules are based on fuzzy thresholds and not hard cutoffs.

Table 4.1: Summary of ADMET Predictor risk codes; thresholds, and rules are not included for brevity.

Category	Rule / Thresholds
ABSN Risk	MW: Molecular Weight too high
	RotB: Too many rotatable bonds
	HBD: Too many H-bond donors
	HBA: Too many H-bond acceptors
	ch: Too much charge (evaluated in terms of TPSA)
	Kow: Too lipophilic
	Peff: Low permeability
	Sw: Low solubility
CYP Risk	1A2: High CLint (In-vitro clearance) of CYP1A2
	2C9: High CLint of CYP2C9
	2C19: High CLint of CYP2C19
	2D6: High CLint of CYP2D6
	3A4: High CLint of CYP3A4
	CL: High CLint measured for liver microsomes or hepatocytes (2 rules)
TOX Risk	hERG: Cardiotoxic liability for hERG
	rat: Acute toxicity in rats
	Xr: Carcinogenicity in chronic rat studies
	Xm: Carcinogenicity in chronic mouse studies
	HEPX: Hepatotoxicity
	MUT: Bacterial mutagenicity
Other Rules	fu: Low fraction unbound
	Vd: High steady-state volume of distribution

4.3   Results

For all comparative results of the ADMET properties prediction of the 15 compounds, ADMET Predictor results will be used as a baseline, since it is the tool with the highest performance and the most reliable. It is possible that the results obtained from the three predictors do not reach a consensus; this is to be expected, since every predictor is based on their own implementation of regression classifiers, machine learning methods, and characteristics calculations. Results then should be interpreted using critical thinking and can only be useful as a general guideline before the application of in-vitro experimental testing.

Absorption & Distribution   **Table 4.2** shows the predicted water solubility (log S) values of all three predictors, with the middle 3 being all 3 methods implemented in SwissADME. ADMET Predictor and pkCSM are the ones that agree the most on the overall solubility of the molecules, with SwissADME's predictors denoting most of the molecules as poorly soluble or insoluble.

Table 4.2: Predicted logS values for selected compounds.

Compound	S. Plus	ESOL	Ali	Silicos-IT	pkCSM
GW7647	-3.66	-7.21	-9.86	-7.16	-3.78
OMDM2	-3.45	-6.38	-9.31	-8.14	-5.51
STK-22	-4.38	-7.38	-8.26	-10.57	-4.62
Compound 1	-3.13	-6.04	-6.93	-8.32	-4.28
AM1172	-3.51	-6.04	-6.93	-8.32	-6.39
SBFI-103	-5.80	-6.75	-7.44	-9.68	-4.90
OMDM1	-3.45	-6.38	-9.31	-8.14	-6.04
STK-15	-5.28	-7.41	-8.32	-10.18	-4.30
Antrafenine	-3.90	-7.96	-8.58	-10.55	-4.49
a-FABP-IN-1	-5.92	-8.45	-10.21	-10.88	-4.25
Compound 34	-5.93	-6.70	-9.02	-6.64	-3.02
AM404	-2.47	-6.70	-9.02	-6.64	-3.99
BMS309403	-5.44	-7.39	-8.37	-10.64	-4.31
ZTY28	-6.80	-7.27	-8.09	-10.76	-4.52
ART26.12	-3.40	-5.70	-6.35	-7.39	-3.40

P-glycoprotein inhibition predicted by ADMET Predictor indicates that every compound except Compound 34 is an inhibitor of the cell membrane transporter. While SwissADME results are not available for this, pkCSM agrees with ADMET Predictor for all compounds except for GW7647, Compound 1, SBFI-103, and ART26.12. P-glycoprotein substrate predictions can be found in **Table 4.3**.

Gastrointestinal absorption, as measured by pkCSM, does not fall under 85% for any compound. Considering that its threshold for poor absorption is <30%, all compounds are expected to perform well in this field. SwissADME also has a low/high threshold for the same prediction; however, it is based on whether the compound falls out or in the egg white of the BOILED-Egg plot, and as such, it is not as reliable for edge cases.

BBB permeation data is not included due to relevance to the clinical case; however, ADMET Predictor and pkCSM mostly agree with one another, with Antrafenine being the only compound able to easily cross the BBB. Antrafenine is an FDA-approved drug that is, in fact, used for analgesic purposes in the central nervous system, so this prediction is validated by a real use case.

Table 4.3: P-glycoprotein substrate predictions.

Compound	S. Plus	SwissADME	pkCSM
GW7647	Yes	Yes	Yes
OMDM2	Yes	No	Yes
STK-22	Yes	No	Yes
Compound 1	Yes	Yes	Yes
AM1172	Yes	Yes	Yes
SBFI-103	Yes	No	No
OMDM1	Yes	No	Yes
STK-15	Yes	No	Yes
Antrafenine	No	Yes	Yes
a-FABP-IN-1	Yes	Yes	Yes
Compound 34	Yes	Yes	Yes
AM404	Yes	Yes	Yes
BMS309403	Yes	Yes	Yes
ZTY28	Yes	Yes	No
ART26.12	Yes	Yes	No

Both ADMET Predictor and pkCSM calculate the volume of distribution at steady state (Vd) and the fraction unbound (fu) of the chemical compounds (**Tables 4.4** and **4.5**).

Unfortunately, it is difficult to obtain conclusive evidence from the results provided by the two predictors. According to pkCSM, values of the Vd should be contained between -0.15 and 0.45, which is true for the ADMET Predictor evaluation excluding 5 compounds, while only 4 molecules subscribe to this rule for pkCSM.

Furthermore, according to ADMET Predictor, the fu is considered low if it is less than [4, 6]%, which is true for all compounds; pkCSM disagrees with this, and

provides generally different numbers. These discrepancies could be attributed to the different types of training datasets used that might not be representative of the compounds analyzed, but more information is needed to reach this conclusion.

Table 4.4: Volume distribution at steady state predictions.

Compound	S. Plus	pkCSM
GW7647	0.29	-1.17
OMDM2	1.50	0.13
STK-22	0.35	-1.27
Compound 1	0.33	-1.40
AM1172	1.29	0.27
SBFI-103	0.28	-2.46
OMDM1	1.50	0.26
STK-15	0.34	-1.49
Antrafenine	7.26	0.85
a-FABP-IN-1	0.32	-1.06
Compound 34	0.25	-0.95
AM404	1.15	0.24
BMS309403	0.26	-1.09
ZTY28	0.30	-1.57
ART26.12	0.29	-1.66

Table 4.5: Fraction unbound predictions.

Compound	S. Plus	pkCSM
GW7647	1.3%	0%
OMDM2	1.5%	11%
STK-22	1.5%	5%
Compound 1	1.2%	6%
AM1172	1.2%	3%
SBFI-103	0.9%	15%
OMDM1	1.5%	11%
STK-15	1.3%	5%
Antrafenine	1.3%	0%
a-FABP-IN-1	0.4%	30%
Compound 34	0.9%	11%
AM404	1.3%	14%
BMS309403	0.8%	31%
ZTY28	0.6%	17%
ART26.12	1.6%	0%

Table 4.6: Absorption risk prediction of ADMET Predictor for the 15 compounds.

Compound	ABSN_Risk	ABSN_Code
1) GW7647	5	Size; RotB; ch; Kow; Sw
2) OMDM2	4	Size; RotB; ch; Kow
3) STK-22	3	Size; Kow; Sw
4) Compound 1	1.8	Size; Kow
5) AM1172	3	Size; RotB; Kow
6) SBFI-103	3.5	Size; RotB; Kow; Sw
7) OMDM1	4	Size; RotB; ch; Kow
8) STK-15	3	Size; Kow; Sw
9) Antrafenine	3.5	Size; RotB; Kow; Sw
10) a-FABP-IN-1	3	Size; Kow; Sw
11) Compound 34	4	Size; ch; Kow; Sw
12) AM404	3.8	Size; RotB; Kow; Sw
13) BMS309403	3	Size; Kow; Sw
14) ZTY28	3	Size; Kow; Sw
15) ART26.12	2.7	Size; Kow; Sw
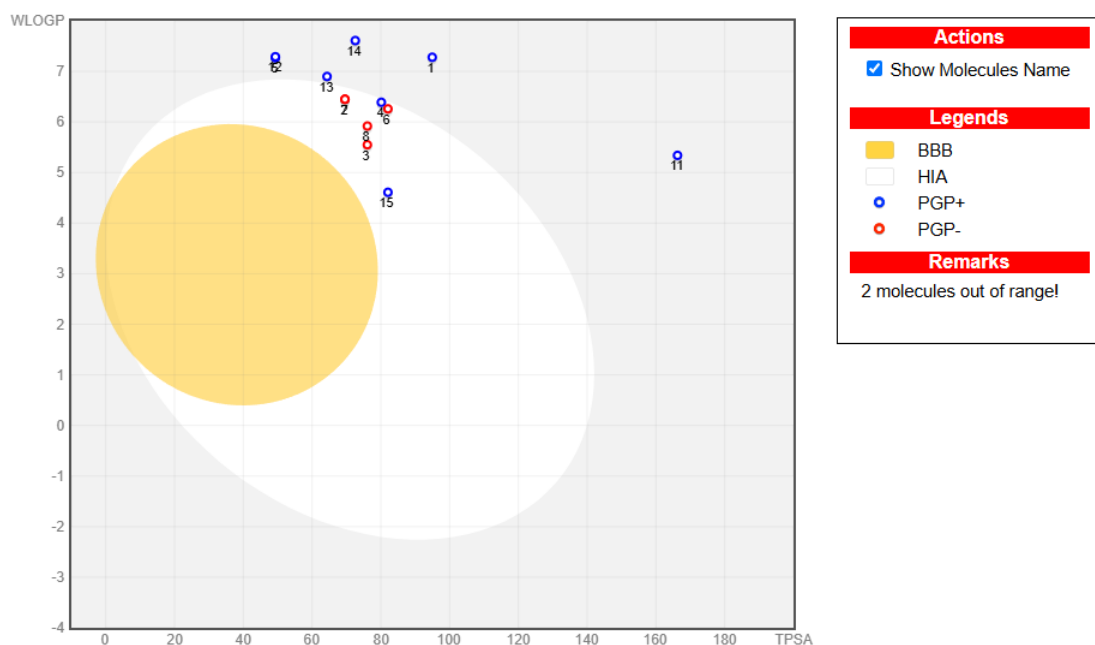
Figure 4.3: BOILED-Egg plot of the 15 compounds. Molecules 9 and 10 are out of range due to their WLOGP being >8. Compounds are indicated with numbers instead of their names for readability (refer to the above table).

Table 4.6 shows the risk scores and codes associated with the absorption of each of the 15 compounds; decimal values are a byproduct of the fuzzy thresholds for rule violations. ABSN_risk is higher than 4 for 9% of the compounds used in the training set for ADMET Predictor construction, and as such can be used as a good metric for potential lead compounds. Most of the 15 compounds highlighted possess a score that is borderline or lower, with only GW7647 clearly breaking too many violations.

SwissADME summarizes all relevant information for absorption and distribution in the BOILED-Egg plot, reported in **Figure 4.3**. Molecules inside the yellow area would be able to cross the BBB, while the ones in the egg white can be absorbed by the gastrointestinal tract. Since this metric is based on TPSA and WLOGP, it is more reliable to use pkCSM and ADMET Predictor, but it still provides a qualitative metric that is visually representative of the dataset. The blue and red circles, on the other hand, represent whether a compound is a substrate of P-glycoprotein or not.

Metabolism All three different predictors provide information regarding multiple and different CYP enzymes, but all three of them highlight whether the chemical compounds inhibit 5 enzymes of the CYP family (CYP1A2, CYP2C9, CYP2C19, CYP2D6, and CYP3A4). The three predictors reach a consensus most of the time, with ADMET Predictor and pkCSM agreeing with each other the most (**Table 4.7**).

As with absorption, ADMET Predictor sums up its predictions regarding the metabolism of the chemical compounds with risk scores and codes, as shown in **Table 4.8**. CYP_risk is higher than 2 for 9% of the compounds used in the training set for construction of the ADMET Predictor. Here, only a few compounds pass that threshold, meaning they would probably be metabolized too quickly before being able to exert their function on the body.

Excretion SwissADME does not provide any predictions related to the excretion of the compounds. Both ADMET Predictor and pkCSM describe whether the compounds are renal uptake Organic Cationic Transporter-2 (OCT2) inhibitors, as shown in **Table 4.9**. Both predictors agree in almost all cases, with two exceptions. Additionally, pkCSM provides some info regarding the renal clearance of the compounds, as shown in **Table 4.10**.

Table 4.7: CYP enzymes inhibition predictions.

Compound	CYP1A2	CYP2C19	CYP2C9	CYP2D6	CYP3A4
GW7647	🔴🟥🔺	🔴🟥🔺	🟢🟥🔺	🔴🟥🔺	🟢🟩🔺
OMDM2	🔴🟥🟢	🔴🟩🔺	🔴🟥🔺	🟢🟩🔺	🔴🟩🔺
STK-22	🔴🟥🔺	🔴🟩🔺	🟢🟥🔺	🔴🟥🔺	🔴🟩🔺
Compound 1	🔴🟥🔺	🔴🟥🔺	🟢🟥🔺	🔴🟥🔺	🔴🟥🔺
AM1172	🔴🟩🔺	🔴🟥🔺	🔴🟥🔺	🔴🟥🔺	🔴🟥🔺
SBFI-103	🔴🟥🔺	🟢🟥🔺	🟢🟥🔺	🔴🟥🔺	🔴🟥🔺
OMDM1	🔴🟥🔺	🔴🟥🔺	🔴🟥🔺	🔴🟥🔺	🔴🟩🔺
STK-15	🔴🟥🔺	🔴🟩🔺	🟢🟩🔺	🔴🟥🔺	🔴🟩🔺
Antrafenine	🔴🟥🔺	🟢🟥🔺	🔴🟥🔺	🟢🟩🔺	🟢🟩🔺
a-FABP-IN-1	🔴🟥🔺	🔴🟥🔺	🔴🟥🔺	🔴🟥🔺	🔴🟥🔺
Compound 34	🔴🟥🔺	🟢🟥🔺	🔴🟩🔺	🔴🟥🔺	🟢🟥🔺
AM404	🔴🟥🔺	🔴🟥🔺	🟢🟥🔺	🔴🟥🔺	🔴🟩🔺
BMS309403	🔴🟥🔺	🔴🟩🔺	🔴🟥🔺	🔴🟥🔺	🔴🟥🔺
ZTY28	🔴🟥🟢	🔴🟥🔺	🟢🟥🔺	🟢🟥🔺	🔴🟥🔺
ART26.12	🔴🟥🔺	🟢🟥🔺	🟢🟥🔺	🔴🟥🔺	🔴🟥🔺

Table 4.8: Metabolism risk prediction of ADMET Predictor for the 15 compounds.

Compound	CYP_Risk	CYP_Code
GW7647	2.9	2D6; 3A4; CL
OMDM2	1	CL
STK-22	2.1	2C9; 3A4; CL
Compound 1	0	-
AM1172	3	2C9; 2D6; CL
SBFI-103	0.5	CL-
OMDM1	1	CL
STK-15	2.4	2C9; 3A4; CL
Antrafenine	2.3	2D6; 3A4; CL
a-FABP-IN-1	0	-
Compound 34	1	2C9-; 2D6-
AM404	2	2D6; CL
BMS309403	1	2C9
ZTY28	1.8	2C9; CL
ART26.12	0	-

Table 4.9: OCT2 inhibition predictions.

Compound	S. Plus	pkCSM
GW7647	No	No
OMDM2	No	No
STK-22	No	No
Compound 1	No	No
AM1172	Yes	No
SBFI-103	No	No
OMDM1	No	No
STK-15	No	No
Antrafenine	Yes	Yes
a-FABP-IN-1	No	No
Compound 34	No	No
AM404	Yes	No
BMS309403	No	No
ZTY28	No	No
ART26.12	No	No

Table 4.10: Renal clearance prediction, expressed in log(ml/min/kg).

Compound	pkCSM
GW7647	0.13
OMDM2	1.95
STK-22	0.23
Compound 1	0.10
AM1172	2.00
SBFI-103	0.27
OMDM1	1.85
STK-15	0.20
Antrafenine	0.46
a-FABP-IN-1	0.16
Compound 34	0.56
AM404	1.85
BMS309403	0.88
ZTY28	0.02
ART26.12	0.49

Toxicity Both ADMET Predictor and pkCSM report different descriptors related to the prediction of compound toxicity in humans and other animals. One thing to evaluate before that is the potential mutagenicity of the chemicals, which can be assessed through the Ames test, an in vitro method that utilizes bacteria to determine whether the compound may be carcinogenic. Both predictors implement their own variation of the Ames test performed through machine learning models, and their results are mostly the same (**Table 4.11**).

Table 4.11: Ames toxicity predictions. Compounds highlighted with "Yes" are toxic according to in silico Ames test predictions from both tools.

Compound	S. Plus	pkCSM
GW7647	No	No
OMDM2	No	No
STK-22	No	No
Compound 1	No	Yes
AM1172	No	No
SBFI-103	No	No
OMDM1	No	No
STK-15	No	No
Antrafenine	No	No
a-FABP-IN-1	No	No
Compound 34	Yes	Yes
AM404	No	No
BMS309403	No	No
ZTY28	No	Yes
ART26.12	No	No

Animal tests related to rats of ADMET Predictor are summarized in **Table 4.15** alongside other TOX_Risk violations of the compounds. Fathead Minnow toxicity is evaluated by both predictors, as shown in **Table 4.12**, and is a metric useful to evaluate the acute toxicity of the compounds. Values for the toxicity are expressed as the logarithm of the local concentration 50 (LC50), which is the concentration of the compound that would kill 50% of the animals in a test. Values under -0.3 are an indicator of acute toxicity, which is agreed upon by both predictors.

Table 4.12: Fathead Minnow predictions expressed as log(LC50).

Compound	S. Plus	pkCSM
GW7647	-2.45	-1.02
OMDM2	-0.55	-1.07
STK-22	-3.17	-3.19
Compound 1	-1.77	-1.66
AM1172	0.35	-1.94
SBFI-103	-2.08	-6.16
OMDM1	-0.55	-1.72
STK-15	-3.06	-3.47
Antrafenine	-2.38	-1.72
a-FABP-IN-1	-3.10	-3.60
Compound 34	-3.25	-4.61
AM404	0.40	-1.39
BMS309403	-2.65	-2.10
ZTY28	-2.59	-4.84
ART26.12	-0.89	-2.45

As for human tests, hepatotoxicity of ADMET Predictor is included in **Table 4.15** once again. Cardiotoxicity is expressed as a qualitative indication of hERG channel inhibition, as shown in **Table 4.13**; both predictors fail to reach a definite consensus on this specific subject, only agreeing in 4 cases. The Maximum Recommended Therapeutic Dose and Maximum Recommended Therapeutic Dose predictions are included in **Table 4.14**. The former is calculated as a threshold value of 0.5 log(mg/kg/day) with a qualitative above/below description, while the latter uses the same unit of measure but is a quantitative prediction that considers a value above 0.477 log(mg/kg/day) as high, vice versa low.

Once again, ADMET Predictor summarizes its predictions regarding the toxicity of the chemical compounds with risk scores and codes, as shown in **Table 4.15**. TOX_risk is higher than 2 for 9% of the compounds used in the training set for construction of the ADMET Predictor. Half of the compounds are below the rule violation limit; only a few are considered cardiotoxic or hepatotoxic, with most of the rule violations being caused by chronic and acute rat toxicity.

Table 4.13: hERG toxicity predictions. Compounds highlighted with "Yes" are considered cardiotoxic according to the predictions of both tools.

Compound	S. Plus	pkCSM
GW7647	No	No
OMDM2	No	Yes
STK-22	No	Yes
Compound 1	Yes	No
AM1172	No	Yes
SBFI-103	Yes	No
OMDM1	No	Yes
STK-15	No	Yes
Antrafenine	Yes	Yes
a-FABP-IN-1	Yes	No
Compound 34	No	No
AM404	No	Yes
BMS309403	Yes	No
ZTY28	Yes	No
ART26.12	No	No

Table 4.14: Maximum Recommended Therapeutic Dose (S. Plus) and Maximum Recommended Tolerated Dose (pkCSM) predictions. For ADMET Predictor, below/above refer to a threshold value of around 0.5 log(mg/kg/day).

Compound	S. Plus	pkCSM
GW7647	Below	-0.22
OMDM2	Above	-1.06
STK-22	Above	-0.08
Compound 1	Below	0.37
AM1172	Above	-0.43
SBFI-103	Above	0.37
OMDM1	Above	-0.33
STK-15	Above	0.04
Antrafenine	Below	0.20
a-FABP-IN-1	Above	0.50
Compound 34	Below	0.25
AM404	Above	-0.70
BMS309403	Above	0.46
ZTY28	Below	0.35
ART26.12	Below	0.62

Table 4.15: Toxicity risk prediction of ADMET Predictor for the 15 compounds.

Compound	TOX_Risk	TOX_Code
GW7647	0	-
OMDM2	0	-
STK-22	2	rat; Xr-; Xm-
Compound 1	2	rat; Xr-; Xm+
AM1172	0.5	hERG-
SBFI-103	2	rat; Xr-; Xm-
OMDM1	0	-
STK-15	2	rat; Xr-; Xm-
Antrafenine	1.5	rat; Xr+; Xm-
a-FABP-IN-1	2.9	hERG; rat; Xr-; Xm+; HEPX+
Compound 34	3	rat; Xr-; Xm+; MUT
AM404	0.5	hERG-
BMS309403	1.9	hERG; rat; Xr-; Xm+; HEPX+
ZTY28	3.5	hERG; rat; Xr-; Xm-; HEPX+
ART26.12	1.5	rat; Xr-

Chapter 5

Conclusion

Discovery of the involvement of FABP12 in the aberrant lipid metabolism of prostate cancer has changed the landscape in the search for innovative therapeutics for advanced versions of this deadly disease. Taking advantage of prostate cancer's peculiar metabolism before it develops into more aggressive forms means targeting the overexpression of FABP12, which can induce signaling responses in the cells by the accumulation of lipid droplets, which can eventually promote EMT.

The University of Alberta researchers, who have discovered FABP12 and its involvement in prostate cancer, will continue to study the protein and are set to find out more about the energy-production, lipid signaling and lipidomics, metabolomics, and epigenetics of the link that connects prostate cancer and FABP12, which eventually leads to the development of an invariably fatal form of the pathology. Part of the investigation finds its purpose and aim in the search for novel drugs that can clinically treat aggressive prostate cancer by inhibition of FABP12.

Discovery of novel drugs is not simple: their development can take 10-15 years and anywhere between 1 to 3 billion dollars. Every step of development is complex and riddled with various problems and issues that can derail progress unexpectedly, especially with the transition from in silico simulations to the preclinical phase. Rational drug design is not a technique or a method that is perfect by any means and requires numerous iterations of development cycles before it can start to give out substantial results.

My own research is proof of this. The docking simulations and ADMET predictions I performed are only possible due to the work of those who came before me, the University of Alberta, due to their research on FABP12, and Enrico Astara, as already mentioned, for providing the AF models necessary to make docking comparisons. My own work, in fact, will be useful to other students and researchers in the same way if it is found to have success in its purpose: to evaluate the quality of the FABP12 AF models and find potential inhibitors of the protein that can be readily accessible for validation of the docking simulations themselves. In vitro

validation of the results is already underway at the University of Alberta and will enlighten whoever is responsible for the next cycles of development as to how to proceed better.

Presuming that any of the chemical compounds highlighted in the results or belonging to the database of 816 known ligands for FABP inhibition would go straight to the preclinical phase would be arrogant, at best. This is quite evident: those same compounds have already been tested, discarded, or chosen for subsequent studies by teams of researchers who certainly have more experience than just one student, and yet only ART26.12 has reached the preclinical stage. If there is an assumption to be made, it is that it's unlikely that any other compound would work exceptionally well for FABP12. Surprises could happen, but they are not to be expected in the least.

However, this doesn't mean that this research is pointless, as this is only the first cycle of development. The docking results obtained are promising for AF models, and analysis of the binding sites and scoring of the poses generated has highlighted 112 compounds that perform better than the control BMS309403. Out of these 112, 12 are commercial and will be used for the in vitro validation. As already mentioned, it is unlikely that any of these will ever reach the preclinical stage; however, the chemical formula of the compounds themselves could be useful if it is discovered that any of them are actually very potent, so that they could be used as a base for ligand derivatization.

Ligand derivatization is a technique that mainly consists of taking parts, functional groups, or even hydrogens or carbon groups within molecules and replacing them with other functional groups with physicochemical characteristics that can improve the performance of the original ligand. This can be used both to adjust the ADMET properties of the compound and lower the risk score, but also to better exploit the steric and donor/acceptor character of amino acids in the binding site of the protein.

This method, however, is not the only one that could be used to push the research forward. The successful compounds could be used as the basis of QSAR, fingerprint, and pharmacophore models, which will identify similar molecules (depending on physicochemical properties, similarity, functional groups, etc.) among public databases of compounds. One such example is ZINC, which contains over a billion chemicals, most of which are purchasable, making subsequent in vitro testing easier to carry out.

Improvements could be made on the proteins themselves: Molecular Dynamics techniques were completely disregarded due to the computational time required to perform significant simulations on 19 different structures. By placing the original protein in an aqueous environment similar to the human body and then performing a cycle of heating and energy minimization, the result should be a conformation of the protein that would more accurately represent reality. Not only that, but

molecular dynamics lets the user implement an explicit water solvation model to improve performance further. The presence of water in the protein structure will more closely model the actual binding mode of FABP12 to its inhibitors, given the common presence of water molecules that are interposed and bound between ligand and receptor.

Molecular Dynamics can also be applied after the docking procedures in the same way as below, for a second cycle of heating and minimization. This should further stabilize the protein and improve performance. Down the line, a specific application of Molecular Dynamics called Thermodynamic Integration could be used to evaluate the difference in free binding energy between two macroscopic states of the protein (two different binding poses with differing ligands) to support the results further. MOE implements a version of Thermodynamic Integration that is capable of sampling the search space created by the presence of multiple ligands without having to perform the simulation on every single combination, which would otherwise quickly get out of control in terms of required time.

The docking simulations themselves could also implement consensus docking, which simply translates to the use of multiple docking software which continue to generate poses until they reach a consensus in the S Scores calculated (or they are timed out by an algorithm).

Concluding, while there is still a lot to be done, the results obtained (especially those related to AlphaFold) are promising and will certainly provide a good starting point for future studies. The implementation of the prediction of ADMET properties in the workflow is borderline mandatory, and while results sometimes differ between tools, they can be very useful to pinpoint potential inhibitors. It would be interesting to perform a more systematic review of all available tools and evaluate their performance against known FDA drugs and future tested compounds for FABP12.

Bibliography

[1] T. I. A. f. R. o. Cancer (IARC), *Global Cancer Observatory, https://gco.iarc.fr/*\
en-US.

[2] F. Bray, M. Laversanne, H. Sung, *et al.*, «Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries», en, *CA: A Cancer Journal for Clinicians*, vol. 74, no. 3, pp. 229–263, May 2024, ISSN: 0007-9235, 1542-4863. DOI: `10.3322/caac.21834`.

[3] *Cancer Today, https://gco.iarc.who.int/today/*.

[4] G. Gandaglia, P. Albers, P.-A. Abrahamsson, *et al.*, «Structured Population-based Prostate-specific Antigen Screening for Prostate Cancer: The European Association of Urology Position in 2019», en, *European Urology*, vol. 76, no. 2, pp. 142–150, Aug. 2019, ISSN: 03022838. DOI: `10.1016/j.eururo.2019.04.033`.

[5] *Cancer Tomorrow, https://gco.iarc.who.int/tomorrow/*.

[6] P. Rawla, «Epidemiology of Prostate Cancer», en, *World Journal of Oncology*, vol. 10, no. 2, pp. 63–89, 2019, ISSN: 1920-4531, 1920-454X. DOI: `10.14740/wjon1191`.

[7] *SEER Explorer Application, https://seer.cancer.gov/statistics-network*.

[8] *PROSTATE CANCER, https://www.efpia.eu/publications/cancer-comparator-report/cancer-types/prostate-cancer/*.

[9] *GCO - SURVCAN, https://gco.iarc.who.int/survival/survcan/*.

[10] R. L. Siegel, K. D. Miller, H. E. Fuchs, *et al.*, «Cancer statistics, 2022», en, *CA: A Cancer Journal for Clinicians*, vol. 72, no. 1, pp. 7–33, Jan. 2022, ISSN: 0007-9235, 1542-4863. DOI: `10.3322/caac.21708`.

[11] M. B. Clements, E. A. Vertosick, L. Guerrios-Rivera, *et al.*, «Defining the Impact of Family History on Detection of High-grade Prostate Cancer in a Large Multi-institutional Cohort», en, *European Urology*, vol. 82, no. 2, pp. 163–169, Aug. 2022, ISSN: 03022838. DOI: `10.1016/j.eururo.2021.12.011`.

[12] M. N. Okobia, J. M. Zmuda, R. E. Ferrell, *et al.*, «Chromosome 8q24 variants are associated with prostate cancer risk in a high risk population of African ancestry», en, *The Prostate*, vol. 71, no. 10, pp. 1054–1063, Jul. 2011, ISSN: 0270-4137, 1097-0045. DOI: 10.1002/pros.21320.

[13] L. A. Mucci, J. B. Hjelmborg, J. R. Harris, *et al.*, «Familial Risk and Heritability of Cancer Among Twins in Nordic Countries», en, *JAMA*, vol. 315, no. 1, p. 68, Jan. 2016, ISSN: 0098-7484. DOI: 10.1001/jama.2015.17703.

[14] A. Plym, Y. Zhang, K. H. Stopsack, *et al.*, «A Healthy Lifestyle in Men at Increased Genetic Risk for Prostate Cancer», en, *European Urology*, vol. 83, no. 4, pp. 343–351, Apr. 2023, ISSN: 03022838. DOI: 10.1016/j.eururo.2022.05.008.

[15] L. Aaron, O. E. Franco, and S. W. Hayward, «Review of Prostate Anatomy and Embryology and the Etiology of Benign Prostatic Hyperplasia», eng, *The Urologic Clinics of North America*, vol. 43, no. 3, pp. 279–288, Aug. 2016, ISSN: 1558-318X. DOI: 10.1016/j.ucl.2016.04.012.

[16] T. B. J. Murray, «The Pathogenesis of Prostate Cancer», in *Prostate Cancer*, Urology Department, Frimley Park Hospital, Portsmouth Rd, Frimley, Camberley GU16 7UJ, UK, S. R. Bott, and K. Lim Ng, Eds., Exon Publications, May 2021, pp. 29–42, ISBN: 978-0-6450017-5-4. DOI: 10.36255/exonpublications.prostatecancer.pathogenesis.2021.

[17] D. G. Bostwick and J. Qian, «High-grade prostatic intraepithelial neoplasia», en, *Modern Pathology*, vol. 17, no. 3, pp. 360–379, Mar. 2004, ISSN: 08933952. DOI: 10.1038/modpathol.3800053.

[18] P. A. Humphrey, «Histopathology of Prostate Cancer», en, *Cold Spring Harbor Perspectives in Medicine*, vol. 7, no. 10, a030411, Oct. 2017, ISSN: 2157-1422. DOI: 10.1101/cshperspect.a030411.

[19] W. G. Nelson, A. M. De Marzo, and W. B. Isaacs, «Prostate Cancer», en, *New England Journal of Medicine*, vol. 349, no. 4, pp. 366–381, Jul. 2003, ISSN: 0028-4793, 1533-4406. DOI: 10.1056/NEJMra021562.

[20] J. I. Epstein, L. Egevad, M. B. Amin, *et al.*, «The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System», en, *American Journal of Surgical Pathology*, vol. 40, no. 2, pp. 244–252, Feb. 2016, ISSN: 0147-5185. DOI: 10.1097/PAS.0000000000000530.

[21] M. A. Barakzai, «Prostatic Adenocarcinoma: A Grading from Gleason to the New Grade-Group System: A Historical and Critical Review», en, *Asian Pacific Journal of Cancer Prevention*, vol. 20, no. 3, pp. 661–666, Mar. 2019, ISSN: 2476-762X. DOI: 10.31557/APJCP.2019.20.3.661.

[22] O. Six, *The basics of PI-RADS scoring and how AI can help*, en-us.

[23] M. E. Tan, J. Li, H. E. Xu, *et al.*, «Androgen receptor: Structure, role in prostate cancer and drug discovery», en, *Acta Pharmacologica Sinica*, vol. 36, no. 1, pp. 3–23, Jan. 2015, ISSN: 1671-4083, 1745-7254. DOI: 10.1038/aps.2014.18.

[24] S. M. Green, E. A. Mostaghel, and P. S. Nelson, «Androgen action and metabolism in prostate cancer», en, *Molecular and Cellular Endocrinology*, vol. 360, no. 1-2, pp. 3–13, Sep. 2012, ISSN: 03037207. DOI: 10.1016/j.mce.2011.09.046.

[25] O. Sartor and J. S. De Bono, «Metastatic Prostate Cancer», en, *New England Journal of Medicine*, vol. 378, no. 7, D. L. Longo, Ed., pp. 645–657, Feb. 2018, ISSN: 0028-4793, 1533-4406. DOI: 10.1056/NEJMra1701695.

[26] K. B. Blagoev, R. Iordanov, M. Zhou, *et al.*, «Drug resistant cells with very large proliferative potential grow exponentially in metastatic prostate cancer», en, *Oncotarget*, vol. 12, no. 1, pp. 15–21, Jan. 2021, ISSN: 1949-2553. DOI: 10.18632/oncotarget.27855.

[27] M. Montanari, S. Rossetti, C. Cavaliere, *et al.*, «Epithelial-mesenchymal transition in prostate cancer: An overview», en, *Oncotarget*, vol. 8, no. 21, pp. 35 376–35 389, May 2017, ISSN: 1949-2553. DOI: 10.18632/oncotarget.15686.

[28] J. Yang, P. Antin, G. Berx, *et al.*, «Guidelines and definitions for research on epithelial–mesenchymal transition», en, *Nature Reviews Molecular Cell Biology*, vol. 21, no. 6, pp. 341–352, Jun. 2020, ISSN: 1471-0072, 1471-0080. DOI: 10.1038/s41580-020-0237-9.

[29] H. Bonkhoff, T. Fixemer, I. Hunsicker, *et al.*, «Estrogen Receptor Expression in Prostate Cancer and Premalignant Prostatic Lesions», en, *The American Journal of Pathology*, vol. 155, no. 2, pp. 641–647, Aug. 1999, ISSN: 00029440. DOI: 10.1016/S0002-9440(10)65160-7.

[30] P. Mak, I. Leav, B. Pursell, *et al.*, «ER Impedes Prostate Cancer EMT by Destabilizing HIF-1 and Inhibiting VEGF-Mediated Snail Nuclear Localization: Implications for Gleason Grading», en, *Cancer Cell*, vol. 17, no. 4, pp. 319–332, Apr. 2010, ISSN: 15356108. DOI: 10.1016/j.ccr.2010.02.030.

[31] R.-Z. Liu and R. Godbout, «An Amplified Fatty Acid-Binding Protein Gene Cluster in Prostate Cancer: Emerging Roles in Lipid Metabolism and Metastasis», en, *Cancers*, vol. 12, no. 12, p. 3823, Dec. 2020, ISSN: 2072-6694. DOI: `10.3390/cancers12123823`.

[32] R.-Z. Liu, W.-S. Choi, S. Jain, *et al.*, «The FABP12/PPAR pathway promotes metastatic transformation by inducing epithelial-to-mesenchymal transition and lipid-derived energy production in prostate cancer cells», en, *Molecular Oncology*, vol. 14, no. 12, pp. 3100–3120, Dec. 2020, ISSN: 1574-7891, 1878-0261. DOI: `10.1002/1878-0261.12818`.

[33] R.-Z. Liu, M. Garg, X.-H. Yang, *et al.*, «Docetaxel-Induced Cell Death Is Regulated by a Fatty Acid-Binding Protein 12-Slug-Survivin Pathway in Prostate Cancer Cells», en, *International Journal of Molecular Sciences*, vol. 25, no. 17, p. 9669, Sep. 2024, ISSN: 1422-0067. DOI: `10.3390/ijms251 79669`.

[34] H. Van Poppel, T. Albreht, P. Basu, *et al.*, «Serum PSA-based early detection of prostate cancer in Europe and globally: Past, present and future», en, *Nature Reviews Urology*, vol. 19, no. 9, pp. 562–572, Sep. 2022, ISSN: 1759-4812, 1759-4820. DOI: `10.1038/s41585-022-00638-6`.

[35] J.-L. Descotes, «Diagnosis of prostate cancer», en, *Asian Journal of Urology*, vol. 6, no. 2, pp. 129–136, Apr. 2019, ISSN: 22143882. DOI: `10.1016/j.ajur. 2018.11.007`.

[36] G. S. Sandhu and G. L. Andriole, «Overdiagnosis of prostate cancer», eng, *Journal of the National Cancer Institute. Monographs*, vol. 2012, no. 45, pp. 146–151, Dec. 2012, ISSN: 1745-6614. DOI: `10.1093/jncimonographs/ lgs031`.

[37] M. S. Litwin and H.-J. Tan, «The Diagnosis and Treatment of Prostate Cancer: A Review», en, *JAMA*, vol. 317, no. 24, p. 2532, Jun. 2017, ISSN: 0098-7484. DOI: `10.1001/jama.2017.7248`.

[38] A. Stabile, F. Giganti, A. B. Rosenkrantz, *et al.*, «Multiparametric MRI for prostate cancer diagnosis: Current status and future directions», en, *Nature Reviews Urology*, vol. 17, no. 1, pp. 41–61, Jan. 2020, ISSN: 1759-4812, 1759-4820. DOI: `10.1038/s41585-019-0212-4`.

[39] J. O. Barentsz, J. Richenberg, R. Clements, *et al.*, «ESUR prostate MR guidelines 2012», en, *European Radiology*, vol. 22, no. 4, pp. 746–757, Apr. 2012, ISSN: 0938-7994, 1432-1084. DOI: `10.1007/s00330-011-2377-y`.

[40] *Initial Treatment of Prostate Cancer, https://www.cancer.org/cancer/*, en.

[41] M. Amiri, S. Yousefnia, F. Seyed Forootan, *et al.*, «Diverse roles of fatty acid binding proteins (FABPs) in development and pathogenesis of cancers», en, *Gene*, vol. 676, pp. 171–183, Nov. 2018, ISSN: 03781119. DOI: 10.1016/j.gene.2018.07.035.

[42] R. L. Smathers and D. R. Petersen, «The human fatty acid-binding protein family: Evolutionary divergences and functions», en, *Human Genomics*, vol. 5, no. 3, p. 170, 2011, ISSN: 1479-7364. DOI: 10.1186/1479-7364-5-3-170.

[43] L. C. Lawrie, S. R. Dundas, S. Curran, *et al.*, «Liver fatty acid binding protein expression in colorectal neoplasia», en, *British Journal of Cancer*, vol. 90, no. 10, pp. 1955–1960, May 2004, ISSN: 0007-0920, 1532-1827. DOI: 10.1038/sj.bjc.6601828.

[44] I. H. McKillop, C. A. Girardi, and K. J. Thompson, «Role of fatty acid binding proteins (FABPs) in cancer development and progression», en, *Cellular Signalling*, vol. 62, p. 109 336, Oct. 2019, ISSN: 08986568. DOI: 10.1016/j.cellsig.2019.06.001.

[45] K. C. B. Yabut, A. Martynova, A. Nath, *et al.*, «Drugs Form Ternary Complexes with Human Liver Fatty Acid Binding Protein 1 (FABP1) and FABP1 Binding Alters Drug Metabolism», en, *Molecular Pharmacology*, vol. 105, no. 6, pp. 395–410, Jun. 2024, ISSN: 0026895X. DOI: 10.1124/molpharm.124.000878.

[46] H. Huang, A. L. McIntosh, G. G. Martin, *et al.*, «FABP1: A Novel Hepatic Endocannabinoid and Cannabinoid Binding Protein», en, *Biochemistry*, vol. 55, no. 37, pp. 5243–5255, Sep. 2016, ISSN: 0006-2960, 1520-4995. DOI: 10.1021/acs.biochem.6b00446.

[47] F. Shahbazi, S. Mohammadzadeh, D. Meister, *et al.*, *Liver-Type Fatty Acid Binding Protein (FABP1) Has Exceptional Affinity for Minor Cannabinoids*, Sep. 2023. DOI: 10.26434/chemrxiv-2023-82ztv.

[48] Q. Ren, Y. Chen, Z. Zhou, *et al.*, «Discovery of the First-in-Class Intestinal Restricted FXR and FABP1 Dual Modulator ZLY28 for the Treatment of Nonalcoholic Fatty Liver Disease», en, *Journal of Medicinal Chemistry*, vol. 66, no. 9, pp. 6082–6104, May 2023, ISSN: 0022-2623, 1520-4804. DOI: 10.1021/acs.jmedchem.2c01918.

[49] J. Storch and A. E. Thumser, «The fatty acid transport function of fatty acid-binding proteins», en, *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, vol. 1486, no. 1, pp. 28–44, Jun. 2000, ISSN: 13881981. DOI: 10.1016/S1388-1981(00)00046-9.

[50] A. Cheng, Y. Shinoda, T. Yamamoto, *et al.*, «Development of FABP3 ligands that inhibit arachidonic acid-induced -synuclein oligomerization», en, *Brain Research*, vol. 1707, pp. 190–197, Mar. 2019, ISSN: 00068993. DOI: 10.1016/j.brainres.2018.11.036.

[51] H. Cai, G. Yan, X. Zhang, *et al.*, «Discovery of highly selective inhibitors of human fatty acid binding protein 4 (FABP4) by virtual screening», en, *Bioorganic & Medicinal Chemistry Letters*, vol. 20, no. 12, pp. 3675–3679, Jun. 2010, ISSN: 0960894X. DOI: 10.1016/j.bmcl.2010.04.095.

[52] G. Chen, H. Xie, M. You, *et al.*, «Structure-based design of potent FABP4 inhibitors with high selectivity against FABP3», en, *European Journal of Medicinal Chemistry*, vol. 264, p. 115 984, Jan. 2024, ISSN: 02235234. DOI: 10.1016/j.ejmech.2023.115984.

[53] R. Sulsky, D. R. Magnin, Y. Huang, *et al.*, «Potent and selective biphenyl azole inhibitors of adipocyte fatty acid binding protein (aFABP)», en, *Bioorganic & Medicinal Chemistry Letters*, vol. 17, no. 12, pp. 3511–3515, Jun. 2007, ISSN: 0960894X. DOI: 10.1016/j.bmcl.2006.12.044.

[54] D.-D. Gao, H.-X. Dou, H.-X. Su, *et al.*, «From hit to lead: Structure-based discovery of naphthalene-1-sulfonamide derivatives as potent and selective inhibitors of fatty acid binding protein 4», en, *European Journal of Medicinal Chemistry*, vol. 154, pp. 44–59, Jun. 2018, ISSN: 02235234. DOI: 10.1016/j.ejmech.2018.05.007.

[55] W. Al-Jameel, X. Gou, S. S. Forootan, *et al.*, «Inhibitor SBFI26 suppresses the malignant progression of castration-resistant PC3-M cells by competitively binding to oncogenic FABP5», en, *Oncotarget*, vol. 8, no. 19, pp. 31 041–31 056, May 2017, ISSN: 1949-2553. DOI: 10.18632/oncotarget.16055.

[56] M. Kaczocha, M. J. Rebecchi, B. P. Ralph, *et al.*, «Inhibition of Fatty Acid Binding Proteins Elevates Brain Anandamide Levels and Produces Analgesia», en, *PLoS ONE*, vol. 9, no. 4, T. J. Price, Ed., e94200, Apr. 2014, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0094200.

[57] T. Ohmachi, H. Inoue, K. Mimori, *et al.*, «Fatty Acid Binding Protein 6 Is Overexpressed in Colorectal Cancer», en, *Clinical Cancer Research*, vol. 12, no. 17, pp. 5090–5095, Sep. 2006, ISSN: 1078-0432, 1557-3265. DOI: 10.1158/1078-0432.CCR-05-2045.

[58] G. Kaloshi, K. Mokhtari, C. Carpentier, *et al.*, «FABP7 expression in glioblastomas: Relation to prognosis, invasion and EGFR status», en, *Journal of Neuro-Oncology*, vol. 84, no. 3, pp. 245–248, Sep. 2007, ISSN: 0167-594X, 1573-7373. DOI: 10.1007/s11060-007-9377-4.

[59] W. Knoll, F. Natali, J. Peters, *et al.*, «Dynamic properties of a reconstituted myelin sheath», en, *Spectroscopy*, vol. 24, no. 6, pp. 585–592, 2010, ISSN: 0712-4813, 1875-922X. DOI: 10.1155/2010/317525.

[60] M. S. Al Fayi, X. Gou, S. S. Forootan, *et al.*, «The increased expression of fatty acid-binding protein 9 in prostate cancer and its prognostic significance», en, *Oncotarget*, vol. 7, no. 50, pp. 82 783–82 797, Dec. 2016, ISSN: 1949-2553. DOI: 10.18632/oncotarget.12635.

[61] R.-Z. Liu, X. Li, and R. Godbout, «A novel fatty acid-binding protein (FABP) gene resulting from tandem gene duplication in mammals: Transcription in rat retina and testis», en, *Genomics*, vol. 92, no. 6, pp. 436–445, Dec. 2008, ISSN: 08887543. DOI: 10.1016/j.ygeno.2008.08.003.

[62] *PDB101: Learn: Guide to Understanding PDB Data: Methods for Determining Structure, https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/methods-for-determining-structure.*

[63] H. M. Berman, J. Westbrook, Z. Feng, *et al.*, «The Protein Data Bank», eng, *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, Jan. 2000, ISSN: 0305-1048. DOI: 10.1093/nar/28.1.235.

[64] *Atomic Coordinate Entry Format Version 3.3, https://www.wwpdb.org/documentation.*

[65] The UniProt Consortium, A. Bateman, M.-J. Martin, *et al.*, «UniProt: The Universal Protein Knowledgebase in 2025», en, *Nucleic Acids Research*, vol. 53, no. D1, pp. D609–D617, Jan. 2025, ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkae1010.

[66] *Chemical Computing Group (CCG) | Computer-Aided Molecular Design, https://www.chemcomp.com/en/index.htm.*

[67] A. Waterhouse, M. Bertoni, S. Bienert, *et al.*, «SWISS-MODEL: Homology modelling of protein structures and complexes», en, *Nucleic Acids Research*, vol. 46, no. W1, W296–W303, Jul. 2018, ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gky427.

[68] M. T. Muhammed and E. Aki-Yalcin, «Homology modeling in drug discovery: Overview, current applications, and future perspectives», en, *Chemical Biology & Drug Design*, vol. 93, no. 1, pp. 12–20, Jan. 2019, ISSN: 1747-0277, 1747-0285. DOI: 10.1111/cbdd.13388.

[69] *SAVESv6.1 - Structure Validation Server, https://saves.mbi.ucla.edu/.*

[70] J. Jumper, R. Evans, A. Pritzel, *et al.*, «Highly accurate protein structure prediction with AlphaFold», en, *Nature*, vol. 596, no. 7873, pp. 583–589, Aug. 2021, ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-021-03819-2.

[71] J. Pereira, A. J. Simpkin, M. D. Hartmann, *et al.*, «High-accuracy protein structure prediction in casp14», en, *Proteins: Structure, Function, and Bioinformatics*, vol. 89, no. 12, pp. 1687–1699, Dec. 2021, ISSN: 0887-3585, 1097-0134. DOI: `10.1002/prot.26171`.

[72] A. O. Stevens and Y. He, «Benchmarking the Accuracy of AlphaFold 2 in Loop Structure Prediction», en, *Biomolecules*, vol. 12, no. 7, p. 985, Jul. 2022, ISSN: 2218-273X. DOI: `10.3390/biom12070985`.

[73] A. Gimeno, M. J. Ojeda-Montes, S. Tomás-Hernández, *et al.*, «The Light and Dark Sides of Virtual Screening: What Is There to Know?», en, *International Journal of Molecular Sciences*, vol. 20, no. 6, p. 1375, Mar. 2019, ISSN: 1422-0067. DOI: `10.3390/ijms20061375`.

[74] A. Kukol and J. M. Walker, Eds., *Molecular Modeling of Proteins*, ser. Methods in Molecular Biology. Totowa, NJ: Humana Press, 2008, vol. 443, ISBN: 978-1-58829-864-5 978-1-59745-177-2. DOI: `10.1007/978-1-59745-177-2`.

[75] K. Vanommeslaeghe, O. Guvench, and A. D. MacKerell, «Molecular Mechanics», en, *Current Pharmaceutical Design*, vol. 20, no. 20, pp. 3281–3292, May 2014, ISSN: 13816128. DOI: `10.2174/13816128113199990600`.

[76] J. Nochebuena, S. Naseem-Khan, and G. A. Cisneros, «Development and application of quantum mechanics/molecular mechanics methods with advanced polarizable potentials», en, *WIREs Computational Molecular Science*, vol. 11, no. 4, e1515, Jul. 2021, ISSN: 1759-0876, 1759-0884. DOI: `10.1002/wcms.1515`.

[77] S. Singh, Q. Bani Baker, and D. B. Singh, «Molecular docking and molecular dynamics simulation», en, in *Bioinformatics*, Elsevier, 2022, pp. 291–304, ISBN: 978-0-323-89775-4. DOI: `10.1016/B978-0-323-89775-4.00014-6`.

[78] W. F. De Azevedo, Ed., *Docking Screens for Drug Discovery*, en, ser. Methods in Molecular Biology. New York, NY: Springer New York, 2019, vol. 2053, ISBN: 978-1-4939-9751-0 978-1-4939-9752-7. DOI: `10.1007/978-1-4939-9752-7`.

[79] L. Ferreira, R. Dos Santos, G. Oliva, *et al.*, «Molecular Docking and Structure-Based Drug Design Strategies», en, *Molecules*, vol. 20, no. 7, pp. 13 384–13 421, Jul. 2015, ISSN: 1420-3049. DOI: `10.3390/molecules200713384`.

[80] *Force Fields in Molecular Dynamics, https://www.compchems.com/*, en, Section: posts, Jul. 2022.

[81] P. Gkeka, «Molecular dynamics studies of peptide-membrane interactions: Insights from coarse-grained models», en, 2010, Accepted: 2011-01-19T10:28:51Z Publisher: The University of Edinburgh.

[82] B. K. Shoichet, A. R. Leach, and I. D. Kuntz, «Ligand solvation in molecular docking», en, *Proteins: Structure, Function, and Genetics*, vol. 34, no. 1, pp. 4–16, Jan. 1999, ISSN: 0887-3585, 1097-0134. DOI: `10.1002/(SICI)1097-0134(19990101)34:1<4::AID-PROT2>3.0.CO;2-6`.

[83] M. Mursal, M. Ahmad, S. Hussain, *et al.*, «Navigating the Computational Seas: A Comprehensive Overview of Molecular Docking Software in Drug Discovery», en, in *Biomedical Engineering*, Č. Podlipnik, Ed., vol. 27, IntechOpen, Apr. 2024, ISBN: 978-0-85466-842-7 978-0-85466-841-0. DOI: `10.5772/intechopen.1004802`.

[84] G. Zhou, D.-V. Rusnac, H. Park, *et al.*, «An artificial intelligence accelerated virtual screening platform for drug discovery», en, *Nature Communications*, vol. 15, no. 1, p. 7761, Sep. 2024, ISSN: 2041-1723. DOI: `10.1038/s41467-024-52061-7`.

[85] R. Han, H. Yoon, G. Kim, *et al.*, «Revolutionizing Medicinal Chemistry: The Application of Artificial Intelligence (AI) in Early Drug Discovery», en, *Pharmaceuticals*, vol. 16, no. 9, p. 1259, Sep. 2023, ISSN: 1424-8247. DOI: `10.3390/ph16091259`.

[86] I. D. Kuntz, J. M. Blaney, S. J. Oatley, *et al.*, «A geometric approach to macromolecule-ligand interactions», en, *Journal of Molecular Biology*, vol. 161, no. 2, pp. 269–288, Oct. 1982, ISSN: 00222836. DOI: `10.1016/0022-2836(82)90153-X`.

[87] N. S. Pagadala, K. Syed, and J. Tuszynski, «Software for molecular docking: A review», en, *Biophysical Reviews*, vol. 9, no. 2, pp. 91–102, Apr. 2017, ISSN: 1867-2450, 1867-2469. DOI: `10.1007/s12551-016-0247-1`.

[88] A. B. Mamonov, M. Moghadasi, H. Mirzaei, *et al.*, «Focused grid-based resampling for protein docking and mapping», en, *Journal of Computational Chemistry*, vol. 37, no. 11, pp. 961–970, Apr. 2016, ISSN: 0192-8651, 1096-987X. DOI: `10.1002/jcc.24273`.

[89] J. M. Paggi, A. Pandit, and R. O. Dror, «The Art and Science of Molecular Docking», en, *Annual Review of Biochemistry*, vol. 93, no. 1, pp. 389–410, Aug. 2024, ISSN: 0066-4154, 1545-4509. DOI: `10.1146/annurev-biochem-030222-120000`.

[90] M. Furuhashi, G. Tuncman, C. Z. Görgün, *et al.*, «Treatment of diabetes and atherosclerosis by inhibiting fatty-acid-binding protein aP2», en, *Nature*, vol. 447, no. 7147, pp. 959–965, Jun. 2007, ISSN: 0028-0836, 1476-4687. DOI: `10.1038/nature05844`.

[91] K. Rajan, H. O. Brinkhaus, M. I. Agea, *et al.*, «DECIMER.ai: An open platform for automated optical chemical structure identification, segmentation and recognition in scientific publications», en, *Nature Communications*, vol. 14, no. 1, p. 5045, Aug. 2023, ISSN: 2041-1723. DOI: 10.1038/s41467-023-40782-0.

[92] *Ketcher, https://lifescience.opensource.epam.com/ketcher/.*

[93] *SMILES to IUPAC - Leskoff, https://www.leskoff.com/s01814-0.*

[94] X.-Y. Chen, Y. Yin, J. Xi, *et al.*, «11-Aza-artemisinin Derivatives Exhibit Anticancer Activities by Targeting the Fatty Acid Binding Protein 6 (FABP6)», en, *Chinese Journal of Chemistry*, vol. 36, no. 12, pp. 1197–1201, Dec. 2018, ISSN: 1001-604X, 1614-7065. DOI: 10.1002/cjoc.201800361.

[95] G. Floresta, V. Pistarà, E. Amata, *et al.*, «Adipocyte fatty acid binding protein 4 (FABP4) inhibitors. A comprehensive systematic review», en, *European Journal of Medicinal Chemistry*, vol. 138, pp. 854–873, Sep. 2017, ISSN: 02235234. DOI: 10.1016/j.ejmech.2017.07.022.

[96] G. Floresta, A. Cilibrizzi, V. Abbate, *et al.*, «3D-QSAR assisted identification of FABP4 inhibitors: An effective scaffold hopping analysis/QSAR evaluation», en, *Bioorganic Chemistry*, vol. 84, pp. 276–284, Mar. 2019, ISSN: 00452068. DOI: 10.1016/j.bioorg.2018.11.045.

[97] G. Floresta, V. Patamia, C. Zagni, *et al.*, «Adipocyte fatty acid binding protein 4 (FABP4) inhibitors. An update from 2017 to early 2022», en, *European Journal of Medicinal Chemistry*, vol. 240, p. 114 604, Oct. 2022, ISSN: 02235234. DOI: 10.1016/j.ejmech.2022.114604.

[98] Y. Shinoda, Y. Wang, T. Yamamoto, *et al.*, «Analysis of binding affinity and docking of novel fatty acid-binding protein (FABP) ligands», en, *Journal of Pharmacological Sciences*, vol. 143, no. 4, pp. 264–271, Aug. 2020, ISSN: 13478613. DOI: 10.1016/j.jphs.2020.05.005.

[99] H.-y. Cai, T. Wang, J.-c. Zhao, *et al.*, «Benzbromarone, an old uricosuric drug, inhibits human fatty acid binding protein 4 in vitro and lowers the blood glucose level in db/db mice», en, *Acta Pharmacologica Sinica*, vol. 34, no. 11, pp. 1397–1402, Nov. 2013, ISSN: 1671-4083, 1745-7254. DOI: 10.1038/aps.2013.97.

[100] H. Cai, T. Wang, Z. Yang, *et al.*, «Combined Virtual Screening and Substructure Search for Discovery of Novel FABP4 Inhibitors», en, *Journal of Chemical Information and Modeling*, vol. 57, no. 9, pp. 2329–2335, Sep. 2017, ISSN: 1549-9596, 1549-960X. DOI: 10.1021/acs.jcim.7b00364.

[101] H. Kühne, U. Obst-Sander, B. Kuhn, *et al.*, «Design and synthesis of selective, dual fatty acid binding protein 4 and 5 inhibitors», en, *Bioorganic & Medicinal Chemistry Letters*, vol. 26, no. 20, pp. 5092–5097, Oct. 2016, ISSN: 0960894X. DOI: 10.1016/j.bmcl.2016.08.071.

[102] Q. Xu, L. Huang, J. Liu, *et al.*, «Design, synthesis and biological evaluation of thiazole- and indole-based derivatives for the treatment of type II diabetes», en, *European Journal of Medicinal Chemistry*, vol. 52, pp. 70–81, Jun. 2012, ISSN: 02235234. DOI: 10.1016/j.ejmech.2012.03.006.

[103] L. Chen, B. Wang, H. Li, *et al.*, «Design, synthesis, and biological evaluation of novel highly selective non-carboxylic acid FABP1 inhibitors», en, *European Journal of Medicinal Chemistry*, vol. 276, p. 116 705, Oct. 2024, ISSN: 02235234. DOI: 10.1016/j.ejmech.2024.116705.

[104] ——, «Designs, synthesis and biological evaluation of novel highly selective non-carboxylic acid FABP1 inhibitors», en, *European Journal of Medicinal Chemistry*, vol. 276, p. 116 705, Oct. 2024, ISSN: 02235234. DOI: 10.1016/j.ejmech.2024.116705.

[105] C. Li, B. Cheng, S. Fang, *et al.*, «Design, syntheses and lipid accumulation inhibitory activities of novel resveratrol mimics», en, *European Journal of Medicinal Chemistry*, vol. 143, pp. 114–122, Jan. 2018, ISSN: 02235234. DOI: 10.1016/j.ejmech.2017.11.017.

[106] Y.-L. He, M.-T. Chen, T. Wang, *et al.*, «Development of FABP4/5 inhibitors with potential therapeutic effect on type 2 Diabetes Mellitus», en, *European Journal of Medicinal Chemistry*, vol. 224, p. 113 720, Nov. 2021, ISSN: 02235234. DOI: 10.1016/j.ejmech.2021.113720.

[107] S. Yang, S. Li, and J. Chang, «Discovery of Cobimetinib as a novel A-FABP inhibitor using machine learning and molecular docking-based virtual screening», en, *RSC Advances*, vol. 12, no. 21, pp. 13 500–13 510, 2022, ISSN: 2046-2069. DOI: 10.1039/D2RA01057G.

[108] Y. Wang, W.-K. Law, J.-S. Hu, *et al.*, «Discovery of FDA-Approved Drugs as Inhibitors of Fatty Acid Binding Protein 4 Using Molecular Docking Screening», en, *Journal of Chemical Information and Modeling*, vol. 54, no. 11, pp. 3046–3050, Nov. 2014, ISSN: 1549-9596, 1549-960X. DOI: 10.1021/ci500503b.

[109] F. Lehmann, S. Haile, E. Axen, *et al.*, «Discovery of inhibitors of human adipocyte fatty acid-binding protein, a potential type 2 diabetes target», en, *Bioorganic & Medicinal Chemistry Letters*, vol. 14, no. 17, pp. 4445–4448, Sep. 2004, ISSN: 0960894X. DOI: 10.1016/j.bmcl.2004.06.057.

[110] H. Su, Y. Zou, G. Chen, *et al.*, «Exploration of Fragment Binding Poses Leading to Efficient Discovery of Highly Potent and Orally Effective Inhibitors of FABP4 for Anti-inflammation», en, *Journal of Medicinal Chemistry*, vol. 63, no. 8, pp. 4090–4106, Apr. 2020, ISSN: 0022-2623, 1520-4804. DOI: `10.1021/acs.jmedchem.9b02107`.

[111] H. Wang, C. Zhu, M. M. Swamynathan, *et al.*, «Fatty acid binding protein 5 inhibitors as novel anticancer agents against metastatic castration-resistant prostate cancer», en, *Bioorganic & Medicinal Chemistry*, vol. 122, p. 118 136, May 2025, ISSN: 09680896. DOI: `10.1016/j.bmc.2025.118136`.

[112] M. W. Elmes, M. Kaczocha, W. T. Berger, *et al.*, «Fatty Acid-binding Proteins (FABPs) Are Intracellular Carriers for 9-Tetrahydrocannabinol (THC) and Cannabidiol (CBD)», en, *Journal of Biological Chemistry*, vol. 290, no. 14, pp. 8711–8721, Apr. 2015, ISSN: 00219258. DOI: `10.1074/jbc.M114.618447`.

[113] M. Kaczocha, S. Vivieca, J. Sun, *et al.*, «Fatty Acid-binding Proteins Transport N-Acylethanolamines to Nuclear Receptors and Are Targets of Endocannabinoid Transport Inhibitors», en, *Journal of Biological Chemistry*, vol. 287, no. 5, pp. 3415–3424, Jan. 2012, ISSN: 00219258. DOI: `10.1074/jbc.M111.304907`.

[114] A. V. Hertzel, K. Hellberg, J. M. Reynolds, *et al.*, «Identification and Characterization of a Small Molecule Inhibitor of Fatty Acid Binding Proteins», en, *Journal of Medicinal Chemistry*, vol. 52, no. 19, pp. 6024–6031, Oct. 2009, ISSN: 0022-2623, 1520-4804. DOI: `10.1021/jm900720m`.

[115] Y. Zhou, M. W. Elmes, J. M. Sweeney, *et al.*, «Identification of Fatty Acid Binding Protein 5 Inhibitors Through Similarity-Based Screening», en, *Biochemistry*, vol. 58, no. 42, pp. 4304–4316, Oct. 2019, ISSN: 0006-2960, 1520-4995. DOI: `10.1021/acs.biochem.9b00625`.

[116] Y. He, H. Dou, D. Gao, *et al.*, «Identification of new dual FABP4/5 inhibitors based on a naphthalene-1-sulfonamide FABP4 inhibitor», en, *Bioorganic & Medicinal Chemistry*, vol. 27, no. 19, p. 115 015, Oct. 2019, ISSN: 09680896. DOI: `10.1016/j.bmc.2019.07.031`.

[117] H. Huang, A. L. McIntosh, G. G. Martin, *et al.*, «Inhibitors of Fatty Acid Synthesis Induce PPAR -Regulated Fatty Acid -Oxidative Genes: Synergistic Roles of L-FABP and Glucose», en, *PPAR Research*, vol. 2013, pp. 1–22, 2013, ISSN: 1687-4757, 1687-4765. DOI: `10.1155/2013/865604`.

[118] U. Tagami, K. Takahashi, S. Igarashi, *et al.*, «Interaction Analysis of FABP4 Inhibitors by X-ray Crystallography and Fragment Molecular Orbital Analysis», en, *ACS Medicinal Chemistry Letters*, vol. 7, no. 4, pp. 435–439, Apr. 2016, ISSN: 1948-5875, 1948-5875. DOI: `10.1021/acsmedchemlett.6b00040`.

[119] B. A. Umaru, Y. Kagawa, S. K. Shil, *et al.*, «Ligand Bound Fatty Acid Binding Protein 7 (FABP7) Drives Melanoma Cell Proliferation Via Modulation of Wnt/-Catenin Signaling», en, *Pharmaceutical Research*, vol. 38, no. 3, pp. 479–490, Mar. 2021, ISSN: 0724-8741, 1573-904X. DOI: `10.1007/s11095-021-03009-9`.

[120] J. Chen, X. Liu, S. Zhang, *et al.*, «Molecular mechanism with regard to the binding selectivity of inhibitors toward FABP5 and FABP7 explored by multiple short molecular dynamics simulations and free energy analyses», en, *Physical Chemistry Chemical Physics*, vol. 22, no. 4, pp. 2262–2275, 2020, ISSN: 1463-9076, 1463-9084. DOI: `10.1039/C9CP05704H`.

[121] T. Barf, F. Lehmann, K. Hammer, *et al.*, «N-Benzyl-indolo carboxylic acids: Design and synthesis of potent and selective adipocyte fatty-acid binding protein (A-FABP) inhibitors», en, *Bioorganic & Medicinal Chemistry Letters*, vol. 19, no. 6, pp. 1745–1748, Mar. 2009, ISSN: 0960894X. DOI: `10.1016/j.bmcl.2009.01.084`.

[122] X. Liu, X. Huang, W. Lin, *et al.*, «New aromatic substituted pyrazoles as selective inhibitors of human adipocyte fatty acid-binding protein», en, *Bioorganic & Medicinal Chemistry Letters*, vol. 21, no. 10, pp. 2949–2952, May 2011, ISSN: 0960894X. DOI: `10.1016/j.bmcl.2011.03.063`.

[123] H. Cai, Q. Liu, D. Gao, *et al.*, «Novel fatty acid binding protein 4 (FABP4) inhibitors: Virtual screening, synthesis and crystal structure determination», en, *European Journal of Medicinal Chemistry*, vol. 90, pp. 241–250, Jan. 2015, ISSN: 02235234. DOI: `10.1016/j.ejmech.2014.11.020`.

[124] S. J. Bae, J. E. Kim, Y. J. Choi, *et al.*, «Novel Function of -Cubebenoate Derived from Schisandra chinensis as Lipogenesis Inhibitor, Lipolysis Stimulator and Inflammasome Suppressor», en, *Molecules*, vol. 25, no. 21, p. 4995, Oct. 2020, ISSN: 1420-3049. DOI: `10.3390/molecules25214995`.

[125] T. T. Schug, D. C. Berry, N. S. Shaw, *et al.*, «Opposing Effects of Retinoic Acid on Cell Growth Result from Alternate Activation of Two Different Nuclear Receptors», en, *Cell*, vol. 129, no. 4, pp. 723–733, May 2007, ISSN: 00928674. DOI: `10.1016/j.cell.2007.02.050`.

[126] Y. He, S. Li, Y. Zhu, *et al.*, «Optimization of potent, selective and orally bioavailable biphenyl scaffold as FABP4 inhibitors for anti-inflammation», en, *European Journal of Medicinal Chemistry*, vol. 253, p. 115 319, May 2023, ISSN: 02235234. DOI: `10.1016/j.ejmech.2023.115319`.

[127] H.-H. Cho, H. S. Park, S.-H. Jang, *et al.*, «Rotenoisin A is a novel anti-adipogenic compound», en, *Bioorganic & Medicinal Chemistry Letters*, vol. 29, no. 1, pp. 89–96, Jan. 2019, ISSN: 0960894X. DOI: `10.1016/j.bmcl.2018.11.008`.

[128] S. Yan, M. W. Elmes, S. Tong, *et al.*, «SAR studies on truxillic acid mono esters as a new class of antinociceptive agents targeting fatty acid binding proteins», en, *European Journal of Medicinal Chemistry*, vol. 154, pp. 233–252, Jun. 2018, ISSN: 02235234. DOI: 10.1016/j.ejmech.2018.04.050.

[129] H. Wang, A. Taouil, M. Awwa, *et al.*, «SAR study on Novel truxillic acid monoester-Based inhibitors of fatty acid binding proteins as Next-Generation antinociceptive agents», en, *Bioorganic Chemistry*, vol. 129, p. 106184, Dec. 2022, ISSN: 00452068. DOI: 10.1016/j.bioorg.2022.106184.

[130] C. Zhu, L. Schutz, K. Jayanetti, *et al.*, «Truxillic acid monoamides as fatty acid binding protein 5 inhibitors», en, *Bioorganic & Medicinal Chemistry*, vol. 94, p. 117464, Oct. 2023, ISSN: 09680896. DOI: 10.1016/j.bmc.2023.117464.

[131] N.-S. Tan, N. S. Shaw, N. Vinckenbosch, *et al.*, «Selective Cooperation between Fatty Acid Binding Proteins and Peroxisome Proliferator-Activated Receptors in Regulating Transcription», en, *Molecular and Cellular Biology*, vol. 22, no. 14, pp. 5114–5127, Jul. 2002, ISSN: 1098-5549. DOI: 10.1128/MCB.22.14.5114-5127.2002.

[132] H. Lan, C. C. Cheng, T. J. Kowalski, *et al.*, «Small-molecule inhibitors of FABP4/5 ameliorate dyslipidemia but not insulin resistance in mice with diet-induced obesity», en, *Journal of Lipid Research*, vol. 52, no. 4, pp. 646–656, Apr. 2011, ISSN: 00222275. DOI: 10.1194/jlr.M012757.

[133] Y. Beniyama, K. Matsuno, and H. Miyachi, «Structure-guided design, synthesis and in vitro evaluation of a series of pyrazole-based fatty acid binding protein (FABP) 3 ligands», en, *Bioorganic & Medicinal Chemistry Letters*, vol. 23, no. 6, pp. 1662–1666, Mar. 2013, ISSN: 0960894X. DOI: 10.1016/j.bmcl.2013.01.054.

[134] R. Ringom, E. Axen, J. Uppenberg, *et al.*, «Substituted benzylamino-6-(trifluoromethyl)pyrimidin-4(1H)-ones: A novel class of selective human A-FABP inhibitors», en, *Bioorganic & Medicinal Chemistry Letters*, vol. 14, no. 17, pp. 4449–4452, Sep. 2004, ISSN: 0960894X. DOI: 10.1016/j.bmcl.2004.06.058.

[135] T. Okada, M. Hiromura, M. Otsuka, *et al.*, «Synthesis of BMS-309403-Related Compounds, Including [14C]BMS-309403, a Radioligand for Adipocyte Fatty Acid Binding Protein», en, *Chemical and Pharmaceutical Bulletin*, vol. 60, no. 1, pp. 164–168, 2012, ISSN: 0009-2363, 1347-5223. DOI: 10.1248/cpb.60.164.

[136] W. T. Berger, B. P. Ralph, M. Kaczocha, *et al.*, «Targeting Fatty Acid Binding Protein (FABP) Anandamide Transporters – A Novel Strategy for Development of Anti-Inflammatory and Anti-Nociceptive Drugs», en, *PLoS ONE*, vol. 7, no. 12, A. Catapano, Ed., e50968, Dec. 2012, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0050968.

[137] Y. Zhou, T. Nie, Y. Zhang, *et al.*, «The discovery of novel and selective fatty acid binding protein 4 inhibitors by virtual screening and biological evaluation», en, *Bioorganic & Medicinal Chemistry*, vol. 24, no. 18, pp. 4310–4317, Sep. 2016, ISSN: 09680896. DOI: 10.1016/j.bmc.2016.07.022.

[138] H. Pei, C. Xie, Y. Liu, *et al.*, «Therapeutic potential of a synthetic FABP4 inhibitor 8g on atherosclerosis in ApoE-deficient mice: The inhibition of lipid accumulation and inflammation», en, *RSC Advances*, vol. 6, no. 58, pp. 52 518–52 527, 2016, ISSN: 2046-2069. DOI: 10.1039/C6RA05637G.

[139] A. Ayo, E. Figueras, T. Schachtsiek, *et al.*, «Tumor-Targeting Peptides: The Functional Screen of Glioblastoma Homing Peptides to the Target Protein FABP3 (MDGI)», en, *Cancers*, vol. 12, no. 7, p. 1836, Jul. 2020, ISSN: 2072-6694. DOI: 10.3390/cancers12071836.

[140] Y. Wang, H.-Q. Lin, C.-Y. Xiao, *et al.*, «Using molecular docking screening for identifying hyperoside as an inhibitor of fatty acid binding protein 4 from a natural product database», en, *Journal of Functional Foods*, vol. 20, pp. 159–170, Jan. 2016, ISSN: 17564646. DOI: 10.1016/j.jff.2015.10.031.

[141] T. Sander, J. Freyss, M. Von Korff, *et al.*, «DataWarrior: An Open-Source Program For Chemistry Aware Data Visualization And Analysis», en, *Journal of Chemical Information and Modeling*, vol. 55, no. 2, pp. 460–473, Feb. 2015, ISSN: 1549-9596, 1549-960X. DOI: 10.1021/ci500588j.

[142] G. Warren, M. Osborn, C. Tsantoulas, *et al.*, «Discovery and Preclinical Evaluation of a Novel Inhibitor of FABP5, ART26.12, Effective in Oxaliplatin-Induced Peripheral Neuropathy», en, *The Journal of Pain*, vol. 25, no. 7, p. 104 470, Jul. 2024, ISSN: 15265900. DOI: 10.1016/j.jpain.2024.01.335.

[143] J. Kim and O. De Jesus, «Medication Routes of Administration», eng, in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2025.

[144] B. Chandrasekaran, S. N. Abed, O. Al-Attraqchi, *et al.*, «Computer-Aided Prediction of Pharmacokinetic (ADMET) Properties», en, in *Dosage Form Design Parameters*, Elsevier, 2018, pp. 731–755, ISBN: 978-0-12-814421-3. DOI: 10.1016/B978-0-12-814421-3.00021-X.

[145] E. Gianti and R. J. Zauhar, «Structure–activity relationships and drug design», en, in *Remington*, Elsevier, 2021, pp. 129–153, ISBN: 978-0-12-820007-0. DOI: 10.1016/B978-0-12-820007-0.00007-6.

[146] H. A. Zhong, «ADMET Properties: Overview and Current Topics», en, in *Drug Design: Principles and Applications*, A. Grover, Ed., Singapore: Springer Singapore, 2017, pp. 113–133, ISBN: 978-981-10-5186-9 978-981-10-5187-6. DOI: 10.1007/978-981-10-5187-6_8.

[147] L. Hutchinson and R. Kirk, «High drug attrition rates—where are we going wrong?», en, *Nature Reviews Clinical Oncology*, vol. 8, no. 4, pp. 189–190, Apr. 2011, ISSN: 1759-4774, 1759-4782. DOI: 10.1038/nrclinonc.2011.34.

[148] C. A. Lipinski, F. Lombardo, B. W. Dominy, *et al.*, «Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings», en, *Advanced Drug Delivery Reviews*, vol. 23, no. 1-3, pp. 3–25, Jan. 1997, ISSN: 0169409X. DOI: 10.1016/S0169-409X(96)00423-1.

[149] J. Zhang, H. Li, Y. Zhang, *et al.*, «Computational toxicology in drug discovery: Applications of artificial intelligence in ADMET and toxicity prediction», en, *Briefings in Bioinformatics*, vol. 26, no. 5, bbaf533, Aug. 2025, ISSN: 1467-5463, 1477-4054. DOI: 10.1093/bib/bbaf533.

[150] S. Kar and J. Leszczynski, «Open access in silico tools to predict the ADMET profiling of drug candidates», en, *Expert Opinion on Drug Discovery*, vol. 15, no. 12, pp. 1473–1487, Dec. 2020, ISSN: 1746-0441, 1746-045X. DOI: 10.1080/17460441.2020.1798926.

[151] G. Moroy, V. Y. Martiny, P. Vayer, *et al.*, «Toward in silico structure-based ADMET prediction in drug discovery», en, *Drug Discovery Today*, vol. 17, no. 1-2, pp. 44–55, Jan. 2012, ISSN: 13596446. DOI: 10.1016/j.drudis.2011.10.023.

[152] F. Wu, Y. Zhou, L. Li, *et al.*, «Computational Approaches in Preclinical Studies on Drug Discovery and Development», *Frontiers in Chemistry*, vol. 8, p. 726, Sep. 2020, ISSN: 2296-2646. DOI: 10.3389/fchem.2020.00726.

[153] L. Fu, S. Shi, J. Yi, *et al.*, «ADMETlab 3.0: An updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality and decision support», en, *Nucleic Acids Research*, vol. 52, no. W1, W422–W431, Jul. 2024, ISSN: 0305-1048, 1362-4962. DOI: 10.1093/nar/gkae236.

[154] A. Daina, O. Michielin, and V. Zoete, «SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules», en, *Scientific Reports*, vol. 7, no. 1, p. 42 717, Mar. 2017, ISSN: 2045-2322. DOI: 10.1038/srep42717.

[155]   D. E. V. Pires, T. L. Blundell, and D. B. Ascher, «pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures», en, *Journal of Medicinal Chemistry*, vol. 58, no. 9, pp. 4066–4072, May 2015, ISSN: 0022-2623, 1520-4804. DOI: `10.1021/acs.jmedchem.5b00104`.

[156]   S. Tian, Y. Djoumbou-Feunang, R. Greiner, *et al.*, «CypReact: A Software Tool for in Silico Reactant Prediction for Human Cytochrome P450 Enzymes», en, *Journal of Chemical Information and Modeling*, vol. 58, no. 6, pp. 1282–1291, Jun. 2018, ISSN: 1549-9596, 1549-960X. DOI: `10.1021/acs.jcim.8b00035`.

[157]   P. Schyman, R. Liu, V. Desai, *et al.*, «vNN Web Server for ADMET Predictions», *Frontiers in Pharmacology*, vol. 8, p. 889, Dec. 2017, ISSN: 1663-9812. DOI: `10.3389/fphar.2017.00889`.

[158]   *ADMET Predictor® - Simulations Plus - Machine Learning- ADMET property prediction, https://www.simulations-plus.com/software/admetpredictor/*, en-US.