Politecnico di Torino

Corso di laurea magistrale in Ingegneria Biomedica: Bionanotecnologie

A.a. 2024/2025

Graduation Session 12 2025

Computational analysis of the structural features of the family of FABP proteins and their role in Prostate Cancer:

a comparative study between Homology Modeling and AI powered predictions aimed to establish the FABP12 protein structure

Relatori:

Jacek Adam Tuszinsky

Candidati:

Enrico Astara

Abstract

Prostate cancer (PCa) is the second most common cancer in men after lung cancer and is one of the leading causes of cancer-associated death in men. With cancer risk strongly increasing with age, PCa incidence exhibits a direct correlation with the development index of any given region. Highly developed regions, with higher life expectancy, such as USA, UK, EU [1], show higher incidence in comparison to less developed countries. These regions, however, exhibit a higher rate of annual increase of incidence and higher PCa mortality rates [2]. Localized PCa has excellent prognosis, with 5-year survivability rates of 60-99, with highly effective options of treatment, such as Androgen Deprivation Therapy (ADT), which inhibits one of the major drivers in the disease advancement, the androgen receptor. These values inevitably decrease drastically as the cancer metastasize, degenerating into metastatic castration-sensitive prostate cancer (mCSPC), and eventually develops resistance to treatment (metastatic castration-resistant prostate cancer (mCRPC)). Fatty acids (FAs) and their metabolism play a crucial role in tissues with high rate of growth as an energy source and as metabolic intermediates for membrane biosynthesis, energy storage and the generation of signalling molecules. Due to their hydrophobic nature, FAs transportation has to be achieved either through membrane diffusion or via transportation by specific proteins. Fatty acid binding proteins (FABPs) are cytosolic proteins regulating all functions of cell lipid transportation and storage, and their overexpression has been found to be a marker for tumor advancement and worsening of prognosis. In particular, the most recently discovered member of the FABP family, FABP12, plays an important role in PCa degeneration from local to metastatic. FABP12 inhibition with general FABP inhibitors has however been proven effective in slowing down the process. Despite having identified the primary structure, FABP12 tertiary structure has yet to be elucidated trough experimental methodologies, such as X-ray diffraction or Nuclear Magnetic Resonance (NMR) analysis. To bridge this gap, in order to further the research for more specific inhibitors for FABP12 that might prove effective in arresting the development of metastasis in PCa, this study proposes the use of in silico predictions with different softwares, Molecular Operating Environment (MOE) being the first and employed to perform Homology modelling, the gold standard for in silico protein structure prediction, and the more recent, AI powered software, AlphaFold. First, the sequences of the other FABPs have been compared among themselves to establish reciprocal similarity degree and which structure to use as template for each FABP in the first testing phase, in order to adjust the software's prediction parameters. Then all sequences are compared with FABP12 sequence to establish the best template, that has also been used as control. Structures obtained

through Homology modelling with MOE and through Alphafold predictions have been compared, using different metrics, with experimentally known structure of other FABPs in order to assess the better methodology. Then the possible FABP12 structures obtained with these two softwares have been once again compared and ranked in order to find the most suitable candidate for ligand interaction simulations, which represents the next step in research for more specific FABP12 inhibitors.

Acknowledgements

I would like to thank my colleagues, Gabriele DeRosa and Alessia Davoglio, with whom I shared a wonderful experience of work, struggles, travels, dinners, and to whom I apologize for my constant lateness; my supervisor, Professor Jacek Adam Tuszinsky, for giving me this opportunity; my family and friends, for supporting me even though I might have tried to be a bit easier to support; and my girlfriend, my wonderful Anna, that made me swear to never step in a laboratory or hurt a single mouse, and to whom I dedicate this work.

Table of Contents

List of Tables

List of Figures

Chapter 1

Prostate Cancer: overview of the disease

1.1 Introduction to the organ

The prostate gland is located beneath the bladder, surrounding the urethra. Its main function, as a male reproductive accessory organ, is the addition of secretion to semen, which constitutes ejaculate and maintains sperm viability. The prostate gland is essential to achieve the necessary conditions for male fertility, but is also a direct target of several prevalent benign and malignant diseases (e.g., prostatitis, benign prostatic hyperplasia (BPH), and prostate cancer) that are potentially linked with impaired fertility status[3]. Prostatitis has the greatest potential to affect fertility [4]. While it was already found that prostatic inflammation plays a role in the development of prostate cancer [4], more recent data also shows its involvement in BPH [5]. The prostate of an adult human is comprised of central, transition, and peripheral zones while also containing fibromuscular and periurethral regions (Figure1.1-a). The transition zone (5-10%) surrounds the urethra proximal to the ejaculatory ducts. A fibromuscular band of tissue separates the transition zone from the remaining glandular compartments. The central zone (20-25%) surrounds the ejaculatory ducts and projects under the bladder base. The peripheral zone constitutes the bulk of the apical, posterior, and lateral aspects of the prostate. While this last region makes the largest contribution to healthy prostate function, it also is the most common origin site (>80%) of neoplasms in the aged prostate [2]. . At the microscopic level (Figure1.1-b), the prostate is composed by a fibromuscular stroma which surrounds a glandular epithelium constituted by two histologically different layers. The secretory luminal layer is made of columnar cells that secrete PSA, prostatic acid phosphatase, and human kallikrein-2 as part of the seminal fluid. Under this layer is a basal layer of cuboidal epithelial and neuroendocrine cells,

which is in turn lined by an extracellular matrix membrane. This last membrane divides the basal cells from the stroma.

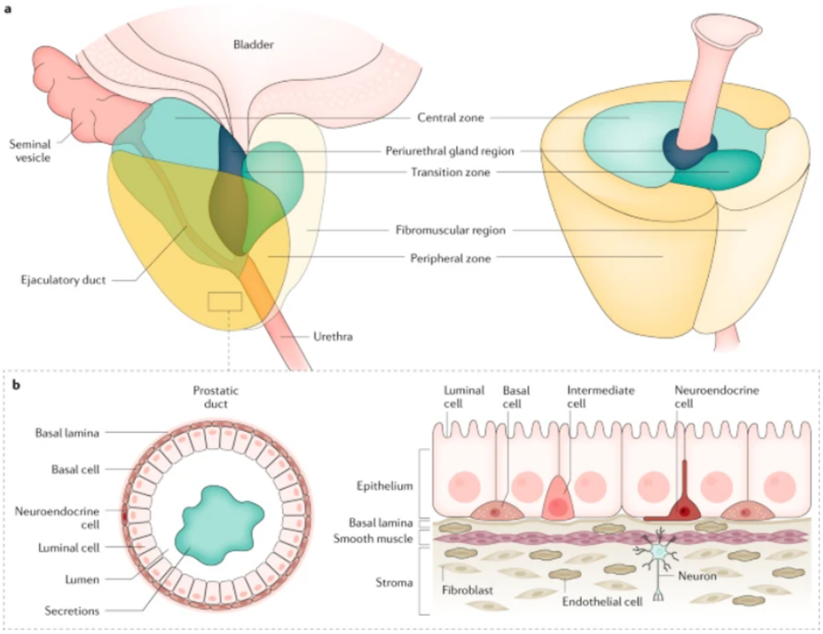

Figure 1.1: a) anatomical structure of the prostate; b) microscopic and histological description.[2]

As shown in Figure1.1-b each region comprises ducts and acini embedded in the stroma, which contains various cell types, predominantly smooth muscle cells but also fibroblasts, which have important roles in prostate development. The ducts and acini comprise a single layer of columnar epithelium (AR+, CK8+, CK18+, PSA+), surrounded by a layer of basal epithelial cells (CK5+, CK14+, p63+), which produce the basement membrane, a layer of extracellular matrix that is anchored to the stromal cells (α-SMA+, vimentin+). Neuroendocrine cells (Syn+, CGA+, NSE+) are also present within the duct [3]. The main function of the stromal compartment is to provide supportive signals to retain or restore gland homeostasis, to maintain the microenvironment of the epithelial compartment. However, studies indicate that activated stroma might have a role in prostate inflammatory processes [6]

1.2 Inhibitory functions of Zn^{2+} and AR mediated metabolic control

An important factor that should be considered about the potential development of prostate cancer is the prostatic epithelial cells metabolism: these are the only healthy human cells that actively produce energy by glycolysis (hallmark of proliferating cancer cells) rather than the Krebs cycle [3]. This property relies on the androgen-related accumulation of Zn^{2+} in the prostatic epithelial cells: the proteins ZIP1–4 (uptake) and ZnT1–10 (release) are the specific zinc transporters that mediate this cycle [7][8]. The Zn^{2+} accumulated in the prostatic epithelial cells accounts for about 4% of the total Zn^{2+} content in the human male body. Its first inhibitory function is blocking the initial step of the Krebs cycle — the oxidation of citrate in isocitrate, as performed by mitochondrial aconitase (mACON), a key enzyme for ATP production by respiration and terminal oxidation by mitochondria, leading to accumulation of citrate [7]. Citrate is a necessary substrate for the free intracytoplasmatic Zn^{2+} pool in order to specifically recognize mACON and not its cytoplasmic isoform[8]. The androgen receptor (AR) is the main intracellular effector of male sex steroids, which regulate intraprostatic accumulation of Zn^{2+} and citrate, inhibition of the Krebs cycle and prostatic fluid release. The AR belongs to the subfamily of nuclear receptors, which have a double function of intracellular receptors and ligand-activated transcription factors[9], allowing them to be activated by 5α-dihydrotestosterone (DHT). This is an androgen obtained by the conversion of testosterone, mediated by the enzyme 5α-reductase, in order to have a molecule with stronger binding affinity with AR [10]. Testosterone's and, therefore DHT's, circulating and intraprostatic levels decrease with age, causing an in-kind trend in the proper functioning of the prostate and its ability to maintain proper tissue levels of intracellular Zn^{2+}, citrate, and KLK-secreted proteins within the prostatic fluid. Moreover, Zn transporters essential for Zn^{2+} uptake (ZIP1–4) and release (ZnT1–10), are also androgen regulated. This not only affects male fertility, but also leads to metabolic changes, since it affects the inhibition of the Krebs cycle, strongly dependent on healthy levels of Zn^{2+}. Without this ion-mediated inhibition there is a raise in production of ATP trough mitochondria oxidation, which favours the development of neoplastic tissue.

1.3 Cancer

Prostate cancer (PCa) affects millions of men globally, with most of the diagnoses in highly developed regions. Localized disease generally has a favourable outcome of a 99% survival rate for 10 years in case of early detection and immediate treatment. Diagnosis usually comes from a prostate biopsy prompted by a blood test to measure prostate-specific antigen (PSA screening) and/or digital rectal examination. Treatment for localized diseases includes active surveillance, radical prostatectomy, ablative radiotherapy, and androgen deprivation therapy (ADT) in case of local relapse or ADT combined with chemotherapy and/or androgen signalling-targeted agents in case of systemic relapse[2]. Prostate cancer risk increases with age, meaning prostate cancer incidence is bound to be higher in regions with high life expectancy. PCa rates in 2022 varied 13-fold for incidence and 9.5-fold for mortality, with the highest incidence rates in Australia/New Zealand, North America, Northern Europe, and Latin America/Caribbean, and the highest mortality rates were in sub-Saharan Africa and Latin America/Caribbean, as shown in Table1 [11]. This highlights a positive correlation between incidence and the human development index (HDI): developed nations tend to present higher life expectancy and therefore higher incidence rates compered to developing ones. The highest incidence rates were observed in Micronesia/Polynesia, the USA, European countries, and Brazil, countries with higher HDI ($r = 0.58$) and per capita GDP ($r = 0.62$). South Asia, Central Asia, and sub-Saharan Africa, which encompass most of the low-income nations, currently have the lowest incidence but the highest rates of incidence increase[12]. Although approximately 90% of PCa cases are found in men without a family history of the disease, familial cancer history and prostate cancer risk are highly correlated. From familial history, various data are gathered to assess risk, such as the number of affected individuals (in case of >3 affected relatives PCa is categorized as a familial disease), the degree of relation and age at disease onset. Men with affected first-degree relatives have a much higher risk of developing the disease[13]. The majority of patients are diagnosed with prostate-limited localized PCa (circa 80%), 5% are diagnosed with distant metastases (often in multiple sites), and 15% of them are diagnosed with locoregional metastases [14]. At the metastatic stage PCa spreads almost exclusively to locoregional lymph nodes and/or the hematogenous spread to the stroma of the bone marrow in the axial skeleton, with more than 80% of distant metastatic lesions are found in the bone tissue [15], leading to physiological and anatomical dysfunction (for example, activating bone remodelling). Patients at this stage experience bone-related effects and are prone to spontaneous fractures resulting in spinal cord compression. Advanced prostate cancer often progresses despite ADT and is then defined as castration resistant. This development of the disease holds a particularly heavy weight in case of late detection, where PCa may

already be in the metastatic phase (5-year rate of survival of 30-40%[14]). In this situation PCa progresses from localized to metastatic castration-sensitive prostate cancer (mCSPC) and metastatic castration-resistant prostate cancer (mCRPC) and is considered a terminal-stage disease [16].

Area	Population size		Cases			Deaths		
	N	%	N	%	ASR	N	%	ASR
Africa	703 122 445		103 050		30.3	55 744		17.3
Eastern Africa	232 760 186	33.1	25 431	24.7	28	14 808	26.6	16.8
Middle Africa	95 067 662	13.5	16 011	15.5	44.2	9401	16.9	27.2
Northern Africa	128 138 567	18.2	16 686	16.2	16.1	6958	12.5	6.9
Southern Africa	34 025 573	4.8	13 691	13.3	59.9	5790	10.4	29.7
Western Africa	213 130 457	30.3	31 231	30.3	36.9	18 787	33.7	23.5
Asia	2 374 486 911		386 424		12.6	120 485		3.8
Eastern Asia	822 363 685	34.6	255 404	66.1	15.1	64 618	53.6	3.5
South-Central Asia	1 060 915 547	44.7	60 131	15.6	6.4	29 194	24.2	3.1
South-Eastern Asia	340 253 165	14.3	42 298	10.9	12.7	17 443	14.5	5.3
Western Asia	150 954 514	6.4	28 591	7.4	24	9230	7.7	7.9
Europe	361 242 024		473 011		59.9	115 182		11.2
Eastern Europe	137 314 603	38	122 189	25.8	48.8	36 407	31.6	13.5
Northern Europe	52 963 352	14.7	95 448	20.2	82.8	21 232	18.4	12.4
Southern Europe	74 143 121	20.5	96 952	20.5	54.4	22 182	19.3	8.1
Western Europe	96 820 948	26.8	158 422	33.5	65.7	35 361	30.7	10.2
Latin America/Caribbean	327 201 180		225 985		58	61 056		13.9

Table 1.1: ASR = age-standardized world rate per 100 000 men. Percentages shown for the population, cases, and deaths are within each region. Source: GLOBOCAN 2022[11]; Part 1

Area	Population size		Cases			Deaths		
	N	%	N	%	ASR	N	%	ASR
Caribbean	21 727 133	6.6	22 271	9.9	73.9	9695	15.9	26.4
Central America	90 053 341	27.5	35 982	15.9	39.8	10 311	16.9	10.2
South America	215 420 706	65.8	167 732	74.2	62.4	41 050	67.2	13.7
North America	184 781 027		255 782		73.5	39 605		8.3
Oceania	21 902 160		23 602		71.9	5358		11.5
Australia/New Zealand	15 386 803	70.3	21 982	93.1	78.1	4818	89.9	11
Melanesia	5 881 650	26.9	1286	5.4	34.3	431	8	14.4
Micronesia	282 593	1.3	117	0.5	43.3	37	0.7	15.4
Polynesia	351 114	1.6	217	0.9	58.9	72	1.3	19.7
World	3 972 735 747		1 467 854		29.4	397 430		7.3

Table 1.2: ASR = age-standardized world rate per 100 000 men. Percentages shown for the population, cases, and deaths are within each region. Source: GLOBOCAN 2022[11]; Part 2

PCa tumorigenesis is believed to be associated with the accumulation of somatic mutations in the prostate epithelial cells' genome. Localized PCa normally present a relatively lower number of genomic aberrations than other types of cancer, in contrast with what happens at the metastatic stage, which exhibits a far higher mutation rate and frequency of copy number alterations (CNAs)[17]. Some of these aberrations, such as Single Nucleotide Polymorphism (SNP), have been suggested as markers for PCa or for its advancement: a recent study has revealed that SNPs in AR were not observed in localized disease, while SNPs in TP53 were significantly more prevalent in mCRPC, and that SPOP SNPs were less prevalent in mCRPC[18]. Recent genome-wide association studies (GWAS) have identified the association of SNP rs11672691 on chromosome 19q13 with aggressive prostate cancer (PCa). Specifically, Gao et al. [19] found an association of the aggressive PCa-associated allele G of rs11672691 with elevated transcript levels of two biologically plausible candidate genes, PCAT19 and CEACAM21, involved in PCa cell growth and tumor progression.

1.4 Treatment

To establish the most effective treatment, the first necessary step is a risk assessment based off PSA screening, DRE, and biopsy results, which give information on cancer staging and life expectancy. Historically, another methodology of classification for PCa was based on the Gleason score. This grading system was devised in the 1960s and 1970s by Dr. Donald F. Gleason and members of the Veterans Administration Cooperative Urological Research Group and is based entirely on the histologic arrangement pattern of carcinoma cells in H&E-stained sections. The method categorizes histologic patterns at relatively low magnification (× 10–40) by the extent of glandular differentiation and the pattern of growth of the tumor in the prostatic stroma. Dr. Gleason defined five grade patterns, shown in the drawing in Figure2. The five basic grade patterns are used to generate a histologic score, which can range from 2 to 10, by adding the primary grade pattern and the secondary grade pattern. The primary pattern is the one that is predominant in area, by simple visual inspection. The secondary pattern is the second most common pattern [20].

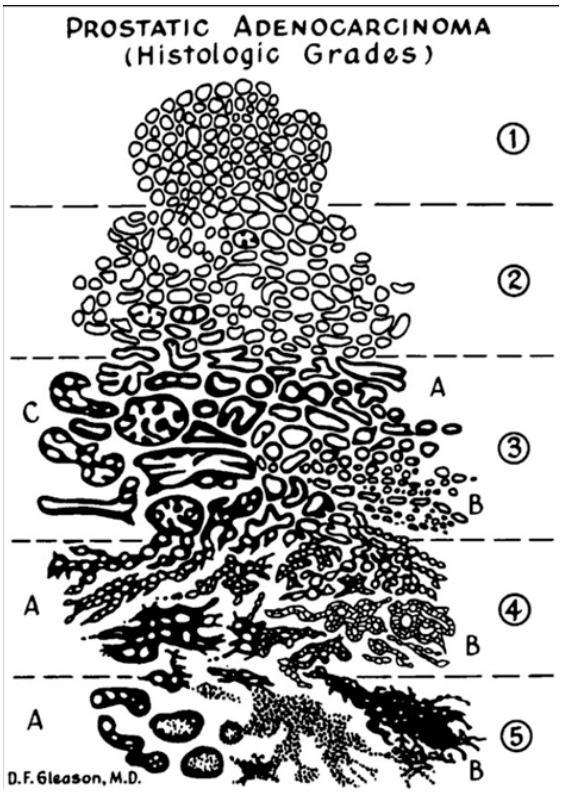

Figure 1.2: Hand-drawn scheme, devised by Dr.Gleason for classification of tumor histological patterns and score determination[20]

However, the grading system was reorganized into the International Society of Urological Pathology (ISUP) grade groups 1–5[21]. Depending on the results the patient may first be deemed suitable for active surveillance, in order to avoid overtreatment of a slow-growing, sometimes even indolent, kind of tumour. Active surveillance is carried out until it's determined that further steps must be taken to halt the development of the disease, keeping it in a curable stage[22], radiotherapy has been shown to be effective in 60% of localized disease cases[23], though both current surgery and radiation therapies are not ideal when only partial or subtotal tissue removal is required instead of radiation or full-organ removal. For metastatic disease the focus must be shifted on systemic treatment, based on circulating androgen reduction and ADT, to which localized PCa also almost universally responds. Antiandrogens were originally given with ADT in a combined androgen blockade. Almost all patients with PCa ultimately develop castration resistant prostate cancer (CRPC), which is refractory to ADT. Second-generation antiandrogens, such as enzalutamide, darolutamide, apalutamide, and abiraterone acetate, and radiotherapy, including external beam radiation therapy (EBRT) with X-ray beams and radiopharmaceuticals, including Ra-223 and 117Lu-PSMA-617 [24], and immunotherapy, including sipuleucel-T, dostarlimab, and pembrolizumab, have been approved and are available for treating patients with mCRPC and have been tested in multiple combinations[25][26].

1.5 Disease degeneration: development of resistance to ADT

The androgen receptor (AR) (NR3C4, nuclear receptor subfamily 3, group C, gene 4) is found in the prostate's luminal cells, fibromuscular stromal cells, and endothelial cells. AR belongs to the steroid hormone group of nuclear receptors with the oestrogen receptor (ER), glucocorticoid receptor (GR), progesterone receptor (PR), and mineralocorticoid receptor (MR). AR is composed of four distinct domains: the N-terminal domain, DNA binding domain (DBD), a hinge region which allows for N- and C-terminal interaction, and a C-terminal ligand binding domain (LBD) [27]. In the cytoplasm, AR is normally bound to several chaperone proteins, members of the heat-shock protein family (e.g., HSP90, HSP70)[28]. Upon activation, androgens bind to the LBD, releasing AR chaperones and allowing AR to homodimerize and translocate to the nucleus, where it acts as a transcription factor[29]. In healthy conditions, AR's transcription activity regulates gene expression and encoding of proteases (e.g., PSA) necessary to maintain prostate homeostasis and its binding to the native specific ligands, 5α-dihydrotestosterone (DHT) and testosterone, and initiates male sexual development and differentiation[30].In the diseased state, AR coordinates the synthesis of PSA, the regulation of lipid metabolism, and drives

tumorigenesis, promoting growth-related transcription program[31]. There are four possible mechanisms of CRPC development: 1) Increased sensitivity of the AR to its agonists, 2) AR mutations that render the receptor responsive to alternate, non-androgen ligands, 3) ligand-independent AR activation, and 4) AR-independent mechanisms[32].

The first mechanism which allows PCa to evolve into CRPC is overexpression of AR due to AR gene amplification. Even with ADt, a small quantity of androgens is still present, and PCa cells with AR amplification can survive and progress to CRPC[32], indicating that CRPC cells may not be strictly androgen independent, but rather, they become more sensitive due to a lowered threshold for androgens[30][32]. Second mechanism is ligand promiscuity due to mutation in the LBD, resulting in loss of specificity. Genes expressing AR are the most frequently mutated in CRPC patients (62.7%), with point mutations, found in 15% to 30% of CRPC patients, that enable activation of AR inducing loss of specificity of the agonist[32]. Epithelial AR can promote growth through induction of fusion between the AR-regulated TMPRSS2 gene promoter and the coding region of the ETS family members erythroblast transformation-specific (ERG) and ETS variant 1 (ETV1)[33], estimated to occur in around 50% of prostate tumors. The TMPRSS2:ETS factor fusion causes androgen responsiveness to ETS transcription factors, which leads to cell-cycle progression [30]. Point mutation of T878A results in activation of AR with Progesterone, oestrogen, flutamide, bicalutamide, and enzalutamide[34][35][36]. The third mechanism involves ligand-independent AR activation: insulin-like growth-factor-1 (IGF-1), keratinocyte growth factor (KGF), and epidermal growth factor (EGF), can activate the AR as a consequence of activating the downstream PI3K/AKT/mTOR pathway, thus creating an "outlaw receptor"[30]. Stromal AR can also promote growth of PCa cells trough cross-talks with growth factors, such as transforming growth factor β (TGFβ), vascular endothelial growth factor (VEGF), insulin-like growth factor (IGF), fibroblast growth factor (FGF), and epidermal growth factor (EGF), which, in turn, activate their respective receptors[33]. Fourth and last mechanism is particularly important, since it does not involve AR. Castration-induced apoptosis causes production of proinflammatory factors by dying cancer cells, leading to infiltration of B and T cells: B cells produce lymphotoxin and factors that increase Stat3 signalling, thus promoting hormone-free survival of prostate cancer cells[37]. Other pathways, like the one involving anti-apoptotic protein Bcl-2[38], or the one involving upregulation of expression of the Glucocorticoid receptor[39], , lead to a similar result, showing a protective effect on PCa cells. These four mechanisms may also act synergistically to enhance AR activation despite ADt, however they do not cause a loss of AR signalling but rather lead to aberrant AR activation trough non-physiological pathways.

Chapter 2

The role of Fatty Acids and of their chaperons in Cancer

2.1 The Fatty Acid Binding Protein family

Fatty acid binding proteins (FABPs) belong to the intracellular lipid-binding protein (iLBP) family and function as chaperones for intracellular hydrophobic ligands and trafficking them throughout cellular compartments, including the peroxisomes, mitochondria, endoplasmic reticulum and nucleus. FABPs are relatively small cytosolic proteins, with molecular weights of approximately 14-15 kDa, with two really distinctive domains: the first is constituted by two alpha-helices in the superior surface, thought to regulate access to the binding pocket; the second is made of ten anti-parallel beta sheets forming a beta-barrel that surrounds the binding pocket[39]. FABPs regulate metabolic pathways, signal transduction, and gene expression by reversely binding and transporting fatty acids and lipid-related molecules, such as fatty acids (FAs) and their acyl-CoA derivatives (FA-CoA) with high affinity but broad specificity, binding long-chain (C16-C20) fatty acids, eicosanoids, bile salts and peroxisome proliferators. In total there are ten FABPs encoded in the human genome, exhibiting a degree of similarity ranging from 20% to 70% , including liver-FABP(L-FABP, FABP1), intestine-FABP (I-FABP, FABP2), heart-FABP (H-FABP, FABP3), adipocyte-FABP (A-FABP, FABP4), epidermal-FABP (E-FABP, FABP5), ileal-FABP(Il-FABP, FABP6), brain-FABP (B-FABP, FABP7), myelin-FABP (M-FABP, FABP8) and testis-FABP (T-FABP, FABP9) and lastly the FABP12, which is currently being studied. FABP10 and FABP11 are not seen in humans, and are expressed only in other species, such as zebrafish(Danio Rerio) and teleost fish (Solea senegalensis)[40][41]. The FABPs' denomination comes from the organ in which they were first identified, but it's not indicative of their exclusive presence in that organ (FABP1 is expressed in

the liver, intestine, pancreas, kidney, lung and stomach). Even though FABPs share a common tertiary structure, differences in primary structures allow to divide this family in three groups: (1) FABP1 and FABP6; (2) FABP3, FABP7, FABP5, FABP8, FABP4 and FABP9; and (3) FABP2. All proteins of group 1 can bind FAs and bulky ligands, such as bile salts, cholesterol and haem; members of group 2 bind FAs and additionally retinoids and eicosanoids; finally, FABP2 binds solely FAs, but in a different conformation than other FABPs (bent instead of U-shaped) [42]. The β-barrel structures displayed by the FABPs is made of two five-stranded β-sheets arranged approximately perpendicular to each other. The β-sheets wrap around a solvent-accessible ligand binding pocket. The strands are linked by β-turns except for strands 1 and 2, which are connected by a with 2 8-10 residue helix–turn–helix motif, and strands 7 and 8, which are joined by an Ω loop[43]. The two helices are thought to act as a "portal" for ligand access, forming a dynamic region characterised by a long-range relationship between α-helix II and the turns between βC-βD and βE-βF loops[44].

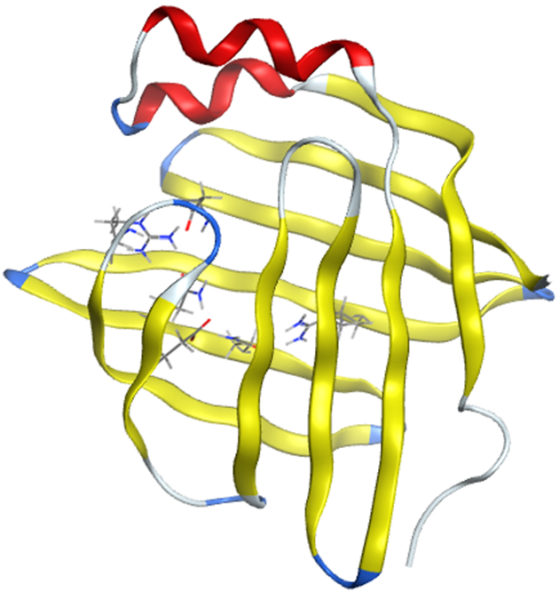

Figure 2.1: FABP8 structure, obtained from loading a .pdb file(PDBID:4BVM) on MOE software

In general, the FABPs' binding pocket is significantly larger than its ligand, with the maximum binding pocket volume reached by FABP1 with 440 Å^3 compared to those of other iLBPs (210–330 Å^3) and other FABPs (FABP8 (330 Å^3), FABP3

(323 Å^3), FABP4 (310 Å^3) or FABP2 (234 Å^3)), allowing the FABP1 to be the only one in its protein family able to bind an additional FA as well as larger hydrophobic molecules[45]. This exceptional feature is due to the presence of seven key residues (S39, N61, T93, S100, T102, N111, and S124) with smaller side chains than in the other iLBP family members[46]. All FABPs bind both saturated and unsaturated long-chain ($\geq^{14} C$) FAs, with the carboxylate group oriented towards the interior of the binding pocket and interacting with Arg and Tyr residues and other side chains[47]. X-ray crystallography and nuclear magnetic resonance (NMR) studies have also revealed FA-binding characteristics within the water-filled cavity, where bound FAs occupy approximately one-third the total pocket volume and share the cavity with many ordered and disordered water molecules. X-ray crystallography and nuclear magnetic resonance (NMR) studies have also revealed FA-binding characteristics within the water-filled cavity, where bound FAs occupy approximately one-third the total pocket volume and share the cavity with many ordered and disordered water molecules [47]. While none of the FABPs shows particular specificity to any of the FAs, binding affinity values, obtained using the ADIFAB method, have been shown to be directly correlated with the FA's hydrophobicity and indicate dissociation constant (K_d) values in the nanomolar range [47][48]. However, functional properties unique to each FABP may be dictated more by structural characteristics of the proteins' surfaces rather than by ligand binding specificity. In vitro studies have already shown that different FABPs transfer FA to membranes by different mechanisms: while FABP1 transfers ligand through membranes by aqueous phase diffusion, most of the other FABPs, such as FABP2, FABP3, FABP4, FABP5, FABP7, and FABP8, operate the transfer with direct membrane interaction [48]. Most determining structural factor for this different transfer mechanism is the FABP helix-turn-helix/portal domain: the amphipathic character of the α-I helices and net positive surface electrostatic potential across the portal region determine protein-ligand interactions[49].

2.2 Lipids in cancer and FABBPs role

Following studies which highlighted correlation between increased body weight and cancer mortality[50][51], lipid metabolism and their role as an energy source for cancer cells was amply researched[52], Lipid metabolism, in particular the synthesis of fatty acids (FAs), is an essential cellular process that converts nutrients into metabolic intermediates for membrane biosynthesis, energy storage and the generation of signalling molecules. Cancers, in order to perform fermentation independently from oxygen availability, express a high rate of glucose uptake, and convert it into lipids, similarly to what happens in the liver[53][54]. Successive studies discovered that cancer cells generate almost all the necessary cellular FAs

through de novo synthesis, and the fatty acid synthase (FASN) was identified as the tumour antigen OA-519 in aggressive breast cancer [55] [56]. FA synthesis is also triggered by deletion of chromosome 8p in breast cancer cells, suggesting the necessity of FAs availability for cancer growth and spreading[57] FA synthesis requires cytoplasmic acetyl-CoA as a metabolic intermediate to provide the necessary substrate: malonyl-CoA, transported by carnitine in the mitochondrial matrix where it undergoes degradation by β-oxidation, which also generates energy at multiple steps[52]. Seven malonyl-CoA molecules and one acetyl-CoA primer are condensed by FASN to obtain palmitate, a 16-carbon saturated FA (16:0), which is then elongated and desaturated to produce molecules of various lengths and degrees of saturation [58][59]. FAs generated through de novo synthesis and through exogenous uptake make up the pool of intracellular FAs that can be used for the synthesis of triacylglycerides for energy storage, glycerophospholipids, cardiolipins and sphingolipids for membrane synthesis, and eicosanoids for signalling processes[52]. Actively proliferating tissues require FAs for the synthesis of structural lipids. Thus, induction of lipid synthesis must be closely connected to cell growth, which is a prerequisite for cell division. Another reason to regulate FA uptake and metabolism in cancer disease can be found in the flexibility exhibited by cancers on switching between FA synthesis, lipid uptake and degradation when exposed to fluctuation in the cancer environment like hypoxia, which increases lipid uptake in breast cancer and glioblastoma cells by inducing the expression of FA binding protein 3 (FABP3) and FABP7 [60][61][62]. Inducing severe deprivation of certain nutrients may, however, promote cancer cells independence on specific metabolic processes. Hypoxia prevents de novo FA synthesis, so cancer cells become more dependent on external lipid uptake; limiting also uptake of exogenous lipids may trigger de novo FA synthesis again, but with alternative substrates such as glutamine or acetate[52]. . FAs are also involved in cancer invasion and the development of drug resistance. Lipids are employed in the modification necessary for activity of many signalling molecules, like WNT proteins, frequently deregulated in cancer[63], nd also function as second messengers or ligands for autocrine receptor signalling, involved in proliferation, migration, inflammation and angiogenesis[64].

FABPs, who were previously believed to act only as lipid chaperones, have been found to be involved in cancer tumorigenesis and progression[40]. In the study of [65], Northern Blot analysis has shown that the gene responsible for expression of epidermal FABP (FABP5) was represented 4.9 ± 0.9- to 16.9 ± 2.1 times more in mRNA of malignant prostate and breast cancer cell lines, compared to benign cell lines. The FABP5 gene also induced the benign cells to disseminate to secondary sites when it was overexpressed in the recipient cells in this system. These results demonstrated that elevated expression of FABP5 can induce metastasis and that

metastatic capability can be transferred in a genetically dominated manner. However, expression of C-FABP did not increase with increasing metastatic potential, indicating that FABP5 might have played a more important role in the initiation of malignant changes in the early stage than in the progression of metastasis. The roles of the other FABPs in cancer disease have also been investigated and are here briefly reported:

- FABP1: mainly present in the liver. It is also present in smaller amounts in the intestines, kidneys and the stomach. FABP1 has already been proven to be involved in steatotic liver and non-alcoholic fatty liver disease[66]While some studies have shown that loss of FABP1 contributes to development of microsatellite unstable colorectal carcinomas [40][67], therefore establishing FABP1 as a cancer suppressor, another study highlighted how its interaction with the VEGF receptor leads to angiogenesis and enhancement of migration properties of cancer cells, promoting metastasis of hepatocellular carcinomas (HCC) [68].

- FABP2: expressed in the small intestine, with the highest levels of expression in the jejunum. FABP2 controls the transfer of fatty acids to prevent the accumulation of non-esterified fatty acids and the alterations to membrane characteristics that would ensue [69]. FABP2 polymorphism studies have shown that a threonine substitution on amino acid 54 disturbed lipid metabolism and was also correlated with insulin resistance, hypertriglyceridemia and excessive triglyceride buildup[70][71]. FABP2 has not been extensively studied in cancer disease progression; however, a study highlighted a negative correlation between FABP2 expression and dietary habits and lipid uptake in colonrectal cancer, indicating that FABP2 may not be a good marker for the disease[72]

- FABP3: most widespread of the FABPs. Predominantly found in the heart, it's also present in other high-energy-demanding tissues, such as lungs, ovary, brain, placenta, mammary gland and stomach. FABP3, therefore, acts as a lipid carrier to direct fatty acids to mitochondria to sustain the energy demand. Abnormal levels of FABP3 are correlated with both heart issues and brain diseases: in the first case excessive levels of FABP3 can lead to low calcium concentration in the heart's sarcoplasmic reticulum [73]; In the second case not only FABP3 accumulates in the brain at an even higher rate than FABP7 (brain FABP) taking part in the later developments of the brain, but its low levels may be involved in Down's syndrome and Alzheimer's disease[74]. High expression of FABP3 has been linked to invasion, cancer stage and overall metastasis [75][76]. FABP3 overexpression has also been highlighted in non-small cell lung carcinoma[76],

- FABP3: most widespread of the FABPs. Predominantly found in the heart,

it's also present in other high-energy-demanding tissues, such as lungs, ovary, brain, placenta, mammary gland and stomach. FABP3, therefore, acts as a lipid carrier to direct fatty acids to mitochondria to sustain the energy demand. Abnormal levels of FABP3 have been found to be correlated with both heart issues and brain diseases: in the first case excessive levels of FABP3 can lead to low calcium concentration in the heart's sarcoplasmic reticulum [73]; In the second case not only FABP3 accumulates in the brain at an even higher rate than FABP7 (brain FABP) taking part in the later developments of the brain, but its low levels may be involved in Down's syndrome and Alzheimer's disease[74]. High expression of FABP3 has been linked to invasion, cancer stage and overall metastasis [75][76]. FABP3 overexpression has also been highlighted in non-small cell lung carcinoma[76], gastric cancer[75], leiomyosarcoma[77] and melanoma[78]. On the other hand, FABP3 plays an important role in disrupting cancer signalling pathways, enhancing cancer cells' sensitivity to drugs, thus aiding cancer suppression and metastasis prevention, especially in breast cancer[79][80].

- FABP4: secreted by both adipocytes and macrophages, with the highest rates of secretion during differentiation of the firsts and activation of the seconds. This leads to enhanced glucose production in hepatic cells, increased insulin secretion [81] and decreased cardiomyocyte contraction [82]. More recent studies also proved the involvement of FABP4 in atherosclerosis via inflammation and lipid buildup in macrophages and foam cells. A steep decrease (circa 60%) in the obstruction of mice's coronary arteries in the absence of FABP4 was highlighted [83][84]. High serum levels of FABP4 in patients with obese breast cancer phenotype have been found to be correlated with cancer size[85]. While FABP4 has previously been reported to be involved in the aggressiveness of various cancers, such as prostate cancer[86], breast cancer[87], cholangiocarcinoma[88], glioblastoma[89] and leukemia[90], it has also been found to be involved, as a transcription factor, in lipid-independent cancer progression[90], and in epthelial to mesenchymal transition (EMT), an hallmark of metastasis, promoted by FABP4 overexpression in cholangiocarcinoma[88] and cervical cancer[91].

- FABP5: mostly expressed in the skin, where it acts as a lipid carrier to maintain the epidermis functionality as a barrier against external agents, is also present in brain tissues and, even if not as much as in the two previous cases, in lungs, kidneys, liver and mammary glands [92]. FABP5 in the skin regulates insulin responses, inflammation and water permeability[93], in the brain is involved in energy homeostasis, neuronal regeneration and neurogenesis[94]. FABP5 is also highly expressed in cancer cells and contributes to proliferation, invasiveness, cancer burden, resistance to therapy and low survival [40]. FABP5 exhibits a

stabilising effect on epidermal growth factor receptor (EGFR), which is a key factor in cancer metastasis[95][96]. A study by Al-Jameel et al.[97] has shown how the inability of FABP5 to bind FAs in prostate cancer significantly slowed down disease progression of CRPC, strengthening the connection between FAs and cancer progression. Recently, FABP5 has been found to be upregulated in the nucleus of most cancer cells, suggesting an involvement in transcription of oncogenic factors. However, this has to be further investigated [40].

- FABP6: commonly found in the ileum, it has a high affinity for bile acids. Absence of FABP6 in male mice facilitates fatty liver disease[98]. Regarding cancer, FABP6 has been researched mostly for its role in colon cancer tumorigenesis and has also been deemed poorly correlated with metastasis after a steep decrease in its expression was highlighted in patients with node metastasis[99].

- FABP7: expressed in the central nervous system, with a notably higher presence in the fetal brain rather than the adult, suggesting a role of FABP7 in early stages of development of the CNS[100]. FABP7 also supports radial glial fiber formation and Schwann's cells' regulation to ensure proper neuron migration[101]. FABP7 has been extensively studied in cancer, especially in glioblastoma, where its overexpression has been found to relate to EGFR amplification, leading to reduced survival times[102]. FABP7 was also studied in breast cancer, highlighting its relationship with the triple negative breast cancer group (particularly aggressive and invasive)[103] , and differences between the nuclear and the cytosolic form of the protein, defining the latter as a marker for poorer prognosis[104].

- FABP8: also known as peripheral myelin protein 2, mP2 and M-FABP, it's mainly expressed in the peripheral nervous system, where it regulates the FAs' content in the myelinic membrane and the Schwann cells [40]. To the author's knowledge, there have been no specific studies able to highlight any role of FABP8 in any form of cancer disease.

- FABP9: expressed in testicular germ cells, is involved in spermatogenesis and prevents the oxidation of FAs in sperm[105]). Its absence causes malformations in the sperm head, without, however, affecting its fertility [106]). FABP9 was highly expressed in highly malignant PCa cell lines PC-3 and PC3-M, but exhibited undetectable levels in benign PNT-2 and other malignant cell lines. FABP9 expression was compared with patients' median survival time, uncovering a direct correlation between a weak expression (60 months survival), a moderate expression (24 months survival) and a strong expression (18 months survival), suggesting the use of FABP9 as a biomarker for PCa

cancer malignancy and patient outcome[107]. Further studies are needed to better discern the role of FABP9 in promoting cancer development.

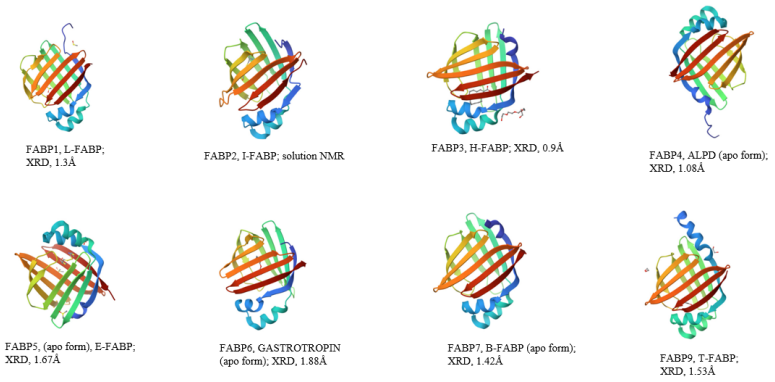

Figure 2.2: Established FABP structures, with associated experimental methodology and, where possible, resolution. Obtained from RCSB Protein Data Bank[108]

2.3 Fatty Acid Binding Protein 12

2.3.1 Discovery

Chromosomal mapping of FABP family members has shown both dispersion and synteny. FABP1-3, 6 and 7 all exist on separate chromosomes in the human genome (FABP1: 2p11; FABP2: 4q28-q31; FABP3: 1p33-p31; FABP6: 5q23-q35; FABP7: 6q22-q23), whereas FABP4, 5, 8 and 9 all co-localise at chromosome (Chr) 8q21 (respectively FABP4: 8q21; FABP5: 8q21.13; FABP8: 8q21.3-q22.1; FABP9: 8q21.13). A deeper analysis within the chromosome 8q211 revealed clustering within a 300,000-base pair (bp) region and an additional gene homologous to FABP8 was identified via prediction programs[109]. The designated FABP12 human gene corresponds to gene name RGD1565000 in rat and gene name 1700008G05Rik in mouse. In the study of R.-Z. Liu et al.[109], the predicted amino acid sequences of rat and mouse FABP12 are 92% identical to each other and $\approx$ 80% identical to that of human FABP12. Identity with other members of the FABP family is significantly lower, ranging from a minimum of 26.8% of similarity with FABP1 to a maximum of 65.91% of similarity with FABP8(Table2.1).

%IDENTITY RELATIVE TO FABP12	
FABP1	32.56
FABP2	27.13
FABP3	51.91
FABP4	55.73
FABP5	51.15
FABP6	25.89
FABP7	48.85
FABP8	65.91
FABP9	57.58

Table 2.1: Identity percentages of the other FABPs realtive to FABP12, obtained via UniProt BLAST(Basic Local Alignment Search)[110]

Being located on the same chromosome, FABP12 gene is closely linked to the other four FABPs (FABP4, FABP5, FABP8, FABP9). However, while it's contiguous with the FABP4 gene in mice and rats, it's separated from FABP4 by a single pseudogene, FTHL11 in the human genome. TBLASTN searches performed with the FABP12 rat genome sequence as query showed the presence of orthologs in other mammals, such as Canis familiaris (Gene name LOC487017, amino acid sequence identity 83%), Equus caballus (LOC100057343, 81%), Pan troglodytes (LOC748819, 80%) and Macaca mulatta (LOC706134, 78%), but also its absence in non-mammalian species[109]. mRNA expression analysis in adult rat and mouse tissues by semi-quantitative RT-PCR suggested FABP12 association with differentiated tissue functions.

2.3.2 Significance of FABP12 in PCa

The genomic region codifying FABP12 has been found to be frequently amplified in PCa, particularly in PC3 cell lines, along with other genes such as cancer Protein D52 (TPD52, 8q21.13)[111], Elongin C (ELOC, 8q21.11)[112], ZBTB10 (8q21.13)[113]. Overexpression of Elongin C found in PC3 is especially interesting because of its gene location in 8q21. The gain of 8q is one of the most common chromosomal alterations in late-stage prostate cancer, found in about 70% to 90% of MCRPC, and it seems to be associated with poor prognosis[114][115][116][117]. In another study by R. Liu et al.[118], gene amplification status of FABP12 was analysed in nine human PCa patient cohorts from cBioportal, showing that patients with metastatic PCa exhibited frequencies of FABP12 amplification ranging from 17% to 30%, while patients with localized PCa had much lower frequencies of FABP12 amplification (5–8%). Comparative analysis of the frequencies of cancers

with FABP12 diploid ($CN = 0$), copy gain ($CN = 1$), and amplification ($CN \geq 2$) between a primary PCa (MSKCC) and a metastatic PCa (SU2C) cohort, showed significantly higher frequencies of FABP12 copy gain (54% vs 15%) and amplification (17% vs 5%) in metastatic PCa compared to primary PCa, but lower frequencies of FABP12 diploid cancers (29% vs 80%)[118]. Patients with FAB12 amplification had a significantly lower survival rate [111][113] and high levels of FABP12 were associated with higher Gleason scores and metastatic status[118], suggesting that amplification of FABP12 may be a marker for poor prognosis in PCa. Finally, they also found out that amplified expression of FABP12 causes morphological changes in PC3 cells, favouring EMT[119] [120], and that, upon PC3 cell motility examination, inhibition of FABP12 with general FABP inhibitor CAS300657-03-8 (T4B) yielded major reduction in cell motility in comparison to control groups of PC3 cells expressing FABP4 or FABP5. This suggests that FABP12 is the main cause of cell motility in these types of PCa models. FABP12 is also upstream of activation of PPARγ, a nuclear receptor implicated in fatty acid signalling and lipid metabolism, has been reported to be a driver of metastasis in PCa[121]. Analysis of a PCa dataset from cBioportal[122] showed that patients with high levels of PPARγ had significantly worse prognosis, with even worse results in the subset that also exhibited high levels of FABP12 [118]. The same study also highlighted FABP12 dependency on PPARγ for increased energy production in cancer cells, with the FABP12-PPARγ pathway as a driver of fatty acid β-oxidation in mitochondria[118].

2.3.3 Researching FABP12 tertiary structure via in silico predictions

Upregulation of FABPs in cancers, inflammatory, metabolic, cardiovascular, neurological, and psychiatric diseases prompted research into employing FABP inhibition as a therapeutic tool[123][124]. The research for inhibitors has, however, to pass through the process of establishing the tridimensional protein structure, with the major problem that even a small protein domain can assume an enormous amount of 3D conformations due to the number of degrees of freedom of the aminoacidic sequence and of its side chains[125]. Traditionally, protein structure is determined with laborious experimental methods, such as X-ray crystallography, nuclear magnetic resonance (NMR), and cryogenic electron microscopy (Cryo-EM), which require purified and homogeneous materials obtained through expensive and time-consuming biochemical explorations, such as optimisation of expression constructs, isotope labelling strategies, buffer conditions, temperature and pH[126]. These issues raised the importance of obtaining in silico of protein structures with at least the same quality as experimentally identified equivalents. The fact we already possess a considerable dataset of FABP structures with a good to high degree of similarity, naturally suggests the use of homology modelling, a technique

that can predict the structure of a protein from a known sequence and experimental structure with a certain degree of homology (>30%)[126]. While homology modelling is certainly a more established technique, which has already yielded successful studies[127][128] [129], it has been gradually losing ground to more recent algorithms and to artificial-intelligence-based methodologies, such as Alphafold[130]. This study employs homology modelling, performed with Molecular Operating Environment (MOE), and AI powered prediction software Alphafold to establish the best tool for protein structure prediction, with the ultimate goal of obtaining a FABP12 structure that could be used to select, or even develop, highly specific FABP12 inhibitors to halt, or at least slow down, the metastatic degeneration of PCa. Homology modelling heavily leverages information provided by the FABP structures already published on RCSB Protein Data Bank [108]and observes sequential patterns within existing structure templates given in input, along with the query sequence, in this case the FABP12 sequence published on UniProt[110], to generate a new structure. AlphaFold does not necessarily rely on any input information besides the query sequence due to the high-resolution crystallographic structures used to train its deep neural network architecture, but several tests performed in this study showed that AF yields better results on predicting not yet established protein structures when given in input a template constituted by a protein structure whose sequence has around 60% degree of similarity to the query sequence. In both cases however, the new structures must be further processed and stabilised with Molecular Dynamics-based processes, before they can be used for any other analysis.

Chapter 3

In Silico Studies

3.1 Materials

Computer-Aided Drug Design (CADD) was introduced at the end of the 20th century and revolutionized drug discovery studies, which often needed decades of trial and error with occasional fortunate discoveries that sped them up, and substantial resources without a guaranteed outcome [131]. CADD is based on the coupling of computer algorithms and chemical and biological databases aimed at simulating and predicting interactions between molecules of interest, usually a protein or DNA sequence in the biological system. CADD development is heavily due to advancements in structural biology, unveiling tridimensional architectures of biomolecules and the growth of computing power, which enormously shortened the necessary timeframe for complex calculations. CADD can fundamentally be divided in two categories: structure-based drug design (SBDD), which relies on knowledge of the three-dimensional structure of the biological target and aims to understand what interactions potential drugs might establish with it; ligand-based drug design works in the exact opposite way, starting from the knowledge of a drug's molecular structure to find new potential candidates. However potent these in silico models may be, they still need crucial experimental on external environmental factors or unexpected biological responses to perform accurate predictions and to validate them[132][133]. This study's aim falls into the scope of the SBDD approach: using the available knowledge on the other FABPs structures, a new structure for FABP12 is predicted, in order to explore the literature and existing open databases, such as ZINC[134], and to find a possible inhibitor, with high specificity towards FABP12. characteristics[135]

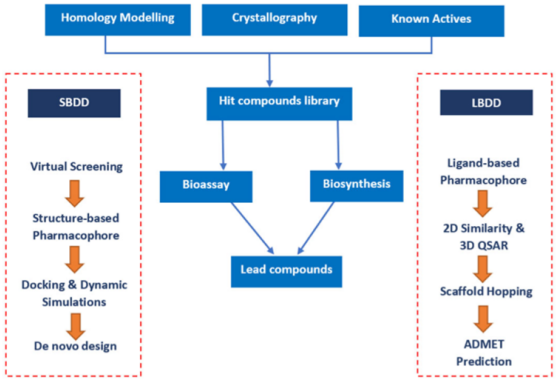

Figure 3.1: Different methodologies employed in Computer Aided Drug Design: while SBDD targets biomolecule structures, LBDD relies on known ligand

3.1.1 The Molecular Operating Environment (MOE) software

Realized by the Chemical Computing Group (CCG)[136], MOE is an is interactive, windows-based chemical computing and molecular modeling tool, and provides a complete tridimensional visualization interface, which opens on startup and allows views of a molecular system from any desired angle. Other secondary windows that can help navigate MOE's applications are the Sequence editor (SEQ), the starting point for MOE's protein modeling and protein structure analysis commands and able to organize and manage residues, chains, and tag groups; the System Manager, a window providing information and control over the systems currently loaded in MOE, presenting a multifunctional list arranged in a hierarchy: the system is a sequence of chains, each of which is a sequence of residues, each of which is a sequence of atoms; the molecular properties of the objects loaded in MOE are displayed in the Atom Manager; finally in the Database Viewer(DBV) it's possible to visualize, create and modify large collections of molecules or the many conformations of a single molecule. Various operations conducted in MOE, such as Homology Modeling or Docking, will generate, as output, a database, which is accessible in the DBV. The Database Viewer must be used carefully, since it provides a direct interface to the data on disk. Any change made in this interface immediately updates the disk file, with backups or undo. Previous data can be irrevocably lost, so it's better to have a copy of any database file that one might need to modify in the Database viewer.

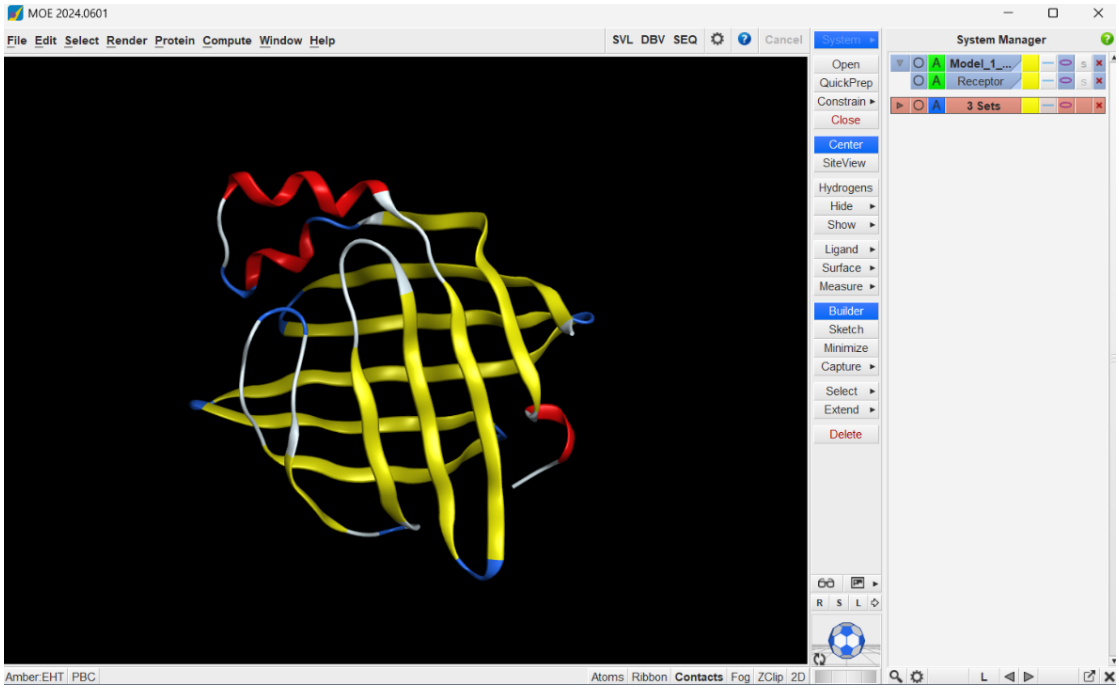

Figure 3.2: View of MOE's Command window, with loaded structure of HM generated FABP12. In the system manager, on the right, it's possible to see all the entities currently loaded into MOE and a quick menu to access various MOE tools, like QuickPrep or Siteview. At the top of the command window there are access buttons to access SVL, SEQ and DBV windows

View of MOE's Command window, with loaded structure of HM-generated FABP12. In the system manager, on the right, it's possible to see all the entities currently loaded into MOE and a quick menu to access various MOE tools, like QuickPrep or Siteview. At the top of the command window, there are access buttons to access SVL, SEQ, and DBV windows The MOE interface can also be command-line based or operated in a hybrid mode, using both the menu and the command line. The Scientific Vector Language (SVL) command window displays the output generated by SVL commands and MOE applications. Furthermore, when MOE web access is required, MOE can also be run as a batch process using MOE/web, which gives access to the MOE/batch application, providing all of the features of MOE, allowing multi-tasking and workflow automation through custom SVL scripts, which can be generated as part of batch files from the parameters provided in the graphical user interface (GUI). This version lacks graphical interfaces. Finally, MOE/cluster (MOE/SMP; Scalable Multi-Processor) enhances the MOE/batch mode and MOE in general, allowing a single MOE session to take advantage of the computational power of multiple devices, including laptops, workstations,

and multi-processor supercomputers, each running a different operating system. Applications with heavy computational costs that can take advantage of this are:

- Conformation import, used to generate the conformation of small molecules
- Database partial charges, used to compute partial charges for a molecular database's entries
- Docking, used to simulate and investigate ligand-receptor interactions
- Flexible alignment of small molecules
- Homology modeling, which is the application of interest for this study
- Model-Evaluation of pharmacophore queries, QSAR models, fingerprint models, and composite models for entries of a molecular database
- Pharmacophore search, conducted by applying a pharmacophore query to molecular conformations
- Potential energy evaluation
- Prediction of residue mutation and property to design a protein structure
- Sampling data to calculate proteins' average properties

3.1.2 Preparing the data

Unprepared files, such as PDB files, may contain missing atoms, alternate geometry or other crystallographic artifacts. These files must therefore be prepared for further computational analysis. Most files containing protein structural data come from X-ray crystallography and NMR files, while representing the minority, tend to contain fewer data-related issues, such as missing atoms and residues. X-ray files present two fundamental issues: the first is missing or unresolved atomic data, which may lead to multiple models, atoms being assigned "partial occupancies" depending on the occurrence of their appearance in different locations, or even total absence of data, causing consequences from occasional absent atoms to entire segments of missing proteins, which have to be modelled before any other process can take place. The second issue resides in the necessity to infer bonding patterns of non-aminoacidic instances from heavy atoms. While bond orders can theoretically be determined, with sufficiently accurate atomic coordinates, from interatomic distances and angles and from knowledge of hydrogen atoms, the flexibility of certain regions of the protein structure, such as loops, and the high mobility of hydrogen atoms, lead to too many positional ambiguities for the process to be reliable. The main goal of MOE's structure preparation application is to avoid manual processing

of biomolecular structures, which is prone to human error, and automate the process.

It is important to know which structures, though flawed, can be used for modeling, and which are bound to present issues that can't be resolved and will affect the quality of the model. Some parameters, at least for structures obtained through X-ray crystallography, allow to discern which structures can be selected:

- Resolution: 2 Å or finer (smaller) is suggestive of a reliable model.
- R, R_{free}: smaller value indicates a better agreement. An R value of 0.5 or greater indicates a poor model.
- Comparison of electron density: regions with low to no electron density close to a region of interest may be indicative of a poor model. Carried out by loading the electron density data into MOE and creating an Electron Density Surface.
- Protein Geometry: The quality of a protein structure with respect to known geometric constraints, such as bond lengths, bond angles, and torsion angle, should also be checked.

The Structure Preparation application allows to check any problems that a structure may have and to categorize them into one of three classes: Issues, Protonation and Warnings. Issues is by far the largest class and comprises Alternates, to which belong atoms with alternate locations and/or ambiguous sequence identities(in this case, the default correction is keeping the atom with the highest occupancy alternate and the deletion of the rest); Termini, which further splits into three subcategories:

- Nucleotide, identifying uncapped breaks in a DNA or RNA chain where there are terminal phosphate groups without an OHE cap(residues without phosphate are not flagged)-the solution consists in temporarily adding an oxygen to the phosphorous to complete the terminal phosphate group, and the oxygen is re-added to the OHE residue.
- Termini identifies uncharged/uncapped protein chain C- or N-termini and empty residues preceding a capped N-terminus/following a capped C-terminus. To correct this issue uncharged/uncapped termini are charged/capped with ACE(acetyl residue) group in case of a N-terminus or with NME(formyl residue) in case of a C-terminus, empty residues are deleted.
- Break identifies the last residue with atoms before a chain break: if the number of consecutive empty residues is between 0 and 7 the system will build a loop otherwise, an ACE or NME residue will be added as a cap, and the empty residues are deleted.

- HCount identifies the correct number of hydrogens on each atom, prompting a deletion ora an addition.

Library is also split into four subcategories: Selenium identifies residues containing a selenium atom that would match an entry in the library if the Se was mutated to S, which is also the necessary correction; Metal, which identifies residues where a metal atom has bonds to protein, DNA or RNA residues-the correction consists in deletion of metal atoms and the reset of the charges of the atom the metal was connected to; Library, which identifies structural differences between the residues in MOE and a library defined in the file $MOE/lib/amber10.mdb; Topology identifies topological problems in a residue, which are corrected according to the aforementioned library. Other categories of the issue class are S-S, which identifies missing or geometrically incorrect disulfide bond and corrects them; Aname(Atom name) which identifies atoms with denomination disagreements that are corrected according to library; UID, which identifies the presence of non-unique residue UIDs(in which case UIDs are reassigned in a chain and whether the whole chain needs to be renumbered; lastly, the Charge category identifies and corrects molecules in the system with incorrect or no charges. The Protonation class highlights issues in residues with possible rotamers, protomers, or tautomeric states, identified with the Protonate3D function in the structure panel, which also provides an almost automatic fix to these issues if provided with inputs on pH, temperature, and salt concentration (mol/L) in the solvent considered for the model. Items in this class also identifies Ligands and enables the user to select possible protomers and tautomers. However, no automatic selection for the best protomers/tautomers is provided. Items in the Warning class cannot be fixed automatically or manipulated through any function in the Structure panel. This includes Clashes (atoms within 65% of the sum of their covalent radii), residues with Cis-amide bonds, and residues with missing forcefields parameters.

3.1.3 A simpler way: the Quickprep panel

The QuickPrep function automates the structure preparation process: it deletes distant solvent, adds hydrogens, installs tethers, calculates charges and performs a first round of refinement. The QuickPrep panel is comprised of a series of items that can be enabled to access different functions. The Prepare item, if enabled, accesses the Structure Preparation function and runs it with default settings. This means that in case of missing atoms (breaks of up to 10 residues and terminal outgaps of up to 5 residues), small loops are built. In case of larger gaps, the system either neutralizes endpoints adjoining empty residues, preserving the empty residues, or the adjoining endpoints are capped and the empty residues are deleted.

The Delete item enables deletion of water molecules farther than 4.5 Å from any receptor or ligand atom, reducing the computational weight that would come from keeping them. The Tether item allows the system to place restraints on the selected entity (ligand, receptor, or solvent) and/or its corresponding hydrogens so that they do not deviate too much from the experimental coordinates during the simulations. The strength and buffer parameters in the Tether item control the allowed deviation distance (strength) and the extent of the flat-bottom restraint (buffer). The Fix item does the exact opposite: if enabled, atoms that are farther from the binding site than the specified distance, will be constrained to be fixed in place. The binding site can be defined with the pulldown menu. The default option uses all ligands. Alternative options consist in using the dummy atoms (which are, however, excluded from the refinement process) obtained with the Site Finder function, or using a custom set of atoms with the selected option. The last item, Refinement, performs energy minimizations using the current parameters in the Potential Setup, which can be set to default unless the simulations call for custom settings, which is not the case of this study. The refinement process ends when the RMSD gradient falls below the value specified by the user. It is possible to keep the restraints selected in the QuickPrep panel by enabling the associated sub-item in the refinement section. This avoids tension and clashes in the conformational space during the refinement process. Enabling the Save Item, the current selction is proposed as the default option for the next QuickPrep, which is useful in case the user needs to run a series of simulations.

3.2 Homology Modeling

Homology Modeling (HM) is a computational structure prediction method based on the biological premise that proteins with high degree of identity between their aminoacidic sequence also possess similar tridimensional structure [137]. The homology modeling procedure normally starts with the use of the query sequence to search for possible template structures in databases such as RCSB PDB Protein Data Bank (RCSB.org)[108] or UniProt[110]. In this case however, these databases were used to recover the data on the already established FABP protein structures. The first and most important portion of data is the aminoacidic sequence of the desired proteins, which can be retrieved in the form of a FASTA file, containing a string of letters representative of each residue and the name, organism of origin, and classification of the protein. The easiest way to obtain FASTA files is through one of UniProt datsets, the most important of which is the Universal Protein Knowledge Base (UniprotKB). It is comprised of both a reviewed set of protein entries (UniProtKB/Swiss-Prot), where each record contains a summary of the experimentally verified, or computationally predicted, functional

information, added and evaluated by an expert biocurator, and the unreviewed set (UniProtKB/TrEMBL). UniProtKB includes annotations to 12,501 Rhea reactions, which are linked to 28,259 857 UniProtKB protein sequence records, including 231,709 reviewed protein sequence records in UniProtKB/Swiss-Prot[110]. After retrieving the aminoacidic sequence from UniProtKB, it was used to search for structures in RCSB PDB. This process lead, for each FABP to the selection of a structure based on two criteria: highest degree of similarity between the sequence published on Uniprot and the sequence published on rcsPDB; highest structure resolution, as reported in Table3.1:

Target	Template	ID%
FABP1	FABP6	36.72
FABP2	FABP7	36.15
FABP3	FABP7	67.18
FABP4	FABP8	67.18
FABP5	FABP8	58.73
FABP6	FABP1	36.72
FABP7	FABP3	67.18
FABP8	FABP9	67.42
FABP9	FABP8	67.42
FABP12	FABP8	65.91

Table 3.1: Identity percentages of each FABP and of the chosen template form Homology Modeling, selected for the highest %ID. Data obtained from UniProt's bLAST[110]

The Homology Modeling module in MOE is articulated in the following steps:

1. Initial Partial Geometry Specification: the template chains (one or more) are used to obtain an initial partial geometry for each target sequence. In regions with conserved identity with the Target and template sequence, all heavy atom coordinates are copied; otherwise, only backbone coordinates are copied. Copying atoms from template includes all the attributes that affect energy minimization, as well as Fixed status, tethers, and restraints. In case a certain residue on the target sequence is aligned with a different residue on the template and the φ angle is positive, no coordinates are copied, and the position is marked as INDEL. When half cystines in the template are both aligned with cystines on the target sequence, a disulfide bond will be copied in the model. Disulfide bonds can, however, also be created in the side chain packing procedure.

2. Insertion and Deletion: after the previous step some residues have yet to be assigned backbone coordinates. These residues may be in loops (insertions in the model with respect to the template), "outgaps" (residues in a model sequence which are aligned before the C-terminus or after the N-terminus of its template), or in deletions (regions where the template has an insertion with respect to the model). In the latter case the copied geometry may be ignored, since it is not likely to be correct on either side of the deletion. This "creates" an insertion, and these cases are referred to as "INDEL". In INDELS backbone geometry gets modeled from high-resolution fragments from PDB, selected depending on the quality of their superposition onto anchor residues on either side of the insertion area. In case no segment is found, the search can back off from indel regions.

3. Loop Selection and Side Chain Packing: creation of a set of independent models. First to be modeled, in random order, are loops. The candidates' data collected during the segment searching stage is analysed with a contact energy function, considering all atoms already modeled and any atoms specified by the user as belonging to the model environment (e.g., a ligand bound to the template, or structural waters). Candidates are then chosen through a Boltzmann-weighted selection process realized with the energies obtained with the contact energy function analysis. The coordinates of the chose candidates are copied in the model. After the loop selection, the sidechains are assembled using data from an extensive rotamer library generated by systematic clustering of conformations from LowModeMD. The sidechains' optimal packing is then selected through a Unary Quadratic Optimization-based process. After backbone and sidechains selection, valence requirements are fulfilled with the addition of hydrogens, and the model goes through minimization to relieve stearic stress on to prepare for scoring. Is then loaded in the output database with a series of descriptive values that may flag geometric issues.

4. The intermediate models (10 by default, this setting has not been changed in this study since it did not produce any significative changes in the results) are scored and ranked. The best emerges as the final model. There are multiple metrics by which the model can be selected:

    - I. Root Mean Square Deviation (a measure of the distance between respective residues of a protein structure when superposed with another protein) of each intermediate model with the average positions of all intermediate models (either all-heavy atom, or Cα-only)
    - II. The electrostatic solvation energy, calculated used a Generalized Born/Volume Integral (GB/VI)

- III. A knowledge-based residue packing quality function obtained from statistical analysis of the usual aminoacidic environment. Residues are classified depending on solvent accessibility, their participation in hydrogen-bonded contacts, and the ratio of polar vs non-polar contacting atoms in its micro-environment. Then, using data from the distributions of each residue type in the different environments, histograms are constructed, allowing to assess the probability for an average protein to exhibit a specific sequence of residues in a definite set of environments. The reported value is the negative Log of this probability, meaning that lower values will indicate higher probability structure-sequence combinations and vice versa. High-resolution structures typically fall in the 2-3 range, while values over 5 indicate unusual residue environments.
- IV. Estimation of the effective atomic contact energy (ACE) [138] [139]: in which the effective energy difference for replacing residue-solvent contacts with residue-residue contacts is based on observed contact frequencies in a database of high-resolution protein structures.

The final model must be inspected and validated using MOE's Protein Geometry tools, in order to evaluate if the model is consistent with average crystal structure values. Possible issues may be due to errors in the alignment used for model building, in which case the alignment might have to be manually adjusted and the model rebuilt.

3.3 Methods

3.3.1 Homology Modeling Workflow

To model each structure, the FASTA file containing the sequence was obtained from UNIPROT [110]t and loaded into MOE. Then the template for Homology Modeling is chosen as the structure with the highest identity percentage according to the results in the Identity Matrix obtained with UNIPROT's BLAST tool, as reported in Table6.1 in the Supplementary Material chapter. The selected structure is loaded into MOE along with its associated.PDB file, taking advantage of MOE's access to PDB and using the PDBId of the desired structure in the search bar. The structure is then corrected using the QuickPrep application with the default settings. The sequences are aligned using the Align/superpose tool in the SEQ and selecting the option "Sequence Only" in the footer, since the default options requires alignment of two sequences and their respective structure. The alignment will produce a new similarity matrix in the SVL command window, showing the same results obtained with BLAST. Without leaving the SEQ panel and choosing the option Protein, a pulldown menu presents, among its option, the Homology Modeling one. Once

selected, it opens a panel where all the modeling parameters for the simulation can be decided. In the Sequence field all the protein chains currently loaded into MOE are presented as options from which the user can choose the sequence that will be modelled. Protein chains that are constituted entirely of empty residues with no atoms are automatically recognized by the system and displayed in this field. In the Template field are displayed all the protein chains currently loaded into MOE to which a structure, displayed in MOE's main window, is associated. These protein chains consist entirely of standard amino acid residues and possess at least three residues with complete main chain coordinates. Once again, the sequence identity of the template with respect to the query sequence, obtained with the SEQ current alignment, is displayed beside the Template field. All the remaining fields, such as the optional selection of atoms that may be used in clash testing when building loops or and/or sidechains, or the optional selection of residues whose geometry may be used in modeling to override the template, are left as default, with no selection of atoms or residues. The C-terminal and N-terminal outgap modeling item was also left not-enabled. This Choice has been made because these features of the Homology Modeling function has been deemed unnecessary for the purposes of this study. Lastly, the scoring metric for the intermediates can be selected, with the default option of the second previously described scoring metric, the Generalized Born/Volume Integral (GB/VI). In this study it was not deemed necessary to keep the intermediates.

3.3.2 Outputs

After having selected all the parameters, pressing OK launches the Homology Modeling application. The Output is a Database which can be viewed in the DBV window, containing both the model and the template, and a series of descriptive values, reported in Table3.2. This process is repeated for each of the FABPs, and each resulting model is once again prepared with QuickPrep for validation against its experimental counterpart, performed using the Align/superpose tool in the SEQ, and calculating the RMSD to assess the quality of the prediction. In the case of FABP12, since there is no experimental structure to validate the prediction, the comparison has been performed with the FABP8, 4BVM protein experimental structure, the template used for its modeling. This result has later been compared to the RMSD calculated by loading into MOE a FABP12 structure obtained through Alphafold prediction and its template, to assess which methodology yielded the best results.

RMSD to Mean	The heavy atom RMSD of each model to the average position of all of the intermediate models.
CA RMSD to Mean	The Cα-only RMSD of each model to the average Cα position of all of the intermediate models.
Packing Score	A statistically based residue packing quality function, expressed on a negative log scale.
GB/VI	Coulomb and Generalized Born [140] interaction energies of the model (including environment atoms).
Contact Energy	Effective atomic contact energies (ACE), based on [139].
U	The total potential energy of the model (including environment atoms).
E sol	The solvation energy of the model (including environment atoms).
E ele	The electrostatic energy of the model (including environment atoms).
E vdw	The van der Waals energy of the model (including environment atoms).
E bond	The bonded energy of the model.
Atom Clashes	A count of the atom-atom contact pairs with repulsion energies exceeding 0.5 kcal/mol.
BB Bond Outliers	A count of the number of backbone bond length outliers (Z-score greater than five) in the homology model.
BB Angle Outliers	A count of the number of backbone bond angle outliers (Z-score greater than five) in the homology model.
BB Torsion Outliers	A count of the number backbone torsion outliers (Z-score greater than five) in the homology model.

Table 3.2: Database of prediction scores produced as output of Homology modeling, along with the .mdb file of the predicted structure

Chapter 4

Alphafold

4.1 Materials

Alphafold2 (AF)is the second computational prediction method employed in this study. The Alphafold system's first version debuted in 2018 at the Critical Assessment of Protein Structure Prediction (CASP13)[141], a biennial competition, run by the structure prediction community, where novel and/or updated computational models and techniques are assessed through blind protein structure prediction, meaning competitors have to predict recently solved structures that have not been deposited in the PDB or publicly disclosed, to benchmark progress in accuracy[142]. AF scored well in all three categories: the Free Modeling (FM), the Template-Based Modeling (TBM), and the hybrid FM/TBM category. Developed by Senior et al. (2020), it consists of a neural network trained to accurately predict distances between pair of residues. The obtained information is then used to construct a potential of mean force[143] to describe the protein's shape. The resulting potential can be optimized through a simple gradient descent algorithm to generate structures rather than through complex sampling procedures. Thus, AF already distanced itself from previous Free Modeling (in which prediction and modeling of a structure are performed without the availability of a homologous structure, with only the aminoacidic sequence) approaches, which relied on fragment assembly, where a structure is created through a stochastic sampling process, e.g. simulated annealing[144][145][146][147]. In the fragment assembly approach, the structure goers through numerous iterations of modifications, with changes being apported to the shape of a short section while retaining changes that lower the potential, minimizing the overall potential and finally obtaining a structure, requiring, however, thousands of such moves to achieve the task. It is with the system's second version, Alphafold2 [148]that prediction accuracy at the atomic level of protein structures, even with no homologous structure available, has been achieved, producing, at

the CASP14 [142], predictions that approach a score of 90 in the global distance test (GDT TS), where a score above 90 is considered roughly equivalent to the accuracy of an experimentally predicted structure. AF greatly outperformed its competitors and earned the 2024 Nobel Prize for chemistry. AF structures had a median backbone accuracy of 0.96 Å r.m.s.d.$_{95}$ (Cα root-mean-square deviation at 95% residue coverage) (95% confidence interval = 0.85–1.16 Å) whereas the next best performing method had a median backbone accuracy of 2.8 Å r.m.s.d.$_{95}$ (95% confidence interval = 2.7–4.0 Å) as measured on CASP domains. For reference, a carbon atom is approximately 1.4 Å wide. The latest iteration of AF, AlphaFold3[149], further improves on the results yielded by AF2 and is capable of high-accuracy prediction of complexes containing nearly all molecular types present in the Protein Data Bank(Berman, 2000). The overall architecture of AF3 is similar to that of AF2, with differences in the main building block of the neural network, Evoformer, where a simpler pairformer module, replaced the evoformer module that gave the block its name. The change streamlined the building process and permitted to accommodate more general chemical structures and to improve the data efficiency of learning.

4.1.1 Accessing Alphafold

AF is a free, open-source AI system, accessible to users in four ways: using the AF source code, which is accessible on Github [150]; accessing through ColabFold [151]; accessing the AlphaFold Protein Structure Database, which has been developed by Google DeepMind in collaboration with EMBL-EBI (European Molecular Biology Laboratory-European Bioinformatics Institute) and contains over 200 million predicted protein structures; last option consists in using a custom software incorporating AF's algorithm and/or data.

4.1.2 Alphafold Database

Every predicted structure has its own page, along with related information and links, including one to the protein's Uniprot page, with access to AF prediction scores for fast assessment of the prediction quality of the structure. The downloadable data is available to users on the form of mmCIF/.pdb files (atomic coordinates) or JSON files (prediction scores). The database can be accessed in different ways depending on the user-case scenario. For occasional access, the AlphaFold Server website interface is sufficient, easy to use and without any coding experience requirements; for downloading large datasets (e.g. proteomes), initiating a File Transfer Protocol (FTP) may be the best option, due to scalability. The Application Programming Interface (API) provides a customisable way to access metadata on AF predicted structures, further improving flexibility and scalability, since it's tasks, including

searching for protein structures or creating advanced queries, fitted to the user personal criteria, can be integrated into other applications. It requires, however, a certain degree of programming literacy and understanding of PAI documentation, making it a more indicated choice for developers, rather than the occasional user. Also, queries may take a long time depending on the amount of requested data. A similar way to access is through Google Big Query: the AF dataset is hosted on Google Cloud Public Datasets and can be downloaded for free under a CC-BY-4.0 license, though Big Query access requires some knowledge of SQL queries. While it's true that BigQuery has a free tier, it offers a limited number of queries per month, leaving it to the user to strategically allocate their resources depending on their necessities. While the AFDB is certainly vast, it may not always constitute the best option to obtain the desired protein data. For example the protein of interest may be outside of the range of lengths included in the database (2,700 for proteomes and Swiss-Prot's reviewed entries and 1,280 for the rest of UniProt; longer proteins, only for human proteome, can be downloaded segmented in fragments, only through FTP); AFDB only contains oligomers(no viral proteins) and only one predicted conformation per protein, so any different typology/conformation of protein of interest will need to be modelled by the user.

4.1.3 Colabfold-based Systems

ColabFold and AlphaFold3 Colab are both online implementations of AF3 that don't require installations on local devices to run any of their operations. AF3Colab is a marginally simplified version of AF3, which makes use a restricted dataset and does not employ homologous protein structures as templates, and does not support any customisation through variation of prediction parameters performed by the user. As a result, its accuracy is lower than normal AF3. ColabFold is a community implementation of AF3 and offers more customisation possibilities, for example, allowing the user to decide the size of the Multiple sequence Alignment (MSA) employed in the predictions, the number of recycles, and even to provide custom MSA and templates [151] as input. Both applications provide a simple interface, similar to a form. Runtime depends on the length of the desired protein and on the type of GPU assigned to you by Colab. In case the user possesses a highly powerful local device, equipped with a cutting-edge GPU, it is possible to set up a local Collab and run ColabFold locally, without having to install the full open-source AF3. The local approach, although more expensive, circumvents limitations of its cloud-based counterparts, mostly set by the quantity of RAM available on the GPU provided by Colab.

4.1.4 Alphafold Source Code

Installing the source code is the most powerful and adaptable way to operate AF3, giving the user all the functionalities previously described for the other options, total control over parameter predictions, and comprehensive outputs. It is, however, highly demanding both in terms of workstation/server hardware and of user competence. All installation instructions and code necessary for running are available in the official GitHub [150]. Running Alphafold on a local device requires a machine capable of operating Linux, 3 TB of disk space for storing the necessary genetic databases, including BFD, MGnify, PDB70, PDB, UniRef30 (FKA UniClust30), and UniRef90. Similarly, AlphaFold-Multimer requires the PDB SEQRES and UniProt databases and a state-of-the-art NVIDIA GPU. While it is possible to run AF3 without a GPU, it will be much slower. In case a similar machine is unavailable, AF3 source code version can also be run on a virtual machine in the cloud, as was decided for the purposes of this study. AF3 was installed on High-Performance Computing (HPC) clusters provided by the Digital Research Alliance of Canada, which have been accessed through a Secure Shell (SSH). From a user perspective, however, even the source code version requires minimal input, strictly correlated with the user case scenario. The only necessary input is the protein sequence, while other inputs, such as templates, MSA, ligands, or bonded atom pairs are totally up to the user's necessity.

4.2 Alphafold achitecture Overview

AF3 architecture is the result of a streamlining of the AF2, though it follows the same principle: from, the input sequence a series of feature maps(or, as they will be referred to going forward "input matrixes") are generated by AF3 to be cyclically updated in the learning process performed by the learning block, with the Pairformer module as the protagonist. The most important of these maps is the Multi-sequence Alignment and the pair representation. While the MSA has a less prominent role in AF3 compared to AF2, it's still crucial for updating the pair representation, which is the main information carrier that will pass through the diffusion module and contribute to generate the predicted structure. A brief overview of the most important input matrixes and of their updating processes is provided here. For a more in-depth explanation of AF3 inner workings, the author recommends the AF3 paper supplementary material[152], and "The Illustrated Alphafold" GitHub page [153], from which some of the explicative figures provided here have been extrapolated.

4.2.1 Input Generation

The input of the aminoacidic sequence prompts AF to build a series of matrixes to store the necessary data for the learning and prediction process. This process is performed by the Data Pipeline code section. The first input matrix to be generated is the Multiple Sequence Alignment (MSA). The MSA is a matrix where each column represents a residue of the query sequence, while the rows represent the sequences in which those residues appear. The MSA allows AF to perform a comparative analysis between different protein sequences, highlighting similarities, differences and evolutionary relations. Close contact between two residues makes it highly likely that a mutation in either of the will cause a mutation in the other, in order to preserve the protein structure, and vice versa, two regions of protein that are evolving and mutating separately are not likely to be close to each other.[154][155][156][157]. Depth (the MSA depth parameter indicates the number of sequences that make up the matrix) and diversity of the MSA are crucial to allow AF to identify co-evolutionary signals and use them to figure out the protein's 3D structure. On the other hand, a shallow MSA, with low variability among its sequences, is more likely to lead AF to yield bad predictions. The Resulting MSA is a tensor with dimensions n,N,c), where n is the number of residues, N is the MSA depth and c is the number of channels, in this case, 64.

At the same time, if any of the sequences in the MSAS also have known structures, individual chains of these structures are also employed to model the final structures, echoing the Homology Modelling building process. Each of these individual chains considered as a template and is represented as a matrix with dimensions "n,n", where n is the number of "tokens" that make up the chain (aminoacidic residues or nucleotides). Each element of the matrix is constituted by the Euclidean distance between all pairs of tokens in the chain, calculated between the representative atom of each token (the Cα atom for amino acids and C1' atom for standard nucleotides). The distances are discretized and converted into a distogram (histogram of distances). Each distogram is then masked in a way that the system can only consider distances within each distogram and not between different distograms. The distograms constitute the elements of the template matrix (Tensor 1t), with dimensions (n,n,c*number of templates), where n is the number of residues and c is the number of channels.

Along with the MSA, a set of pair representations is created, which separately represents the relationships (e.g., distance, potential interactions) between every pair of amino acids/atoms in the complex. The first step to creating the pair representation is to consider separately all the atoms in the target sequence ad their features. Then, for each atom, a reference conformational isomer, a 3D arrangement

of atoms in a molecule that is generated by sampling rotations about single bonds, is calculated by combining experimental data and torsion angle preferences. The conformer information is then associated to each atom's charge, atomic number, and other identifiers, generating an atom-level representation matrix of the target sequence (that will be referred to as Matrix 1s), with dimensions (N,c), where N is the number of atoms and c is the number of channels, in this case, 128. This matrix is used to initialize an atom-level pair representation matrix where each element stores the distance between every single atom and another (that will be referred to as Tensor 1p), with dimensions (N,N,c), where N is the number of atoms and c is the number of channels, in this case, 128. The atom-level representations are updated in the Atom Transformer Block. Since however AF3 works on token-level representation, an in-depth explanation of its inner workings fall beyond the scope of this thesis. For more information the author recommends this material [153]. The

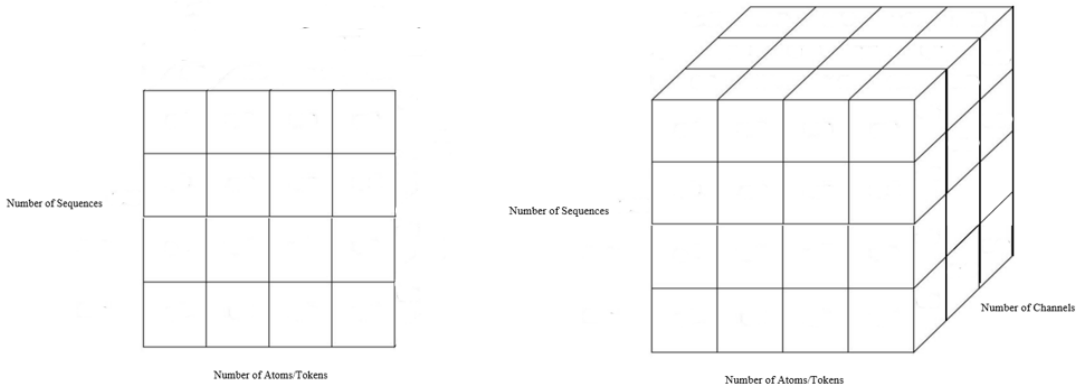

Figure 4.1: Example of the matrices (single representation) and tensors(pair representation) used to store Atom-level or Token-level data(depending on the level of representation). The first input entities created are matrix 1s and tensor 1p; The MSA and the templates data are also stored in tensors.

learning section of AF3 works on a token level (note that, being the study about proteins, the terms token/s and residue/s will be used interchangeably), meaning that it needs representation of residues or nucleotides to predict a structure, not of their atoms. The mean over all the atoms in the Matrix 1s associated to the same token is considered, leaving other atoms unchanged, and used, along with token-level features and statistics from our MSA (where available), to create a token-level matrix(that will be referred to as Matrix 2s), with dimensions (n,c) where n is the number of residues and c is the number of channels, which is 384 plus the necessary channel to accommodate results of concatenations with MSA

data. Each element of Matrix 2s represents a residue of the target sequence, which will be used during the structure prediction section. From Matrix 2s, a second matrix is obtained through linear projection (that will be referred to as Matrix 2.1s), with dimensions (n,c) where n is the number of tokens and c is the number of channels, again 384. This is the starting representation of our sequence that will be updated in the representation learning section. Matrix 2.1s is then used to initialize the token-level pair representation. First, Matrix 2.1s goes through linear projection so that the channel dimension of sequence representation match that of the pair representation Tensor 1p (384 → 128). Then, with the addition of atomic data, for each singular residue atom encoded in Matrix 2s and from the associated atom- level pair representation encoded in Tensor 1p, we obtain each element of the token-level Pair representation. The pair representation can be represented as a tridimensional matrix with dimensions (n,n,c), where n is the number of tokens, c is the number of channels, which is 128. Raw atom coordinates are predicted by a diffusion module, which takes in input coarse arrays, which depict a per-residue representation of the input sequence, of the pair representation, and of the single representation. On the other hand, fine arrays depict a per-atom representation.

4.2.2 Representation learning block

The pair (Tensor 2p) and single (Matrix 2s) representation are cyclically updated during the learning process. The pair representation goes through the template module and the MSA module, then the result of the MSA module and the single representation go through the pairformer module and the results are fed back into the loop (10 cycles, default setting).

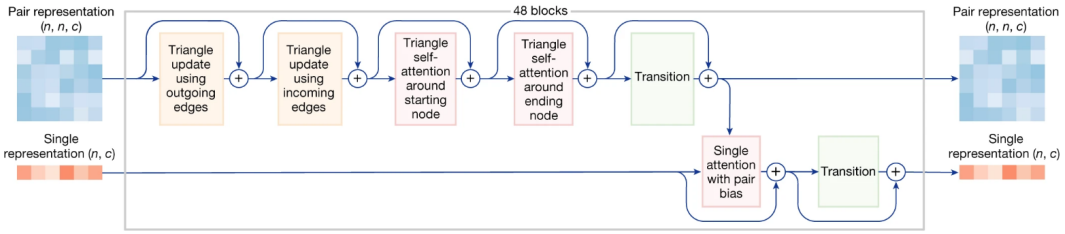

Figure 4.2: Updating cycle of the single and pair representation[152]

In the template module each template (one-element-wide slices of the Tensor 1t) goes through a linear projection and is added to a linear projection of the pair representation (Tensor 2p), generating a new template matrix, Matrix 2t, which goes through the pairformer twice and a final linear projection, finally obtaining

the Tensor 2.1t, containing the adjourned templates.

Updating Pair representation with the MSA

In the MSA module the MSA representation is incorporated in the pair representation (Tensor 2p) via Outer product mean. After subsampling the MSA (which could have reached a depth of 16k sequences, making it to computationally heavy to work with) and adding a projected version of the single atom representation (Matrix 2.1s), we obtain a new MSA, with dimensions (n,s,c), where n is the number of residues, s is the number of sequences and c is the number of channels(this matrix has two channels since is constituted by two equal "slices", the first being the subsampled first channel of the old MSA, the second being the projection of Matrix 2.1s). Two columns of the new MSA are then compared, revealing informations about the associated residues and their positional relationships in the sequence. Let it be that i and j are the indices identifying the columns of the MSA and, therefore, the residues; The s index identifies the rows of the MSA, therefore the sequences, and has a value that ranges from 0 to the MSA depth; the c index identifies the channel considered, and ranges from 0 to 1. For each sequenc,e the outer product is performed for each pair of residues, meaning that on a given s row, all elements with column index i, going from channel 0 to 1. The products are then averaged across the sequences, obtaining another outer product, between the average values in column i and j. The resulting matrix is flattened along the channel dimension, obtaining an array with dimensions (1,1, c^2). This array is linearly projected so that the number of channels is equal to the number of channels of the pair representation (128) and is then added to the pair representation (Tensor 2p). The MSA module significantly reduces the complexity of AF2 Evoformer module, allowing the sharing of information across sequences.

Updating MSA with Pair representation

This specific update pattern is called row-wise gated self-attention using only pair bias. To explain the update flow, consider the MSA tensor. Take a whole row of index s, where s is the index of the considered sequence, obtaining a matrix with dimensions (n,1,c) (basically, a horizontal slice of the MSA). From this matrix, we consider one singular element with index i, which will be represented by an array with length equal to the number of channels. Both the matrix and the array go through separate linear projections so that both entities have 2 channels. The array also goes through sigmoid. The obtained matrix and array will now be referred to as S and V. Now, we consider the pair representation, and we take the row with the same i index as the column from which we obtained V. The row is extracted and linearly projected, obtaining an array with a length equal to the number of residues (the channels are collapsed into one singular channel). This array goes

through Softmax operation, which converts it into a probability distribution: all the array's elements now have a value between 0 and 1, and their sum equals 1. This array will be referred to as P. The next step is multiplying each channel in S by P. The resulting matrix then goes through sum across residues, obtaining the array R. The array R is the n multiplied by V, obtaining the array Y. Since this process has been performed in parallel for each one of the attention heads employed in this process(Abramson et al., 2024, supplementary information), we now have sixteen Y arrays equal to the number of residues in the target sequence. The arrays are concatenated, obtaining an array with a length of c*16, which is then linearly projected to obtain an array with a length equal to the number of channels in the MSA. The new array is the updated version of the array with index i from which we started. The process is repeated for each residue in the MSA, which is represented by the array that runs across the channels.

4.2.3 Pairformer MOdule

The updated pair and single representations are now used to update each other. The representations are updated, in the pairformer module, in a similar way to how they were updated in AF2 [148]. The update patterns are based on the representation of residues on a triangular diagram, with each vertex corresponding to a node representing a residue, with edges corresponding to the pair representation of the two residues on the vertices at the edge extremities.

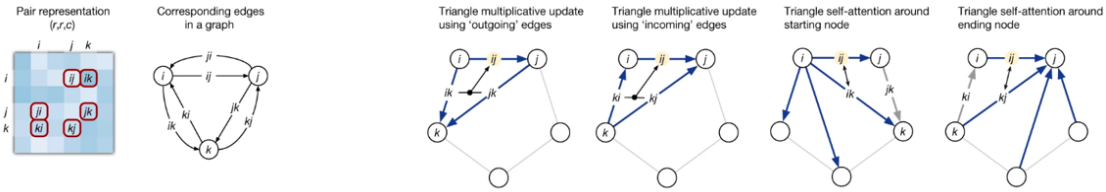

Figure 4.3: Triangle multiplicative update and triangle self-attention. Each circle represents a residue, and each directed edge is one of the elements in the pair representation, constituted by the correlated characteristics of the two residues. In the presented example, element ij is being updated. [148]

To represent geometrical constraints that allow the pair representation to describe a 3D protein structure, including the triangle inequality on distances, which states that the sum of the lengths of any two sides of a triangle must be greater than the length of the remaining side, AF3 employs first triangle multiplicative update, then triangle self-attention. To better explain the update pattern, the case in which the ij edge is being updated shall be taken into consideration. The first stage of triangle multiplicative update uses the values represented on the outgoing

edges of nodes i and j to update the ij edge(Figure6.1), while the second stage uses incoming edges to once again update the ij edge (Figure6.2). Then, the first stage of triangle self-attention focuses on the edges outgoing from the starting node, which is discernible as the node i from the first index of the ij edge (Figure6.3), and the second stage focuses on the edges incoming to the ending node (Figure6.4). Finally, the adjourned pair representation is used to update the single representation through the single attention with pair bias process. Graphical explanation of the triangle multiplicative and triangle self-attention updating patterns is provided in the supplementary material.

After being updated, both the pair and the single representation go through SwingLu transition block, which reduces complexity of the model. Given an input the system creates two linear projections (P1 and P2) so that their column dimension is quadrupled. One of these projections (P2) then goes through a Swish non-linearity, a non-monotonic activation function defined as:

$$f(x) = x \cdot \text{sigmoid}(x) \tag{4.1}$$

Then the two linear projections (P1 and now P2*) are element-wise multiplied, and the product is projected back down to the original size of the input. The pair and single representations are now ready to pass in the Structure Prediction block.

4.2.4 Structure Prediction

In AF3 the entire structure prediction is based on atom-level diffusion. Conceptually, this means iteratively adding random noise at multiple time steps to the refined representations, creating a sequence of T (number of iterations) variants of each datapoint, train the model, and predict which noise was added. During training, at each timestep (t=, the model is given the data variant generated at the previous timestep (t-1), and predicts what noise was added between t-1 and t. A gradient step is taken on the predicted noise compared to the actually added noise. Then at the inference time, the system considers the data point with the complete noise that has been added. For each time step, the system predicts the noise it “thinks” has been added at that time step, and that noise signal is removed from the datapoint. After T subtractions, we obtain a “denoised” datapoint that should resemble the input. Each iteration produces three inputs: the current noisy datapoint, the representation of the iteration, and the information to be conditioned at the given timestep.

Diffusion module

In AF3 each input that gets iteratively randomized is a matrix containing the coordinates of each atom in the structure of interest, randomly rotating and translating them, is the refined Pair and single representation(referred to as 2p* and 2.1s* from here on), and the initial atom and token level single representation that have not been updated in the pairformer (Matrix 1s and 2s). The process teaches the model that any conformation of the structure is equally valid. Before denoising is handled by the Diffusion module, a small amount of noise is added to guarantee diversity of prediction. The complete process is articulated in four steps: Preparation token-level conditioning tensors; Preparation atom-level conditioning tensors, update them using the Atom Transformer, and aggregate them back to token-level; Apply attention at the token-level, and project back to atoms; Apply attention at the atom-level to predict atom-level noise updates. To initialize the token-level positioning tensor for the pair representation, 2p* is concatenated with its relative positional encodings (relative token distances), obtaining a tensor with double the channels of 2p*(c=128). This tensor is then projected back down so that it has 128 channels and goes through several transition blocks. In the same fashion, the single-atom token representation 2.1s* is concatenated with Matrix 2s and projected back down to its former size. Then, a Fourier embedding is created, an array based on the noise at the current timestep. The Fourier embedding is added to the matrix, and the result goes through several transition blocks. To initialize atom-level conditioning tensors, the system updates both the single-atom-level and the pair-atom-level representations (so 1s and 1p) using 2.1s* and 2p* respectively. Both 2.s* and 2p* are broadcast to atom-level and linearly projected to have the same dimensions a 1s and 1p, respectively. Then this projection of 2.1s* and 2p* is added element-wise to 1s and 1p, respectively, creating the atom-level conditioning tensors 1Sx and 1Px. In the next step the system projects the matrix containing the atom coordinates(that will now be referred to as X) (with noise depending on the current time-step) so that it has the same dimensions as 1Sx, effectively creating "dimensionless" coordinates with unit variance. This new atom coordinates matrix is added element-wise to 1Sx, meaning that the system now knows the tridimensional coordinates of each atom at the current time step. This matrix is then updated through the Atom transformer and reaggregated to token-level, obtaining the matrix 1Sx*, which captures coordinates and sequence information. At the end of the process, we obtained: the updated atom representation 1Sx; the updated atom pair representation 1Px; atom representation based on pairformer results 2.1s*; the previously described 1Sx*. 1Sx* is subsequently updated through the Diffusion transformer, which mirrors the Atom transformers but acts on tokens. Briefly, 1Sx* and a copy of 1Sx go through a series of bocks, the first being the Adaptive LayerNorm block (see (Abramson et al.[152]) supplementary material).

The result then goes through a Single Attention with Pair Bias process along with a copy of 1px. The result of this operation goes through a conditioned gating based on the 1Sx matrix. This means that the 1Sx matrix goes through a Sigmoid transformation and the result is multiplied element-wise with the input of the conditioned gating. The result then goes through 3 to 4 cycles of conditioned transition based on 1Px, obtaining the adjourned matrix 1Sx**. In the last step of the process, we return to the atom-level, broadcasting 1Sx** to match 1s dimensions and, updating the 1s original matrix using the broadcasted 1Sx** and running the Atom Transformer. Now, all the conditioning tensors are used to generate the necessary coordinate updates for the atoms in the target structure. Since these updates are dimensionless, they are scaled back into their units and applied to X. We have obtained a prediction for the tridimensional coordinates of the atoms in our target structure and, therefore, a structure prediction.

4.3 Methods

4.3.1 User-level Inputs

Obtaining a prediction on a protein structure using AF3 source code involves running two scripts: the Data Pipeline script, which must receive the correct input, a properly written JSON file, and the Inference script, which receives as input the "*data.json" file produced as output by the data pipeline. Other code parameters, such as the depth of the MSA, the number of refining cycles in the Pairformer module or the number of random seeds, which give more possible starting conditions to the neural network, allowing a more diverse exploration of the solution space, can all be manually modified by the user, depending on what tests ought to be run. The input Json file is articulated in multiple sections, which determine the type of prediction that AF3 will run. The model seeds have to be passed to AF3 as an array of integers (minimum length=1). The length of the model seeds array is the number of protein structures AF3 will predict from a single JSON input file. All the predicted structures will be accessible in the output file, but AF3 also ranks the structures and highlights the best one. After several tests, ten was the number of random seeds chosen for this study, since it provided a good prediction diversity with no significant increase in computation complexity/time. The second section is constituted by a series of dictionaries, which can be omitted in case all necessary information about the target sequence and possible templates has to be indicated. While not strictly necessary, templates can provide additional information. During the testing stage multiple predictions have been run, leading to the decision to use a template to obtain the final structures. The third section, "bondedAtomPairs" allows specifying, with a list, which atoms are to be considered covalently bound during the whole process. The, in the fourth section, "userCCD",

it's possible to either provide a string with user-provided chemical components dictionary. This mode not only allows the use of custom molecules when SMILES is not sufficient, but also binding a custom molecule with other entities; alternatively, with "userCCDpath", it's possible to write the path to the file, which must be in CCD mmCIF format, containing the string. These two sections, however, are an advanced setting that is redundant for does not concern this study, so the "bondedAtomPairs" list can be left empty and "userCCD" can be set to "null". Finally, in the dialect and version section it has to be indicated for what version of AF3 the JSON is intended for "alphafold3", which is the case for this study, or "alphafoldserver", for which the content of the JSON file is fundamentally different and not described heare. In the last section the version of the JSON input file is defined. While simple, this section is important since allows the option of specifying external MSA and templates using the fields described below. Version 3 and 4 added the option of specifying external user-provided CCD using the newly added field "userCCDPath", the option of specifying textual description of protein chains, RNA chains, DNA chains, or ligands, respectively. These two versions, however, do not concern this study.

Protein-specific Input Script

The "protein" Section is the most important for this study since it involves the target sequence, which has to be provided as a string, easily obtainable from the FASTA file available on PDB or UniProt. The second subsection is named "modifications" and allows to write a list of optional post-translational modifications, with each modification is specified using its CCD code and 1-based residue position. Since however, the aim is to predict an unknown protein structure, this section is also redundant, and the list can be left empty. The provided sequence has to be given an Id, which can simply be a character (not influential). In the subsection "unpairedMSA"/"unpairedMsaPath" it is possible to provide a string contining a custom MSA or the path to a file containing a custom MSA for the target sequence. The "pairedMSA"/ "pairedMsaPath" deserves a bit of attention: MSA a pairing is important only when predicting multimers, since there is a need to concatenate MSAs for the individual chains along the sequence dimension. However, the pairing of the MSA must be performed with great care, since the simple row-wise concatenation and padding of the various MSAs so that they have the same length, might cause sequences belonging to different organisms to end up on the same row, worsening AF3 performances. The pairedMSA option is not recommended; in the case the user may want to use it, it's better to use an appropriate software to handle concatenation. "pairedMSApath" works exactly as its unpaired counterpart. All these four options are however advanced and redundant, since all we know about the FABP12 is its sequence, and we rely on AF3 to build the MSA, so they

are omitted in the JSON files used in this study. What is truly of interests is the structural template section. First, it's necessary to provide either a string containing the mmCIF file of the selected template structure ("mmCIF") or the path to the mmCIF file. Then it's necessary to fill two arrays ("queryindices" and "templateindices") containing O-based indices in the query sequence, defining the mapping from query residues to template residues. The mapping has been obtained via a Python script, and describes the indices, not accounting for the gaps that resulted from the alignment, of the identical residues in the sequences.


```
{
  "name": "FABP12_A6NFH5_Temp:FABP8_t1",
  "modelSeeds": [891, 756, 522, 189, 762, 972, 739, 328, 204, 722],
  "sequences": [
    {
      "protein": {
        "id": ["A"],
        "sequence": "MIDQLQGTWKSISCENSEDYMKELGIGRASRKLGRLAKPTVTISTDGDVITIKTKSIFKNNEISFKLGEEFEEITPGGHKTKSKVTLDKESLIQVQDWIQVQDWDGKETTITRKLVDGKMVVESTVNSVICTRTYEKVSSNSVSNS",

        "templates": [
          {
           "mmcifPath": "scratch/alphafold/input/tmpl.cif",
           "queryIndices": [0, 4, 6, 7, 8, 9, 12, 14, 15, 18, 19, 20, 21, 23, 24, 26, 28, 30, 31, 32, 33, 35, 36, 37, 38, 39, 40, 42, 43, 46, 47, 49, 50, 51, 53, 55, 57, 58, 59, 61, 62, 63,
64, 65, 66, 67, 69, 70, 71, 72, 74, 79, 80, 81, 82, 84, 85, 86, 90, 91, 93, 94, 95, 97, 98, 99, 100, 101, 102, 103, 104, 106, 107, 108, 109, 111, 112, 113, 114, 116, 122, 124, 125, 126, 128,
129, 130, 131],
           "templateIndices": [0, 5, 6, 7, 8, 9, 12, 14, 15, 18, 19, 20, 21, 23, 24, 26, 28, 30, 31, 32, 33, 35, 36, 37, 38, 39, 40, 42, 43, 46, 47, 49, 50, 51, 53, 55, 57, 58, 59, 61, 62,
63, 64, 65, 66, 67, 69, 70, 71, 72, 74, 79, 80, 81, 82, 84, 85, 86, 90, 91, 93, 94, 95, 97, 98, 99, 100, 101, 102, 103, 104, 106, 107, 108, 109, 111, 112, 113, 114, 116, 122, 124, 125, 126,
128, 129, 130, 131]
          }
        ],
        "modifications": []
      }
    }
  ],
  "bondedAtomPairs": [],
  "userCCD": null,
  "dialect": "alphafold3",
  "version": 2
}
```


Figure 4.4: Example of an input *.JSON file. In this case, the input used to generate the FABP12 structure with its chosen template, FABP8

The known FABP structures were used as control to fine-tune the previously mentioned code parameters. Several tests were run using different combination of parameters, raising or reducing the MSA depth (max=1024, default setting; min=128); changing the number of cycles (max=20; min=10, default setting). In the end it was decided the use of the higher option for both parameters, since this combination did not show a significative increase in computational times compared to other, possibly less performing, options.

4.3.2 Outputs

AF3 produces in output a directory containing a .cif file of the predicted structure and two JSON files containing AF3 confidence scores, which represent AF3's estimate of the quality of the predicted structure compared to experimental similar structures in the accessible databases:

- pLDDt: predicted local distance difference test (pLDDT) is a per-residue measure of local confidence. It is scaled from 0 to 100, with higher scores indicating higher confidence and usually a more accurate prediction. The pLDDT score can vary significantly along a protein chain. This means the

software can be very confident in the structure of some parts of the protein, but less confident in other regions.

- Predicted aligned error (PAE): is a measure of how confident AlphaFold is in the relative position of two residues within the predicted structure. It's defined as the expected positional error at residue X, measured in Ångströms (Å), if the predicted and actual structures were aligned on residue Y. It measures the confidence of the software in the packing of the domains and that the relative placement of the domains in the predicted structure is correct. The lower the PAE, the better the prediction.

- pTM: integrated measure of how well the overall structure of the complex has been predicted. It is the predicted Template Modeling score for a superposition between the predicted structure and the hypothetical true structure. A score above 0.5 means the overall predicted fold for the complex might be similar to the true structure. Below 0.5 means the predicted structure is likely wrong. pTM score should be interpreted cautiously. In a situation where one of the interacting proteins is larger, and its structure is predicted correctly, while a smaller partner protein is predicted incorrectly. The resulting pTM score of the complex could be dominated by the larger protein and show a pTM score above 0.5.

- ipTM: measures the accuracy of the predicted relative positions of the subunits forming the protein-protein complex. Values higher than 0.8 represent confident, high-quality predictions, while values below 0.6 suggest likely a failed prediction. ipTM values between 0.6 and 0.8 are a grey zone where predictions could be correct or wrong.

However, pTM and ipTM are mainly necessary when using Alphafold Multimer to predict protein complexes. While pTM can still be calculated, it was not considered a prominent indicator of prediction quality. ipTM cannot be calculated when predicting single-subunit structures such as the FABPs. AlphaFold also gives a general ranking score, between 0 and 1, which summarizes the others, and from this point forward this ranking score (RS) will be used to address the quality of each AlphaFold prediction.

4.3.3 Testing

Each structure was first predicted without employing templates, and a second test was run employing, as a template, the FABP with the highest degree of identity, therefore using the exact same conditions as it was done for Homology modeling. The use of template structures with an identity percentage equal or superior to 58% produced significantly better predictions. For some structures, namely FABP1,

FABP2 and FABP6, however, it was not possible to identify a viable template structure. This allowed to coherently compare the two methodologies and decide the better approach for the prediction of FABP12. Each FABP structure was obtained and loaded into MOE to be validated against its Homology Modeling and experimental counterparts using RMSD as measuring parameter, following the same preparation process explained in Chapter 3.

Chapter 5

Results and Discussion

5.1 Results

	FABP1	FABP2	FABP3	FABP4	FABP5
RMSD to Mean	0.446393	0.521152	0	0	0
CA RMSD to Mean	0.360676	0.418036	0	0	0
Contact Energy	-86.3951	-89.4004	-103.734	-103.454	-93.8448
Packing Score	2.21299	2.14705	2.144665	2.100362	2.125764
GB/VI	-5646.62	-6854.98	-5939.97	-5976.68	-6407.06
U	-917.553	-1612.84	-1212.29	-1218.91	-1386.51
E sol	-741.365	-768.648	-721.317	-736.276	-960.752
E ele	-2878.2	-3597.57	-3086.82	-3024.06	-3277.25
E vdw	-406.153	-474.02	-502.108	-497.793	-481.196
E bond	2366.8	2458.751	2376.636	2302.95	2371.937
Atom Clashes	0	0	0	0	0
BB Bond Outliers	0	0	0	0	0
BB Angle Outliers	2	1	0	0	0
BB Torsion Outliers	2	1	3	2	2
Rotamer Outliers	0	0	1	0	0

Table 5.1: Homology Modeling scores for experimental FABPs, obtained for methodology evaluation; Part 1

	FABP6	FABP7	FABP8	FABP9
RMSD to Mean	0.341471	0	0	0
CA RMSD to Mean	0.237414	0	0	0
Contact Energy	-104.643	-112.279	-98.1657	-102.582
Packing Score	2.12181	2.060816	2.052301	2.078849
GB/VI	-5550.82	-6244.67	-6301.88	-6036.07
U	-1094.47	-1381.19	-1403.08	-1206.56
E sol	-312.406	-860.347	-1003.96	-957.048
E ele	-2904.38	-3212.54	-3224.73	-3095.63
E vdw	-471.928	-493.757	-487.779	-519.616
E bond	2281.843	2325.111	2309.421	2408.687
Atom Clashes	0	0	0	0
BB Bond Outliers	0	0	0	0
BB Angle Outliers	0	2	1	2
BB Torsion Outliers	1	0	4	3
Rotamer Outliers	1	1	0	0

Table 5.2: Homology Modeling scores for experimental FABPs, obtained for methodology evaluation; Part 2

	Template AF	**General Ranking Score**	**pLDDT**	**pTM**
FABP1	///	0.9	94.66	0.9
FABP2	///	0.91	94.66	0.91
FABP3	FABP7 (6L90)	0.95	96.98	0.95
FABP4	FABP8 (4BVM)	0.94	96.28	0.94
FABP5	FABP8 (4BVM)	0.93	95.4	0.93
FABP6	///	0.9	93.71	0.9
FABP7	FABP3	0.93	96.02	0.93
FABP8	FABP9	0.94	96.4	0.94
FABP9	FABP8	0.93	95.87	0.93

Table 5.3: AF3 confidence scores for test-modeling experimental structures

Target	Template AF	Template HM (PDB id)	%Id	RMSD VS AF (Å)	RMSD VS HM (Å)
FABP1	///	8WKH	36.7	1.03	1.997
FABP2	///	6L90	36.2	1.185	1.1899
FABP3	FABP7 (6L90)	6L90	67.2	0.229	1.112
FABP4	FABP8 (4BVM)	4BVM	67.2	0.25	1.001
FABP5	FABP8 (4BVM)	4BVM	58.7	0.59	1.348
FABP6	///	6DO6	36.7	2.331	2.321
FABP7	FABP3	3WVM	67.2	0.358	1.009
FABP8	FABP9	7FWK	67.4	0.2	1.177
FABP9	FABP8	4BVM	67.4	0.448	1.26

Table 5.4: Comparison between HM and AF3 prediction of experimental structures:

	FABP12 HM
RMSD to Mean	0
CA RMSD to Mean	0
Contact Energy	-88.47572
Packing Score	2.0328569
GB/VI	-6036.889
U	-997.7667
E sol	-1231.176
E ele	-2903.372
E vdw	-469.7021
E bond	2375.3069
Atom Clashes	0
BB Bond Outliers	0
BB Angle Outliers	1
BB Torsion Outliers	2
Rotamer Outliers	0

Table 5.5: Homology modeling results for FABP12:

	Template	General Ranking Score	pLDDT	pTM
FABP12	FABP8(4BVM)	0.94	92.59	0.9

Table 5.6: AF3 prediction results for FABP12

Target	Template AF(PDBid)	Template HM(PDBid)	%Id	RMSD AF VS HM(Å)
FABP12	FABP8 (4BVM)	4BVM	65.4	0.95

Table 5.7: FABP12 HM and AF3 prediction results comparison.

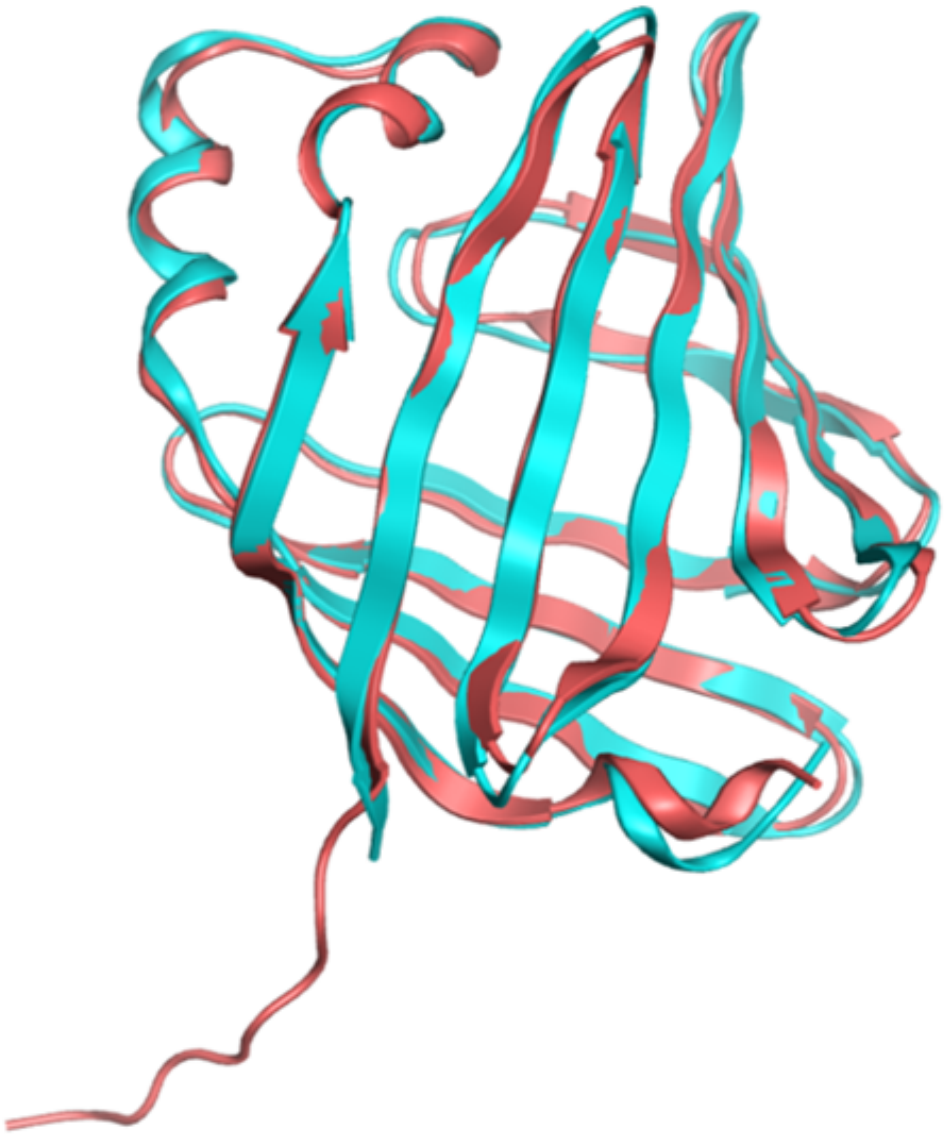

Figure 5.1: Superposition of FABP12 structure obtained through homology modelling (in blue) and Alphafold (in red), obtained via MOE

5.2 Discussion

To accomplish the goal to predict the tridimensional structure of FABP12, the Homology modeling methodology, performed using the MOE software, and the AI-powered predictions obtained with AlphaFold3 software were compared, and the results were validated using the already established tridimensional structures the other FABPs as control. The crystallized structures of established FABPs were obtained from open-source databases (such as PDB and UniProt). The structures were selected, where possible, as non-complexed structures obtained with the same experimental methodology (X-RAY diffraction) and with the lowest possible resolution (between 0.9 and 1.88 Å). As previously stated, for each FABP structure, the template used was another FABP with the greatest degree of sequence identity, as shown in the following table. The FABP8 was also used as a template to generate the possible new structure for the FABP12 (ID% 65.91).

Target	Template	ID%
FABP1	FABP6	36.72
FABP2	FABP7	36.15
FABP3	FABP7	67.18
FABP4	FABP8	67.18
FABP5	FABP8	58.73
FABP6	FABP1	36.72
FABP7	FABP3	67.18
FABP8	FABP9	67.42
FABP9	FABP8	67.42
FABP12	FABP8	65.91

Table 5.8: Established FABP structures and their chosen template, employed for methodology validation. While necessary in HM, the template was not used in the final prediction test for FABP1, FABP2, and FABP6 structures in AF3, because of the low identity percentage.

In reviewing the results, Alphafold clearly emerged as the superior methodology, yielding predictions significantly closer, in terms of RMSD, to their experimental counterparts, with an average RMSD of 0.74Å, against the average 1.38 RMSD of the Homology Modeling structures. However, the use of templates with a sequence identity of 58% or more, significantly improved the prediction accuracy of AF3. Excluding FABP1, FABP2, and FABP6, for which it was not possible to find a suitable template, since their maximum identity percentage with any of the other FABPs does not exceed 35%, the average RMSD of the AF3 structures drops to 0.35 Å, against the 1.15 Å average RMSD of the Homology modeling structures.

The sole exception is FABP6, for which HM performed slightly better than AF. This might be linked both with the absence of a good template, at least by the standards decided in this study, and the fact that, on average, FABP6 exhibits the lowest degree of identity to the other FABPs, as shown in Table5.8.

Average FABP	ID%
FABP1	30.02
FABP2	30.06
FABP3	49.68
FABP4	50.29
FABP5	43.9
FABP6	27.35
FABP7	47.15
FABP8	51.84
FABP9	48.068
FABP12	46.30

Table 5.9: Average identity percentage, calculated by the average of each row of the identity matrix, excluding, for each FABP, the percentage of identity with itself

This not only highlights the higher accuracy of AF3 predictions, but also the importance of good quality templates in both methodologies. The first test performed on the Alphafold software were conducted simply by giving a target FABP sequence in input and leaving the other parameters' default settings. Ranking Scores in this step of the process were between 0.75 and 0.83. The most significant change was raising the number of random seeds, settling for a number of 10 random seeds, which was considered a good compromise between result quality and computational cost. Modifying other code parameters did not cause significative positive changes in the general final scores. Higher depths of MSA yielded better results, though the final RS did not exceed 0.83. The use of templates to better direct the predictions was what caused the most significative changes in the score results, yielding an average score of 0.94 with none of the predicted structures having a score <0.93. Once the template were employed, any variation of other code parameters, such as different combinations of MSA, number of cycles, and random seeds) did not lead to significant differences, showing a plateauing of prediction confidence in AF3 output scores. Confidence scores for the AF3 predicted structures are reported in Table5.2.
Results were validated against structures obtained with the gold standard of in silico modelling, the homology modelling methodology, and the crystallized experimental structures. HM requires both a target sequence and a template to predict a

structure. For each FABP, the same template used for AF predictions were also used for HM. Both AF and HM structure were loaded on the MOE software, and the RMSD (root mean square deviation) was used to determine the difference between predicted and experimentally observed atom position. In the case of FABP12, the AF structure and its HM corresponsive were confronted. Results are shown in the tables reported in the previous section, with a visual representation of the superimposition of FABP12's AF and HM structures.

To obtain a graphical representation of the FABP structures obtained with the AF software, the PAE viewer free online tool was used. This allows a user-friendly interface to easily understand the results.

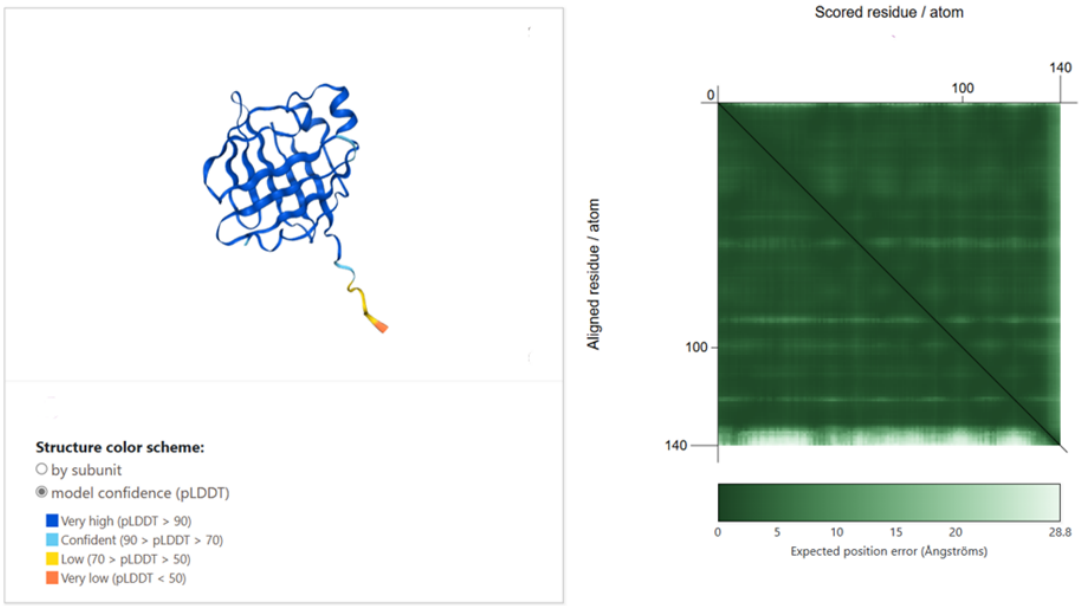

Figure 5.2: Image of FABP12, color-coded according to pLDDT, and associated PAE diagram, obtained on PAE viewer

5.3 Conclusions and Future perspectives

Prostate cancer is the most preminent malignance in men, second only to lung cancer [1], and identification of biomarkers, such as PSA, correlated with its diagnosis already revolutionized the epidemiology of the disease [158]. The FABP family has been proven, in multiple studies, to be connected with tumorigenesis, spreading, and overall cancer aggressiveness. Upregulation of FABPs, in fact, is connected to reduced disease survivability [159] [102], in EMT, which is the gateway to metastatic disease degeneration [88] [103], which begs the question if they can confidently be used as markers of poor cancer prognosis. Not all FABPs, in fact, behave in the same way: FABP1, FABP2, and FABP6 exhibit an ambiguous behaviour, in some cases even acting as tumor suppressors or showing a negative correlation between their expression and metastasis manifestation [99] [67][72], suggesting more extensive research. On the other hand, the inhibition of FABPs, which translates to their inability to bind FAs, has shown some effectiveness in slowing down disease progression in CRPC [97] and, in particular, inhibition of FABP12 with non-specific FABP inhibitor, T4B, in PC3 cell line model for PCa, caused a more significant reduction of cell motility compared to inhibition of PC3 cells expressing FABP4 and/or FABP5[118], suggesting FABP12's viability as a therapeutic target for non-hormonal PCa treatment. While experimental methodologies remain the gold standard for protein structure establishment, in silico predictions can circumvent their issues of costs and necessary resources, such as time, highly trained personnel, and equipped laboratories. In this study, two methodologies were compared, Homology Modeling and AI predictions, in order to determine which of them could produce a better FABP12 structure, leveraging the data already available on the other FABP structures. Testing and validation processes have shown that, while Homology Modeling is still a reliable tool for structure prediction, AI predictions, obtained via AlphaFold3 software, showed a far greater accuracy, consistently providing more accurate results, especially with high-identity templates available. The average RMSD value for AF3 predicted structures was well under the 1 Å threshold, comparable with experimental structures. The exception presented by FABP6 strengthens the idea that the base for good predictions is in the source data and in the quantity of shared residues between target sequences and in structures used as templates and/or input data for a neural network. Successful prediction of FABP12 structure should not, however, dampen the efforts to obtain an experimentally resolved FABP12 via X-RAY crystallography or solution NMR, eventually validating the results of in silico studies such as this and help with refine them. The structural model might now enable future studies to experiment with highly specific testable predictions, such as mutational analysis of residues in the binding pocket and how such mutations might affect fatty acid, ligand interaction, and overall metabolic activity and signalling (mutational prediction and analysis

can be performed with a conjunction of the same softwares used for this study, AF3 and MOE). These studies will help clarifying the role of FABP12 in PCa and will go hand-in-hand with in vitro experiments on prostate cancer cell lines Predicted structures can in no way completely substitute experimentally obtained ones, which will need to be eventually obtained in future studies, allowing physical validation and even more accurate predictions of structures and of target structure interaction with a given ligand. They can, however, be useful to perform the first stages of research for possible inhibitors, which represents the natural prosecution of this study. Predicted structures, in fact, do not exist in a vacuum, but constitute the base of a molecular dynamics process named Virtual Screening

5.3.1 Future perspectives:Molecular Dynamics, Docking

After establishing a structural foundation, future studies might follow various roads for development, eventually reaching possible highly specific inhibitors for FABP12, providing an alternative/combination therapy with ADT. The structures obtained in this study will need to go through Molecular Dynamics simulations and study the conformational evolution of the target systems in different thermodynamic conditions, assessing the flexibility of the overall structure, in particular of the 2 α-helixes motifs, which constitute the gateway to the binding site. This means tracking the potential energy states of the target system, in our case a protein, in an environment as similar as possible to the physiological one, therefore requiring a pre-emptive solvation step to model water interactions. The atoms' coordinates and velocities variations are evaluated as a function of the applied thermodynamic parameters at each time step, usually in the order of magnitude of the femtosecond (10^{-15}). This process is necessary to reveal conformational states in the target protein which are relevant to drug design. Ligand interactions analysis will need to be in rounds of increasing accuracy, with the first step consisting in creating a robust database of chemical compounds which have been tested for FABPs, taking advantage of libraries such as ZINC [134], a free downloadable database that contains over 230 million purchasable compounds already prepared for ligand-interactions simulations and over 750 million purchasable compounds that allow quick analogue search. Ligand interaction simulations are performed through the Docking process, which analyses the optimal spatial conformation and relative position between a target system and a ligand, evaluating the affinity of the coupled system. The Docking evaluation is fundamentally a brute-strength screening method, that greatly reduces the scope of ligand research, highlighting compounds with similar structures, and evaluating binding affinity and selectivity across the FABP protein family. Once again, the research will leverage the principle that similar chemical structure implies similar structure and behaviour, and move onto Pharmacophore-based searches, repurposing compounds with a high degree of similarity to the ones

found in Docking evaluations and finding a common base structure (scaffold) from which might be possible to derivatize a ligand (Pharmacophore modeling). A library of similar compounds can also be useful to perform QSAR (Quantitative Structure-Activity Relationship) and search for a mathematical relationship between the relative biological properties of a set of molecules and one or more of their calculated or physical properties. Various softwares can perform these tasks, such as DOCK [160], Autodock [161] and MOE, which is, for now, arguably the most complete one in the number of available softwares. Finally, compounds emerging from this thorough selection process will be able to go forward to in vitro testing, producing feedback that will refine new iterations of prediction/virtual screening, until the conjunction of in silico and in vitro studies will highlight a set of compounds viable for clinical trials.

5.3.2 Final Remarks

This study demonstrates that AlphaFold3, a cutting-edge AI-based structural prediction software, can successfully complement and even overperform classical computational methods when applied to structurally unresolved proteins. By combining the strengths of HM and AF3, this study provides a structurally informed foundation for the study of FABP12, opening the way toward deeper functional understanding and to new potential drug research that might benefit from the data presented here to find and/or design inhibitor compounds to serve as new therapeutic tools against CRPC. The project not only contributes to the field of lipid-mediated tumorigenesis, and of innovative therapy resaerch but also exemplifies how computational structural biology is reshaping molecular research in the post-AI era.

Chapter 6

Supplementary Materials

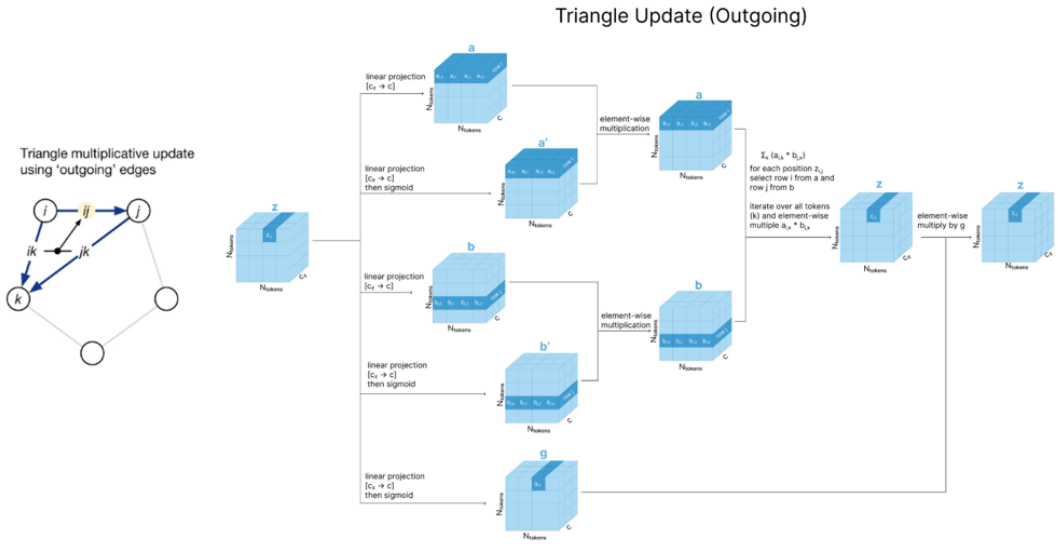

Figure 6.1: Visual representation of pair representation update: on the left, edges diagram highlighting ij element being updated; on the right, practical operations performed on the pair representation tensor. Images taken from [148] and [153]

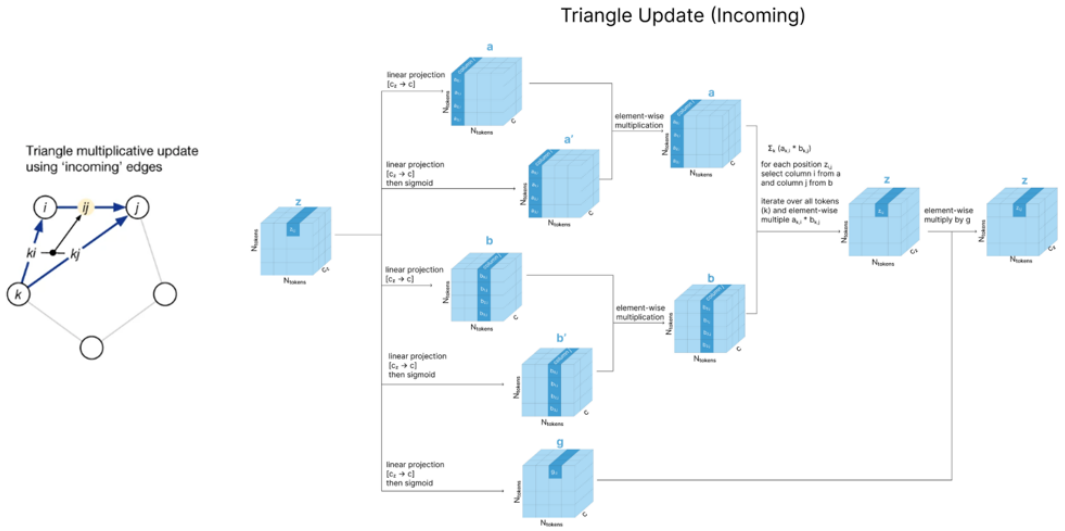

Figure 6.2: Visual representation of pair representation update: on the left, edges diagram highlighting ij element being updated; on the right, practical operations performed on the pair representation tensor. Images taken from [148] and [153]

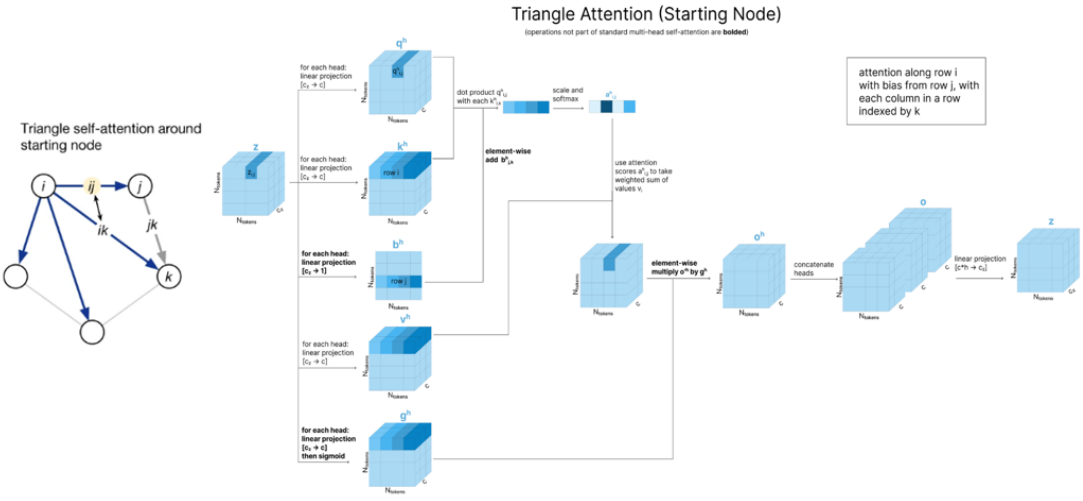

Figure 6.3: Visual representation of pair representation update: on the left, edges diagram highlighting ij element being updated; on the right, practical operations performed on the pair representation tensor. Images taken from [148] and [153]

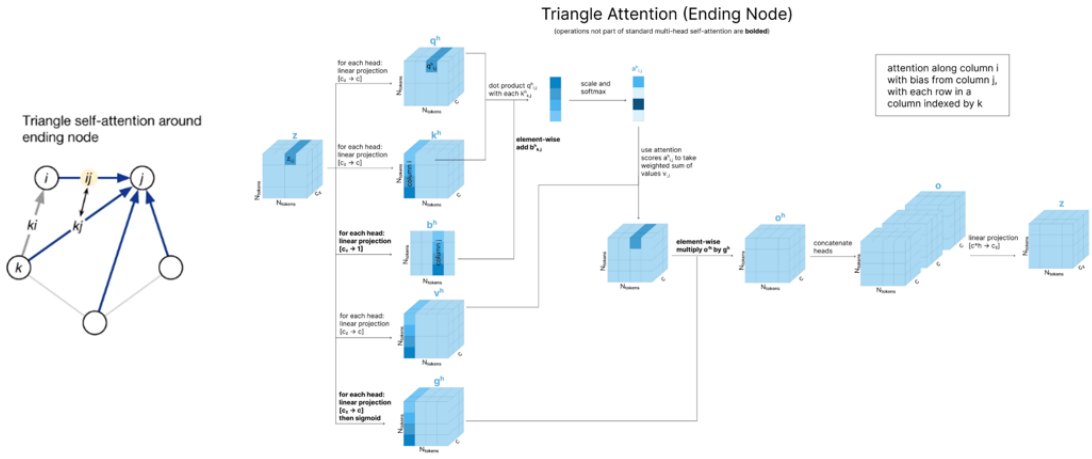

Figure 6.4: Visual representation of pair representation update: on the left, edges diagram highlighting ij element being updated; on the right, practical operations performed on the pair representation tensor. Images taken from [148] and [153]

Identity Matrix										
	FABP1	FABP2	FABP3	FABP4	FABP5	FABP6	FABP7	FABP8	FABP9	FABP12
FABP1	100	25.37	29.67	27.91	28.4	36.72	30.4	29.76	29.41	32.56
FABP2	25.37	100	33.59	33.85	27.13	26.52	36.15	31.78	29.23	26.92
FABP3	29.67	33.59	100	64.89	50	30.77	67.18	63.36	55.73	51.91
FABP4	27.91	33.85	64.89	100	55.2	26.92	56.82	67.18	64.12	55.73
FABP5	28.4	27.13	50	55.2	100	26.67	47.15	58.73	51.15	50.77
FABP6	36.72	26.52	30.77	26.92	26.67	100	25.27	23.2	25.27	24.83
FABP7	30.4	36.15	67.18	56.82	47.15	25.27	100	59.54	53.03	48.85
FABP8	29.76	31.78	63.36	67.18	58.73	23.2	59.54	100	67.42	65.65
FABP9	29.41	29.23	55.73	64.12	51.15	25.27	53.03	67.42	100	57.25
FABP12	32.56	27.13	51.91	55.73	51.15	25.89	48.85	65.91	57.58	100

Table 6.1: Matrix showing the identity percentage of each FABP with each other, with the identity percentage of each FABP relative to itself on the diagonal. The matrix, obtained on UniProt-s BLAST[110], has been used to select the templates for the prediction of each FABP, according to the standards decided for this study, both for HM and AF3

Bibliography

[1] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel, Lindsey A. Torre, and Ahmedin Jemal. «Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries». In: *CA: A Cancer Journal for Clinicians* 68 (6 Nov. 2018), pp. 394–424. ISSN: 0007-9235. DOI: 10.3322/caac.21492 (cit. on pp. i, 57).

[2] Richard J. Rebello, Christoph Oing, Karen E. Knudsen, Stacy Loeb, David C. Johnson, Robert E. Reiter, Silke Gillessen, Theodorus Van der Kwast, and Robert G. Bristow. «Prostate cancer». In: *Nature Reviews Disease Primers* 7 (1 Feb. 2021), p. 9. ISSN: 2056-676X. DOI: 10.1038/s41572-020-00243-0. URL: https://www.nature.com/articles/s41572-020-00243-0 (cit. on pp. i, 1, 2, 4).

[3] Paolo Verze, Tommaso Cai, and Stefano Lorenzetti. «The role of the prostate in male fertility, health and disease». In: *Nature Reviews Urology* 13 (7 July 2016), pp. 379–386. ISSN: 1759-4812. DOI: 10.1038/nrurol.2016.89. URL: https://www.nature.com/articles/nrurol.2016.89 (cit. on pp. 1–3).

[4] Florian M.E. Wagenlehner, Johny E. Elkahwaji, Ferran Algaba, Truls Bjerklund-Johansen, Kurt G. Naber, Rudolf Hartung, and Wolfgang Weidner. «The role of inflammation and infection in the pathogenesis of prostate carcinoma». In: *BJU International* 100 (4 Oct. 2007), pp. 733–737. ISSN: 1464-4096. DOI: 10.1111/j.1464-410X.2007.07091.x (cit. on p. 1).

[5] Vincenzo Ficarra, Marta Rossanese, Michele Zazzara, Gianluca Giannarini, Maria Abbinante, Riccardo Bartoletti, Vincenzo Mirone, and Francesco Scaglione. «The Role of Inflammation in Lower Urinary Tract Symptoms (LUTS) due to Benign Prostatic Hyperplasia (BPH) and Its Potential Impact on Medical Therapy». In: *Current Urology Reports* 15 (12 Dec. 2014), p. 463. ISSN: 1527-2737. DOI: 10.1007/s11934-014-0463-9 (cit. on p. 1).

[6] Cera M Nieto, Leah C Rider, and Scott D Cramer. «Influence of stromal–epithelial interactions on androgen action». In: *Endocrine-Related Cancer* 21 (4 Aug. 2014), T147–T160. ISSN: 1351-0088. DOI: 10.1530/ERC-14-0138 (cit. on p. 2).

[7] Renty B Franklin, Beatrice Milon, Pei Feng, and Leslie C Costello. «Zinc and zinc transporters in normal prostate and the pathogenesis of prostate cancer.» In: *Frontiers in bioscience : a journal and virtual library* 10 (Sept. 2005), pp. 2230–9. ISSN: 1093-9946. DOI: 10.2741/1692 (cit. on p. 3).

[8] M.-C. Franz, P. Anderle, M. Bürzle, Y. Suzuki, M.R. Freeman, M.A. Hediger, and G. Kovacs. «Zinc transporters in prostate cancer». In: *Molecular Aspects of Medicine* 34 (2-3 Apr. 2013), pp. 735–741. ISSN: 00982997. DOI: 10.1016/j.mam.2012.11.007 (cit. on p. 3).

[9] Ken-ichi Takayama and Satoshi Inoue. «Transcriptional network of androgen receptor in prostate cancer progression». In: *International Journal of Urology* 20 (8 Aug. 2013), pp. 756–768. ISSN: 0919-8172. DOI: 10.1111/iju.12146 (cit. on p. 3).

[10] Z X Zhou, M V Lane, J A Kemppainen, F S French, and E M Wilson. «Specificity of ligand-dependent androgen receptor stabilization: receptor domain interactions influence ligand dissociation and receptor stability.» In: *Molecular Endocrinology* 9 (2 Feb. 1995), pp. 208–218. ISSN: 0888-8809. DOI: 10.1210/mend.9.2.7776971 (cit. on p. 3).

[11] Elizabeth J. Schafer, Mathieu Laversanne, Hyuna Sung, Isabelle Soerjomataram, Alberto Briganti, William Dahut, Freddie Bray, and Ahmedin Jemal. «Recent Patterns and Trends in Global Prostate Cancer Incidence and Mortality: An Update». In: *European Urology* 87 (3 Mar. 2025), pp. 302–313. ISSN: 03022838. DOI: 10.1016/j.eururo.2024.11.013. URL: https://linkinghub.elsevier.com/retrieve/pii/S0302283824027076 (cit. on pp. 4–6).

[12] Martin C.S. Wong, William B. Goggins, Harry H.X. Wang, Franklin D.H. Fung, Colette Leung, Samuel Y.S. Wong, Chi Fai Ng, and Joseph J.Y. Sung. «Global Incidence and Mortality for Prostate Cancer: Analysis of Temporal Patterns and Trends in 36 Countries». In: *European Urology* 70 (5 Nov. 2016), pp. 862–874. ISSN: 03022838. DOI: 10.1016/j.eururo.2016.05.043. URL: https://linkinghub.elsevier.com/retrieve/pii/S0302283816302512 (cit. on p. 4).

[13] L.E. Johns and R.S. Houlston. «A systematic review and meta-analysis of familial prostate cancer risk». In: *BJU International* 91 (9 June 2003), pp. 789–794. ISSN: 1464-4096. DOI: 10.1046/j.1464-410X.2003.04232.x (cit. on p. 4).

[14] Rebecca L. Siegel, Kimberly D. Miller, and Ahmedin Jemal. «Cancer statistics, 2018». In: *CA: A Cancer Journal for Clinicians* 68 (1 Jan. 2018), pp. 7–30. ISSN: 0007-9235. DOI: 10.3322/caac.21442 (cit. on pp. 4, 5).

[15] Richard B. Berish, Aymon N. Ali, Patrick G. Telmer, John A. Ronald, and Hon S. Leong. «Translational models of prostate cancer bone metastasis». In: *Nature Reviews Urology* 15 (7 July 2018), pp. 403–421. ISSN: 1759-4812. DOI: 10.1038/s41585-018-0020-2 (cit. on p. 4).

[16] Katherine S. Virgo et al. «Second-Line Hormonal Therapy for Men With Chemotherapy-Naïve, Castration-Resistant Prostate Cancer: American Society of Clinical Oncology Provisional Clinical Opinion». In: *Journal of Clinical Oncology* 35 (17 June 2017), pp. 1952–1964. ISSN: 0732-183X. DOI: 10.1200/JCO.2017.72.8030 (cit. on p. 5).

[17] Adam Abeshouse et al. «The Molecular Taxonomy of Primary Prostate Cancer». In: *Cell* 163 (4 Nov. 2015), pp. 1011–1025. ISSN: 00928674. DOI: 10.1016/j.cell.2015.10.025 (cit. on p. 6).

[18] Michael Fraser, Julie Livingstone, Jeffrey L. Wrana, Antonio Finelli, Housheng Hansen He, Theodorus van der Kwast, Alexandre R. Zlotta, Robert G. Bristow, and Paul C. Boutros. «Somatic driver mutation prevalence in 1844 prostate cancers identifies ZNRF3 loss as a predictor of metastatic relapse». In: *Nature Communications* 12 (1 Oct. 2021), p. 6248. ISSN: 2041-1723. DOI: 10.1038/s41467-021-26489-0 (cit. on p. 6).

[19] Ping Gao et al. «Biology and Clinical Implications of the 19q13 Aggressive Prostate Cancer Susceptibility Locus». In: *Cell* 174 (3 July 2018), 576–589.e18. ISSN: 00928674. DOI: 10.1016/j.cell.2018.06.003 (cit. on p. 6).

[20] Peter A Humphrey. «Gleason grading and prognostic factors in carcinoma of the prostate». In: *Modern Pathology* 17 (3 Mar. 2004), pp. 292–306. ISSN: 08933952. DOI: 10.1038/modpathol.3800054 (cit. on p. 7).

[21] Lars Egevad, Brett Delahunt, John R Srigley, and Hemamali Samaratunga. «International Society of Urological Pathology (ISUP) grading of prostate cancer – An ISUP consensus on contemporary grading». In: *APMIS* 124 (6 June 2016), pp. 433–435. ISSN: 0903-4641. DOI: 10.1111/apm.12533 (cit. on p. 8).

[22] William T. Lowrance, James A. Eastham, Caroline Savage, A.C. Maschino, Vincent P. Laudone, Christopher B. Dechet, Robert A. Stephenson, Peter T. Scardino, and Jaspreet S. Sandhu. «Contemporary Open and Robotic Radical Prostatectomy Practice Patterns Among Urologists in the United States». In: *Journal of Urology* 187 (6 June 2012), pp. 2087–2093. ISSN: 0022-5347. DOI: 10.1016/j.juro.2012.01.061 (cit. on p. 8).

[23] Ulrika Harmenberg, Freddie C. Hamdy, Anders Widmark, Bo Lennernäs, and Sten Nilsson. «Curative radiation therapy in prostate cancer». In: *Acta Oncologica* 50 (sup1 June 2011), pp. 98–103. ISSN: 0284-186X. DOI: 10.3109/0284186X.2010.576115 (cit. on p. 8).

[24] Oliver Sartor et al. «Lutetium-177–PSMA-617 for Metastatic Castration-Resistant Prostate Cancer». In: *New England Journal of Medicine* 385 (12 Sept. 2021), pp. 1091–1103. ISSN: 0028-4793. DOI: 10.1056/NEJMoa2107322 (cit. on p. 8).

[25] Bo-Ren Wang, Yu-An Chen, Wei-Hsiang Kao, Chih-Ho Lai, Ho Lin, and Jer-Tsong Hsieh. «Developing New Treatment Options for Castration-Resistant Prostate Cancer and Recurrent Disease». In: *Biomedicines* 10 (8 Aug. 2022), p. 1872. ISSN: 2227-9059. DOI: 10.3390/biomedicines10081872 (cit. on p. 8).

[26] Sobia Wasim, Sang-Yoon Lee, and Jaehong Kim. «Complexities of Prostate Cancer». In: *International Journal of Molecular Sciences* 23 (22 Nov. 2022), p. 14257. ISSN: 1422-0067. DOI: 10.3390/ijms232214257. URL: https://www.mdpi.com/1422-0067/23/22/14257 (cit. on p. 8).

[27] Helen E. MacLean, Garry L. Warne, and Jeffrey D. Zajac. «Localization of functional domains in the androgen receptor». In: *The Journal of Steroid Biochemistry and Molecular Biology* 62.4 (July 1997), pp. 233–242. ISSN: 09600760. DOI: 10.1016/S0960-0760(97)00049-6 (cit. on p. 8).

[28] Laia Querol Cano, Derek N. Lavery, and Charlotte L. Bevan. «Mini-review: Foldosome regulation of androgen receptor action in prostate cancer». In: *Molecular and Cellular Endocrinology* 369 (1-2 Apr. 2013), pp. 52–62. ISSN: 03037207. DOI: 10.1016/j.mce.2013.01.023 (cit. on p. 8).

[29] Meghan A. Rice, Sanjay V. Malhotra, and Tanya Stoyanova. «Second-Generation Antiandrogens: From Discovery to Standard of Care in Castration Resistant Prostate Cancer». In: *Frontiers in Oncology* 9 (Aug. 2019). ISSN: 2234-943X. DOI: 10.3389/fonc.2019.00801 (cit. on p. 8).

[30] MH Eileen Tan, Jun Li, H Eric Xu, Karsten Melcher, and Eu-leong Yong. «Androgen receptor: structure, role in prostate cancer and drug discovery». In: *Acta Pharmacologica Sinica* 36 (1 Jan. 2015), pp. 3–23. ISSN: 1671-4083. DOI: 10.1038/aps.2014.18 (cit. on pp. 8, 9).

[31] Alastair D. Lamb, Charlie E. Massie, and David E. Neal. «The transcriptional programme of the androgen receptor (AR) in prostate cancer». In: *BJU International* 113 (3 Mar. 2014), pp. 358–366. ISSN: 1464-4096. DOI: 10.1111/bju.12415 (cit. on p. 9).

[32] Kazutoshi Fujita and Norio Nonomura. «Role of Androgen Receptor in Prostate Cancer: A Review». In: *The World Journal of Men's Health* 37 (3 2019), p. 288. ISSN: 2287-4208. DOI: 10.5534/wjmh.180040 (cit. on p. 9).

[33] Ayesha A. Shafi, Aihua E. Yen, and Nancy L. Weigel. «Androgen receptors in hormone-dependent and castration-resistant prostate cancer». In: *Pharmacology & Therapeutics* 140 (3 Dec. 2013), pp. 223–238. ISSN: 01637258. DOI: 10.1016/j.pharmthera.2013.07.003 (cit. on p. 9).

[34] Mary-Ellen Taplin, Glenn J. Bubley, Todd D. Shuster, Martha E. Frantz, Amy E. Spooner, George K. Ogata, Harold N. Keer, and Steven P. Balk. «Mutation of the Androgen-Receptor Gene in Metastatic Androgen-Independent Prostate Cancer». In: *New England Journal of Medicine* 332 (21 May 1995), pp. 1393–1398. ISSN: 0028-4793. DOI: 10.1056/NEJM199505253322101 (cit. on p. 9).

[35] Toru Yoshida, Hidefumi Kinoshita, Takehiko Segawa, Eijiro Nakamura, Takahiro Inoue, Yousuke Shimizu, Toshiyuki Kamoto, and Osamu Ogawa. «Antiandrogen Bicalutamide Promotes Tumor Growth in a Novel Androgen-Dependent Prostate Cancer Xenograft Model Derived from a Bicalutamide-Treated Patient». In: *Cancer Research* 65 (21 Nov. 2005), pp. 9611–9616. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-05-0817 (cit. on p. 9).

[36] Nada Lallous et al. «Functional analysis of androgen receptor mutations that confer anti-androgen resistance identified in circulating cell-free DNA from prostate cancer patients». In: *Genome Biology* 17 (1 Dec. 2016), p. 10. ISSN: 1474-760X. DOI: 10.1186/s13059-015-0864-1 (cit. on p. 9).

[37] Massimo Ammirante, Jun-Li Luo, Sergei Grivennikov, Sergei Nedospasov, and Michael Karin. «B-cell-derived lymphotoxin promotes castration-resistant prostate cancer». In: *Nature* 464 (7286 Mar. 2010), pp. 302–305. ISSN: 0028-0836. DOI: 10.1038/nature08782 (cit. on p. 9).

[38] Sunita R Setlur and Mark A Rubin. «Current Thoughts on the Role of the Androgen Receptor and Prostate Cancer Progression». In: *Advances in Anatomic Pathology* 12 (5 Sept. 2005), pp. 265–270. ISSN: 1072-4109. DOI: 10.1097/01.pap.0000184179.50672.f7 (cit. on p. 9).

[39] Vivek K. Arora et al. «Glucocorticoid Receptor Confers Resistance to Antiandrogens by Bypassing Androgen Receptor Blockade». In: *Cell* 155 (6 Dec. 2013), pp. 1309–1322. ISSN: 00928674. DOI: 10.1016/j.cell.2013.11.012 (cit. on pp. 9, 10).

[40] Shilpa Gurung, Katherine Po Sin Chung, Terence Kin-Wah Lee, et al. «Emerging role of fatty acid binding proteins in cancer pathogenesis». In: *Histology and histopathology* 34.1 (2019), pp. 1–12 (cit. on pp. 10, 13–16).

[41] Norbert H Haunerland and Friedrich Spener. «Properties and physiological significance of fatty acid binding proteins». In: 2003, pp. 99–122. DOI: 10.1016/S1569-2558(03)33007-3 (cit. on p. 10).

[42] Agata Chmurzyńska. «The multigene family of fatty acid-binding proteins (FABPs): Function, structure and polymorphism». In: *Journal of Applied Genetics* 47 (1 Mar. 2006), pp. 39–48. ISSN: 1234-1983. DOI: 10.1007/BF03194597 (cit. on p. 11).

[43] Anna Marie C. Marcelino, Robert G. Smock, and Lila M. Gierasch. «Evolutionary coupling of structural and functional sequence information in the intracellular lipid-binding protein family». In: *Proteins: Structure, Function, and Bioinformatics* 63 (2 May 2006), pp. 373–384. ISSN: 0887-3585. DOI: 10.1002/prot.20860 (cit. on p. 11).

[44] James C. Sacchettini, Jeffrey I. Gordon, and Leonard J. Banaszak. «Crystal structure of rat intestinal fatty-acid-binding protein». In: *Journal of Molecular Biology* 208 (2 July 1989), pp. 327–339. ISSN: 00222836. DOI: 10.1016/0022-2836(89)90392-6 (cit. on p. 11).

[45] Yan He, Xiaomin Yang, Hsin Wang, Rima Estephan, Fouad Francis, Sarala Kodukula, Judith Storch, and Ruth E. Stark. «Solution-State Molecular Structure of Apo and Oleate-Liganded Liver Fatty Acid-Binding Protein». In: *Biochemistry* 46 (44 Nov. 2007), pp. 12543–12556. ISSN: 0006-2960. DOI: 10.1021/bi701092r (cit. on p. 12).

[46] James Thompson, Nate Winter, Daniel Terwey, Judy Bratt, and Leonard Banaszak. «The Crystal Structure of the Liver Fatty Acid-binding Protein». In: *Journal of Biological Chemistry* 272 (11 Mar. 1997), pp. 7140–7150. ISSN: 00219258. DOI: 10.1074/jbc.272.11.7140 (cit. on p. 12).

[47] Gary V. Richieri, Ronald T. Ogata, and Alan M. Kleinfeld. «Fatty acid interactions with native and mutant fatty acid binding proteins». In: *Molecular and Cellular Biochemistry* 192 (1-2 Feb. 1999), pp. 77–85. ISSN: 0300-8177. DOI: 10.1023/A:1006826405151 (cit. on p. 12).

[48] Judith Storch and Alfred E.A Thumser. «The fatty acid transport function of fatty acid-binding proteins». In: *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1486 (1 June 2000), pp. 28–44. ISSN: 13881981. DOI: 10.1016/S1388-1981(00)00046-9 (cit. on p. 12).

[49] Vince J. LiCata and David A. Bernlohr. «Surface properties of adipocyte lipid-binding protein: Response to lipid binding, and comparison with homologous proteins». In: *Proteins: Structure, Function, and Genetics* 33 (4 Dec. 1998), pp. 577–589. ISSN: 08873585. DOI: 10.1002/(SICI)1097-0134(19981201)33:4<577::AID-PROT10>3.0.CO;2-2 (cit. on p. 12).

[50] Cassandra Arroyo-Johnson and Krista D. Mincey. «Obesity Epidemiology Worldwide». In: *Gastroenterology Clinics of North America* 45 (4 Dec. 2016), pp. 571–579. ISSN: 08898553. DOI: 10.1016/j.gtc.2016.07.012 (cit. on p. 12).

[51] Eugenia E. Calle, Carmen Rodriguez, Kimberly Walker-Thurmond, and Michael J. Thun. «Overweight, Obesity, and Mortality from Cancer in a Prospectively Studied Cohort of U.S. Adults». In: *New England Journal of Medicine* 348 (17 Apr. 2003), pp. 1625–1638. ISSN: 0028-4793. DOI: 10.1056/NEJMoa021423 (cit. on p. 12).

[52] Florian Röhrig and Almut Schulze. «The multifaceted roles of fatty acid synthesis in cancer». In: *Nature Reviews Cancer* 16.11 (Nov. 2016), pp. 732–749. ISSN: 1474-175X. DOI: 10.1038/nrc.2016.89 (cit. on pp. 12, 13).

[53] Otto Warburg. «Über den Stoffwechsel der Carcinomzelle». In: *Die Naturwissenschaften* 12 (50 Dec. 1924), pp. 1131–1137. ISSN: 0028-1042. DOI: 10.1007/BF01504608 (cit. on p. 12).

[54] Grace Medes, Alice Thomas, and Sidney Weinhouse. «Metabolism of Neoplastic Tissue. IV. A Study of Lipid Synthesis in Neoplastic Tissue Slices in Vitro*». In: *Cancer Research* 13.1 (Jan. 1953), pp. 27–29. ISSN: 0008-5472. eprint: https://aacrjournals.org/cancerres/article-pdf/13/1/27/2370645/cr0130010027.pdf (cit. on p. 12).

[55] F P Kuhajda, K Jenner, F D Wood, R A Hennigar, L B Jacobs, J D Dick, and G R Pasternack. «Fatty acid synthesis: a potential selective target for antineoplastic therapy.» In: *Proceedings of the National Academy of Sciences* 91.14 (July 1994), pp. 6379–6383. ISSN: 0027-8424. DOI: 10.1073/pnas.91.14.6379 (cit. on p. 13).

[56] M. Ookhtens, R. Kannan, I. Lyon, and N. Baker. «Liver and adipose tissue contributions to newly formed fatty acids in an ascites tumor». In: *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 247 (1 July 1984), R146–R153. ISSN: 0363-6119. DOI: 10.1152/ajpregu.1984.247.1.R146 (cit. on p. 13).

[57] Yanyan Cai et al. «Loss of Chromosome 8p Governs Tumor Progression and Drug Response by Altering Lipid Metabolism». In: *Cancer Cell* 29.5 (May 2016), pp. 751–766. ISSN: 15356108. DOI: 10.1016/j.ccell.2016.04.003 (cit. on p. 13).

[58] A JAKOBSSON, R WESTERBERG, and A JACOBSSON. «Fatty acid elongases in mammals: Their regulation and roles in metabolism». In: *Progress in Lipid Research* 45 (3 May 2006), pp. 237–249. ISSN: 01637827. DOI: 10.1016/j.plipres.2006.01.004 (cit. on p. 13).

[59] R. A. Igal. «Stearoyl-CoA desaturase-1: a novel key player in the mechanisms of cell proliferation, programmed cell death and transformation to cancer». In: *Carcinogenesis* 31 (9 Sept. 2010), pp. 1509–1515. ISSN: 0143-3334. DOI: 10.1093/carcin/bgq131 (cit. on p. 13).

[60] Hironobu Yasui, Shingo Matsumoto, Nallathamby Devasahayam, Jeeva P Munasinghe, Rajani Choudhuri, Keita Saito, Sankaran Subramanian, James B Mitchell, and Murali C Krishna. «Low-field magnetic resonance imaging to visualize chronic and cycling hypoxia in tumor-bearing mice.» In: *Cancer research* 70.16 (Aug. 2010), pp. 6427–36. ISSN: 1538-7445. DOI: 10.1158/0008-5472.CAN-10-1350 (cit. on p. 13).

[61] Karim Bensaad et al. «Fatty Acid Uptake and Lipid Storage Induced by HIF-1α Contribute to Cell Growth and Survival after Hypoxia-Reoxygenation». In: *Cell Reports* 9.1 (Oct. 2014), pp. 349–365. ISSN: 22111247. DOI: 10.1016/j.celrep.2014.08.056 (cit. on p. 13).

[62] Carine Michiels, Céline Tellier, and Olivier Feron. «Cycling hypoxia: A key feature of the tumor microenvironment». In: *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1866.1 (Aug. 2016), pp. 76–86. ISSN: 0304419X. DOI: 10.1016/j.bbcan.2016.06.004 (cit. on p. 13).

[63] Arnout Schepers and Hans Clevers. «Wnt signaling, stem cells, and cancer of the gastrointestinal tract.» In: *Cold Spring Harbor perspectives in biology* 4 (4 Apr. 2012), a007989. ISSN: 1943-0264. DOI: 10.1101/cshperspect.a007989 (cit. on p. 13).

[64] Jong Bae Park, Chang Sup Lee, Jin-Hyeok Jang, Jaewang Ghim, Youn-Jae Kim, Sungyoung You, Daehee Hwang, Pann-Ghill Suh, and Sung Ho Ryu. «Phospholipase signalling networks in cancer». In: *Nature Reviews Cancer* 12 (11 Nov. 2012), pp. 782–792. ISSN: 1474-175X. DOI: 10.1038/nrc3379 (cit. on p. 13).

[65] Chun Jing, Carol Beesley, Christopher S. Foster, Philip S. Rudland, Hiroshi Fujii, Terno Ono, Haijuan Chen, Paul H. Smith, and Youqiang Ke. «Identification of the Messenger RNA for Human Cutaneous Fatty Acid-binding Protein as a Metastasis Inducer1». In: *Cancer Research* 60.9 (May 2000), pp. 2390–2398. ISSN: 0008-5472. eprint: https://aacrjournals.org/cancerres/article-pdf/60/9/2390/3240456/ch090002390p.pdf (cit. on p. 13).

[66] Anping Chen et al. «Liver fatty acid binding protein (L-Fabp) modulates murine stellate cell activation and diet-induced nonalcoholic fatty liver disease». In: *Hepatology* 57 (6 June 2013), pp. 2202–2212. ISSN: 0270-9139. DOI: 10.1002/hep.26318 (cit. on p. 14).

[67] Stephanie M Wood, Anthony J Gill, Alexander S Brodsky, Shaolei Lu, Kenneth Friedman, Galina Karashchuk, Kara Lombardo, Dongfang Yang, and Murray B Resnick. «Fatty acid-binding protein 1 is preferentially lost in microsatellite instable colorectal carcinomas and is immune modulated via

the interferon pathway». In: *Modern Pathology* 30 (1 Jan. 2017), pp. 123–133. ISSN: 08933952. DOI: 10.1038/modpathol.2016.170 (cit. on pp. 14, 57).

[68] Chung-Yu Ku, Yu-Huei Liu, Hsuan-Yuan Lin, Shao-Chun Lu, and Jung-Yaw Lin. «Liver fatty acid-binding protein (L-FABP) promotes cellular angiogenesis and migration in hepatocellular carcinoma». In: *Oncotarget* 7 (14 Apr. 2016), pp. 18229–18246. ISSN: 1949-2553. DOI: 10.18632/oncotarget.7571 (cit. on p. 14).

[69] Norbert H Haunerland and Friedrich Spener. «Fatty acid-binding proteins – insights from genetic manipulations». In: *Progress in Lipid Research* 43 (4 July 2004), pp. 328–349. ISSN: 01637827. DOI: 10.1016/j.plipres.2004.05.001 (cit. on p. 14).

[70] L J Baier et al. «An amino acid substitution in the human intestinal fatty acid binding protein is associated with increased fatty acid binding, increased fat oxidation, and insulin resistance.» In: *Journal of Clinical Investigation* 95 (3 Mar. 1995), pp. 1281–1287. ISSN: 0021-9738. DOI: 10.1172/JCI117778 (cit. on p. 14).

[71] Emile Levy, Daniel Ménard, Edgard Delvin, Simona Stan, Grant Mitchell, Marie Lambert, Ehud Ziv, Juan Carlos Feoli-Fonseca, and Ernest Seidman. «The Polymorphism at Codon 54 of the FABP2 Gene Increases Fat Absorption in Human Intestinal Explants». In: *Journal of Biological Chemistry* 276 (43 Oct. 2001), pp. 39679–39684. ISSN: 00219258. DOI: 10.1074/jbc.M105713200 (cit. on p. 14).

[72] Ikuko Kato, Susan Land, Adhip P. Majumdar, Jill Barnholtz-Sloan, and Richard K. Severson. «Functional polymorphisms to modulate luminal lipid exposure and risk of colorectal cancer». In: *Cancer Epidemiology* 34 (3 June 2010), pp. 291–297. ISSN: 18777821. DOI: 10.1016/j.canep.2010.02.010 (cit. on pp. 14, 57).

[73] Wenjie Li, Shaoran Zhang, Siwei Zhou, Lingling Jiang, and Wei Wang. «Cardiac Fatty Acid Binding Protein (FABP3) Depletes SR Calcium Load in Ventricular Myocytes». In: *Biophysical Journal* 112 (3 Feb. 2017), 424a. ISSN: 00063495. DOI: 10.1016/j.bpj.2016.11.2266 (cit. on pp. 14, 15).

[74] M. S. Cheon, S. H. Kim, M. Fountoulakis, and G. Lubec. «Heart type fatty acid binding protein (H-FABP) is decreased in brains of patients with Down syndrome and Alzheimer's disease». In: 2003, pp. 225–234. DOI: 10.1007/978-3-7091-6721-2_20 (cit. on pp. 14, 15).

[75] Takeaki Hashimoto et al. «Expression of Heart-Type Fatty Acid-Binding Protein in Human Gastric Carcinoma and Its Association with Tumor Aggressiveness, Metastasis and Poor Prognosis». In: *Pathobiology* 71.5 (2004), pp. 267–273. ISSN: 1015-2008. DOI: 10.1159/000080061 (cit. on pp. 14, 15).

[76] Zhiyuan Tang et al. «Elevated expression of FABP3 and FABP4 cooperatively correlates with poor prognosis in non-small cell lung cancer (NSCLC).» In: *Oncotarget* 7.29 (July 2016), pp. 46253–46262. ISSN: 1949-2553. DOI: 10.18632/oncotarget.10086 (cit. on pp. 14, 15).

[77] Ben Davidson, Vera Maria Abeler, Ellen Hellesylt, Arild Holth, Ie-Ming Shih, Tone Skeie-Jensen, Li Chen, Yanqin Yang, and Tian-Li Wang. «Gene expression signatures differentiate uterine endometrial stromal sarcoma from leiomyosarcoma». In: *Gynecologic Oncology* 128 (2 Feb. 2013), pp. 349–355. ISSN: 00908258. DOI: 10.1016/j.ygyno.2012.11.021 (cit. on p. 15).

[78] Annett Linge, Susan Kennedy, Deirdre O'Flynn, Stephen Beatty, Paul Moriarty, Michael Henry, Martin Clynes, Annemarie Larkin, and Paula Meleady. «Differential Expression of Fourteen Proteins between Uveal Melanoma from Patients Who Subsequently Developed Distant Metastases versus Those Who Did Not». In: *Investigative Opthalmology & Visual Science* 53 (8 July 2012), p. 4634. ISSN: 1552-5783. DOI: 10.1167/iovs.11-9019 (cit. on p. 15).

[79] Hung Huynh, Lesley Alpert, and Michael Pollak. «Silencing of the Mammary-derived Growth Inhibitor (MDGI) Gene in Breast Neoplasms Is Associated with Epigenetic Changes1». In: *Cancer Research* 56.21 (Nov. 1996), pp. 4865–4870. ISSN: 0008-5472. eprint: https://aacrjournals.org/cancerres/article-pdf/56/21/4865/2462108/cr0560214865.pdf (cit. on p. 15).

[80] Jonna Nevo et al. «Mammary-Derived Growth Inhibitor Alters Traffic of EGFR and Induces a Novel Form of Cetuximab Resistance». In: *Clinical Cancer Research* 15 (21 Nov. 2009), pp. 6570–6581. ISSN: 1078-0432. DOI: 10.1158/1078-0432.CCR-09-0773 (cit. on p. 15).

[81] Haiming Cao et al. «Adipocyte Lipid Chaperone aP2 Is a Secreted Adipokine Regulating Hepatic Glucose Production». In: *Cell Metabolism* 17 (5 May 2013), pp. 768–778. ISSN: 15504131. DOI: 10.1016/j.cmet.2013.04.012 (cit. on p. 15).

[82] Valéria Lamounier-Zepter, Christiane Look, Julio Alvarez, Torsten Christ, Ursula Ravens, Wolf-Hagen Schunck, Monika Ehrhart-Bornstein, Stefan R. Bornstein, and Ingo Morano. «Adipocyte Fatty Acid–Binding Protein Suppresses Cardiomyocyte Contraction». In: *Circulation Research* 105 (4 Aug. 2009), pp. 326–334. ISSN: 0009-7330. DOI: 10.1161/CIRCRESAHA.109.200501 (cit. on p. 15).

[83] Mireille Vasseur-Cognet and M. Daniel Lane. «Trans-acting factors involved in adipogenic differentiation». In: *Current Opinion in Genetics & Development* 3 (2 1993), pp. 238–245. ISSN: 0959437X. DOI: 10.1016/0959-437X(93)90029-O (cit. on p. 15).

[84] Liza Makowski et al. «Lack of macrophage fatty-acid–binding protein aP2 protects mice deficient in apolipoprotein E against atherosclerosis». In: *Nature Medicine* 7 (6 June 2001), pp. 699–705. ISSN: 1078-8956. DOI: 10.1038/89076 (cit. on p. 15).

[85] K. Hancke, D. Grubeck, N. Hauser, R. Kreienberg, and J. M. Weiss. «Adipocyte fatty acid-binding protein as a novel prognostic factor in obese breast cancer patients». In: *Breast Cancer Research and Treatment* 119.2 (Jan. 2010), pp. 367–377. ISSN: 0167-6806. DOI: 10.1007/s10549-009-0577-9 (cit. on p. 15).

[86] Hisanori Uehara, Tetsuyuki Takahashi, Mina Oha, Hirohisa Ogawa, and Keisuke Izumi. «Exogenous fatty acid binding protein 4 promotes human prostate cancer cell progression». In: *International Journal of Cancer* 135.11 (Dec. 2014), pp. 2558–2568. ISSN: 0020-7136. DOI: 10.1002/ijc.28903 (cit. on p. 15).

[87] Sandra Guaita-Esteruelas, Alba Bosquet, Paula Saavedra, Josep Gumà, Josefa Girona, Eric W.-F. Lam, Kepa Amillano, Joan Borràs, and Lluís Masana. «Exogenous FABP4 increases breast cancer cell proliferation and activates the expression of fatty acid transport proteins». In: *Molecular Carcinogenesis* 56.1 (Jan. 2017), pp. 208–217. ISSN: 08991987. DOI: 10.1002/mc.22485 (cit. on p. 15).

[88] Jihua Nie, Jingying Zhang, Lili Wang, Lunjie Lu, Qian Yuan, Fangmei An, Shuyu Zhang, and Yang Jiao. «Adipocytes promote cholangiocarcinoma metastasis through fatty acid binding protein 4». In: *Journal of Experimental & Clinical Cancer Research* 36.1 (Dec. 2017), p. 183. ISSN: 1756-9966. DOI: 10.1186/s13046-017-0641-y (cit. on pp. 15, 57).

[89] O. Cataltepe, M. C. Arikan, E. Ghelfi, C. Karaaslan, Y. Ozsurekci, K. Dresser, Y. Li, T. W. Smith, and S. Cataltepe. «Fatty acid binding protein 4 is expressed in distinct endothelial and non-endothelial cell populations in glioblastoma». In: *Neuropathology and Applied Neurobiology* 38.5 (Aug. 2012), pp. 400–410. ISSN: 0305-1846. DOI: 10.1111/j.1365-2990.2011.01237.x (cit. on p. 15).

[90] F Yan et al. «Fatty acid-binding protein FABP4 mechanistically links obesity with aggressive AML by enhancing aberrant DNA methylation in AML cells». In: *Leukemia* 31.6 (June 2017), pp. 1434–1442. ISSN: 0887-6924. DOI: 10.1038/leu.2016.349 (cit. on p. 15).

[91] Jiangbo Jin, Ziyu Zhang, Song Zhang, Xinyu Chen, Zhen Chen, Ping Hu, Jianbin Wang, and Caifeng Xie. «Fatty acid binding protein 4 promotes epithelial-mesenchymal transition in cervical squamous cell carcinoma

through AKT/GSK3β/Snail signaling pathway». In: *Molecular and Cellular Endocrinology* 461 (Feb. 2018), pp. 155–164. ISSN: 03037207. DOI: 10.1016/j.mce.2017.09.005 (cit. on p. 15).

[92] Denis Khnykin, Jeffrey H. Miner, and Frode Jahnsen. «Role of fatty acid transporters in epidermis». In: *Dermato-Endocrinology* 3 (2 Apr. 2011), pp. 53–61. ISSN: 1938-1980. DOI: 10.4161/derm.3.2.14816 (cit. on p. 15).

[93] Yoshiyuki Kusakari et al. «Decreased keratinocyte motility in skin wound on mice lacking the epidermal fatty acid binding protein gene». In: *Molecular and Cellular Biochemistry* 284 (1-2 May 2006), pp. 183–188. ISSN: 0300-8177. DOI: 10.1007/s11010-005-9048-8 (cit. on p. 15).

[94] Shuiliang Yu, Liraz Levi, Ruth Siegel, and Noa Noy. «Retinoic Acid Induces Neurogenesis by Activating Both Retinoic Acid Receptors (RARs) and Peroxisome Proliferator-activated Receptor / (PPAR/)». In: *Journal of Biological Chemistry* 287 (50 Dec. 2012), pp. 42195–42205. ISSN: 00219258. DOI: 10.1074/jbc.M112.410381 (cit. on p. 15).

[95] Liraz Levi, Glenn Lobo, Mary Kathryn Doud, Johannes von Lintig, Darcie Seachrist, Gregory P. Tochtrop, and Noa Noy. «Genetic Ablation of the Fatty Acid–Binding Protein FABP5 Suppresses HER2-Induced Mammary Tumorigenesis». In: *Cancer Research* 73 (15 Aug. 2013), pp. 4770–4780. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-13-0384 (cit. on p. 16).

[96] Catherine A Powell, Mohd W Nasser, Helong Zhao, Jacob C Wochna, Xiaoli Zhang, Charles Shapiro, Konstantin Shilo, and Ramesh K Ganju. «Fatty acid binding protein 5 promotes metastatic potential of triple negative breast cancer cells through enhancing epidermal growth factor receptor stability.» In: *Oncotarget* 6 (8 Mar. 2015), pp. 6373–85. ISSN: 1949-2553. DOI: 10.18632/oncotarget.3442 (cit. on p. 16).

[97] Waseem Al-Jameel et al. «Inhibitor SBFI26 suppresses the malignant progression of castration-resistant PC3-M cells by competitively binding to oncogenic FABP5.» In: *Oncotarget* 8 (19 May 2017), pp. 31041–31056. ISSN: 1949-2553. DOI: 10.18632/oncotarget.16055 (cit. on pp. 16, 57).

[98] Luis B. Agellon, Laurie Drozdowski, Lena Li, Claudiu Iordache, Le Luong, M. Tom Clandinin, Richard R.E. Uwiera, Matthew J. Toth, and Alan B.R. Thomson. «Loss of intestinal fatty acid binding protein increases the susceptibility of male mice to high fat diet-induced fatty liver». In: *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1771 (10 Oct. 2007), pp. 1283–1288. ISSN: 13881981. DOI: 10.1016/j.bbalip.2007.08.004 (cit. on p. 16).

[99] Takahiro Ohmachi, Hiroshi Inoue, Koshi Mimori, Fumiaki Tanaka, Atsushi Sasaki, Tatsuo Kanda, Hiroshi Fujii, Katsuhiko Yanaga, and Masaki Mori. «Fatty Acid Binding Protein 6 Is Overexpressed in Colorectal Cancer». In: *Clinical Cancer Research* 12 (17 Sept. 2006), pp. 5090–5095. ISSN: 1078-0432. DOI: 10.1158/1078-0432.CCR-05-2045 (cit. on pp. 16, 57).

[100] Rong-Zong Liu, Raja Mita, Michael Beaulieu, Zhihua Gao, and Roseline Godbout. «Fatty acid binding proteins in brain development and disease». In: *The International Journal of Developmental Biology* 54 (8-9 2010), pp. 1229–1239. ISSN: 0214-6282. DOI: 10.1387/ijdb.092976rl (cit. on p. 16).

[101] Shyra J. Miller et al. «Brain Lipid Binding Protein in Axon-Schwann Cell Interactions and Peripheral Nerve Tumorigenesis». In: *Molecular and Cellular Biology* 23 (6 Mar. 2003), pp. 2213–2224. ISSN: 1098-5549. DOI: 10.1128/MCB.23.6.2213-2224.2003 (cit. on p. 16).

[102] Mian Xie, Xiaojun Wu, Jinjun Zhang, Chaosheng He, Shenhai Wei, Junyao Huang, Xinge Fu, and Yingying Gu. «The Prognostic Significance of Notch1 and Fatty Acid Binding Protein 7 (FABP7) Expression in Resected Tracheobronchial Adenoid Cystic Carcinoma: A Multicenter Retrospective Study». In: *Cancer Research and Treatment* 50 (4 Oct. 2018), pp. 1064–1073. ISSN: 1598-2998. DOI: 10.4143/crt.2017.337 (cit. on pp. 16, 57).

[103] Rong-Zong Liu, Kathryn Graham, Darryl D Glubrecht, Raymond Lai, John R Mackey, and Roseline Godbout. «A fatty acid-binding protein 7/RXR pathway enhances survival and proliferation in triple-negative breast cancer». In: *The Journal of Pathology* 228 (3 Nov. 2012), pp. 310–321. ISSN: 0022-3417. DOI: 10.1002/path.4001 (cit. on pp. 16, 57).

[104] Alaa T. Alshareeda, Emad A. Rakha, Christopher C. Nolan, Ian O. Ellis, and Andrew R. Green. «Fatty acid binding protein 7 expression and its sub-cellular localization in breast cancer». In: *Breast Cancer Research and Treatment* 134 (2 July 2012), pp. 519–529. ISSN: 0167-6806. DOI: 10.1007/s10549-012-2083-8 (cit. on p. 16).

[105] Tatsuo Kido, Satoru Arata, Ryusuke Suzuki, Tomohiko Hosono, Yoshinobu Nakanishi, Jun-ichi Miyazaki, Izumu Saito, Toshio Kuroki, and Seiji Shioda. «The testicular fatty acid binding protein PERF15 regulates the fate of germ cells in PERF15 transgenic mice». In: *Development, Growth & Differentiation* 47 (1 Jan. 2005), pp. 15–24. ISSN: 0012-1592. DOI: 10.1111/j.1440-169x.2004.00775.x (cit. on p. 16).

[106] Vimal Selvaraj, Atsushi Asano, Jennifer L. Page, Jacquelyn L. Nelson, Kumar S.D. Kothapalli, James A. Foster, J. Thomas Brenna, Robert S. Weiss, and Alexander J. Travis. «Mice lacking FABP9/PERF15 develop sperm head abnormalities but are fertile». In: *Developmental Biology* 348 (2 Dec. 2010),

pp. 177–189. ISSN: 00121606. DOI: 10.1016/j.ydbio.2010.09.019 (cit. on p. 16).

[107] Majed Saad Al Fayi, Xiaojun Gou, Shiva S Forootan, Waseem Al-Jameel, Zhengzheng Bao, Philip R Rudland, Philip A Cornford, Syed A Hussain, and Youqiang Ke. «The increased expression of fatty acid-binding protein 9 in prostate cancer and its prognostic significance.» In: *Oncotarget* 7.50 (Dec. 2016), pp. 82783–82797. ISSN: 1949-2553. DOI: 10.18632/oncotarget.12635 (cit. on p. 17).

[108] H. M. Berman. «The Protein Data Bank». In: *Nucleic Acids Research* 28.1 (Jan. 2000), pp. 235–242. ISSN: 13624962. DOI: 10.1093/nar/28.1.235 (cit. on pp. 17, 20, 27).

[109] Rong-Zong Liu, Xiaodong Li, and Roseline Godbout. «A novel fatty acid-binding protein (FABP) gene resulting from tandem gene duplication in mammals: transcription in rat retina and testis». In: *Genomics* 92 (6 Dec. 2008), pp. 436–445. ISSN: 08887543. DOI: 10.1016/j.ygeno.2008.08.003 (cit. on pp. 17, 18).

[110] Alex Bateman et al. «UniProt: the Universal Protein Knowledgebase in 2025». In: *Nucleic Acids Research* 53.D1 (Jan. 2025), pp. D609–D617. ISSN: 0305-1048. DOI: 10.1093/nar/gkae1010 (cit. on pp. 18, 20, 27, 28, 30, 63).

[111] Mark A. Rubin et al. «Overexpression, Amplification, and Androgen Regulation of TPD52 in Prostate Cancer». In: *Cancer Research* 64 (11 June 2004), pp. 3814–3822. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-03-3881 (cit. on pp. 18, 19).

[112] Kati Porkka, Outi Saramäki, Minna Tanner, and Tapio Visakorpi. «Amplification and Overexpression of Elongin C Gene Discovered in Prostate Cancer by cDNA Microarrays». In: *Laboratory Investigation* 82 (5 May 2002), pp. 629–637. ISSN: 00236837. DOI: 10.1038/labinvest.3780457 (cit. on p. 18).

[113] Mark van Duin et al. «High-resolution array comparative genomic hybridization of chromosome arm 8q: Evaluation of genetic progression markers for prostate cancer». In: *Genes, Chromosomes and Cancer* 44 (4 Dec. 2005), pp. 438–449. ISSN: 1045-2257. DOI: 10.1002/gcc.20259 (cit. on pp. 18, 19).

[114] M L Cher, G S Bova, D H Moore, E J Small, P R Carroll, S S Pin, J I Epstein, W B Isaacs, and R H Jensen. «Genetic alterations in untreated metastases and androgen-independent prostate cancer detected by comparative genomic hybridization and allelotyping.» In: *Cancer research* 56.13 (July 1996), pp. 3091–102. ISSN: 0008-5472 (cit. on p. 18).

[115] N N Nupponen, L Kakkola, P Koivisto, and T Visakorpi. «Genetic alterations in hormone-refractory recurrent prostate carcinomas.» In: *The American journal of pathology* 153.1 (July 1998), pp. 141–8. ISSN: 0002-9440. DOI: 10.1016/S0002-9440(10)65554-X (cit. on p. 18).

[116] Janneke C Alers, Jenneke Rochat, Pieter-Jaap Krijtenburg, Wim C J Hop, Ries Kranse, Carla Rosenberg, Hans J Tanke, Fritz H Schröder, and Herman van Dekken. «Identification of Genetic Markers for Prostatic Cancer Progression». In: *Laboratory Investigation* 80.6 (June 2000), pp. 931–942. ISSN: 00236837. DOI: 10.1038/labinvest.3780096 (cit. on p. 18).

[117] Kati Porkka, Outi Saramäki, Minna Tanner, and Tapio Visakorpi. «Amplification and Overexpression of Elongin C Gene Discovered in Prostate Cancer by cDNA Microarrays». In: *Laboratory Investigation* 82.5 (May 2002), pp. 629–637. ISSN: 00236837. DOI: 10.1038/labinvest.3780457 (cit. on p. 18).

[118] Rong-Zong Liu et al. «The FABP12/PPAR pathway promotes metastatic transformation by inducing epithelial-to-mesenchymal transition and lipid-derived energy production in prostate cancer cells». In: *Molecular Oncology* 14 (12 Dec. 2020), pp. 3100–3120. ISSN: 1574-7891. DOI: 10.1002/1878-0261.12818 (cit. on pp. 18, 19, 57).

[119] Douglas Hanahan and Robert A. Weinberg. «Hallmarks of Cancer: The Next Generation». In: *Cell* 144 (5 Mar. 2011), pp. 646–674. ISSN: 00928674. DOI: 10.1016/j.cell.2011.02.013 (cit. on p. 19).

[120] Micaela Montanari et al. «Epithelial-mesenchymal transition in prostate cancer: an overview.» In: *Oncotarget* 8 (21 May 2017), pp. 35376–35389. ISSN: 1949-2553. DOI: 10.18632/oncotarget.15686 (cit. on p. 19).

[121] Imran Ahmad et al. «Sleeping Beauty screen reveals Pparg activation in metastatic prostate cancer». In: *Proceedings of the National Academy of Sciences* 113 (29 July 2016), pp. 8290–8295. ISSN: 0027-8424. DOI: 10.1073/pnas.1601571113 (cit. on p. 19).

[122] Barry S. Taylor et al. «Integrative Genomic Profiling of Human Prostate Cancer». In: *Cancer Cell* 18 (1 July 2010), pp. 11–22. ISSN: 15356108. DOI: 10.1016/j.ccr.2010.05.026 (cit. on p. 19).

[123] William George Warren, Myles Osborn, Andy Yates, Karen Wright, and Saoirse E. O'Sullivan. «The emerging role of fatty acid binding protein 5 (FABP5) in cancers». In: *Drug Discovery Today* 28 (7 July 2023), p. 103628. ISSN: 13596446. DOI: 10.1016/j.drudis.2023.103628 (cit. on p. 19).

[124] Saoirse Elizabeth O'Sullivan and Martin Kaczocha. «FABP5 as a novel molecular target in prostate cancer». In: *Drug Discovery Today* 25 (11 Nov. 2020), pp. 2056–2061. ISSN: 13596446. DOI: 10.1016/j.drudis.2020.09.018 (cit. on p. 19).

[125] Furkan Ayberk Binbay, Dhruv Chetanbhai Rathod, Ajay Abisheck Paul George, and Diana Imhof. «Quality Assessment of Selected Protein Structures Derived from Homology Modeling and AlphaFold». In: *Pharmaceuticals* 16 (12 Nov. 2023), p. 1662. ISSN: 1424-8247. DOI: 10.3390/ph16121662 (cit. on p. 19).

[126] Xiao-chen Bai, Tamir Gonen, Angela M. Gronenborn, Anastassis Perrakis, Andrea Thorn, and Jianyi Yang. «Challenges and opportunities in macromolecular structure determination». In: *Nature Reviews Molecular Cell Biology* 25 (1 Jan. 2024), pp. 7–12. ISSN: 1471-0072. DOI: 10.1038/s41580-023-00659-y (cit. on pp. 19, 20).

[127] W.J. Browne, A.C.T. North, D.C. Phillips, Keith Brew, Thomas C. Vanaman, and Robert L. Hill. «A possible three-dimensional structure of bovine -lactalbumin based on that of hen's egg-white lysozyme». In: *Journal of Molecular Biology* 42 (1 May 1969), pp. 65–86. ISSN: 00222836. DOI: 10.1016/0022-2836(69)90487-2 (cit. on p. 20).

[128] Jens Carlsson, Ryan G Coleman, Vincent Setola, John J Irwin, Hao Fan, Avner Schlessinger, Andrej Sali, Bryan L Roth, and Brian K Shoichet. «Ligand discovery from a dopamine D3 receptor homology model and crystal structure». In: *Nature Chemical Biology* 7 (11 Nov. 2011), pp. 769–778. ISSN: 1552-4450. DOI: 10.1038/nchembio.662 (cit. on p. 20).

[129] Thijs Beuming and Woody Sherman. «Current Assessment of Docking into GPCR Crystal Structures and Homology Models: Successes, Challenges, and Guidelines». In: *Journal of Chemical Information and Modeling* 52 (12 Dec. 2012), pp. 3263–3277. ISSN: 1549-9596. DOI: 10.1021/ci300411b (cit. on p. 20).

[130] Amy O. Stevens and Yi He. «Benchmarking the Accuracy of AlphaFold 2 in Loop Structure Prediction». In: *Biomolecules* 12 (7 July 2022), p. 985. ISSN: 2218-273X. DOI: 10.3390/biom12070985 (cit. on p. 20).

[131] Sarfaraz K. Niazi and Zamara Mariam. «Computer-Aided Drug Design and Drug Discovery: A Prospective Analysis». In: *Pharmaceuticals* 17.1 (Dec. 2023), p. 22. ISSN: 1424-8247. DOI: 10.3390/ph17010022 (cit. on p. 21).

[132] Nalini Schaduangrat, Samuel Lampa, Saw Simeon, Matthew Paul Gleeson, Ola Spjuth, and Chanin Nantasenamat. «Towards reproducible computational drug discovery». In: *Journal of Cheminformatics* 12.1 (Dec. 2020), p. 9. ISSN: 1758-2946. DOI: 10.1186/s13321-020-0408-x (cit. on p. 21).

[133] Changzhen Fu et al. «Molecular Drug Simulation and Experimental Validation of the CD36 Receptor Competitively Binding to Long-Chain Fatty Acids by 7-Ketocholesteryl-9-carboxynonanoate». In: *ACS Omega* 8.31 (Aug. 2023), pp. 28277–28289. ISSN: 2470-1343. DOI: 10.1021/acsomega.3c02082 (cit. on p. 21).

[134] John J. Irwin, Khanh G. Tang, Jennifer Young, Chinzorig Dandarchuluun, Benjamin R. Wong, Munkhzul Khurelbaatar, Yurii S. Moroz, John Mayfield, and Roger A. Sayle. «ZINC20—A Free Ultralarge-Scale Chemical Database for Ligand Discovery». In: *Journal of Chemical Information and Modeling* 60 (12 Dec. 2020), pp. 6065–6073. ISSN: 1549-9596. DOI: 10.1021/acs.jcim.0c00675 (cit. on pp. 21, 58).

[135] Wenchao Lu, Rukang Zhang, Hao Jiang, Huimin Zhang, and Cheng Luo. «Computer-Aided Drug Design in Epigenetics». In: *Frontiers in Chemistry* 6 (Mar. 2018). ISSN: 2296-2646. DOI: 10.3389/fchem.2018.00057 (cit. on p. 21).

[136] ULC Chemical Computing Group. *Molecular Operating Environment (MOE), 2024.0601 Chemical Computing Group ULC, 910-1010 Sherbrooke St. W., Montreal, QC H3A 2R7, 2025.* June 2024 (cit. on p. 22).

[137] Andrei N Lupas and Robert B Russell. *Structural Bioinformatics*. 2012 (cit. on p. 27).

[138] Sanzo Miyazawa and Robert L. Jernigan. «Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation». In: *Macromolecules* 18.3 (Mar. 1985), pp. 534–552. ISSN: 0024-9297. DOI: 10.1021/ma00145a039 (cit. on p. 30).

[139] Chao Zhang, George Vasmatzis, James L Cornette, and Charles DeLisi. «Determination of atomic desolvation energies from the structures of crystallized proteins». In: *Journal of Molecular Biology* 267.3 (Apr. 1997), pp. 707–726. ISSN: 00222836. DOI: 10.1006/jmbi.1996.0859 (cit. on pp. 30, 32).

[140] Paul Labute. «The generalized Born/volume integral implicit solvent model: Estimation of the free energy of hydration using London dispersion instead of atomic surface area». In: *Journal of Computational Chemistry* 29.10 (July 2008), pp. 1693–1698. ISSN: 0192-8651. DOI: 10.1002/jcc.20933 (cit. on p. 32).

[141] Andriy Kryshtafovych, Torsten Schwede, Maya Topf, Krzysztof Fidelis, and John Moult. «Critical assessment of methods of protein structure prediction (CASP)—Round XIII». In: *Proteins: Structure, Function, and Bioinformatics* 87.12 (Dec. 2019), pp. 1011–1020. ISSN: 0887-3585. DOI: 10.1002/prot.25823 (cit. on p. 33).

[142] Andrew W. Senior et al. «Improved protein structure prediction using potentials from deep learning». In: *Nature* 577.7792 (Jan. 2020), pp. 706–710. ISSN: 0028-0836. DOI: 10.1038/s41586-019-1923-7 (cit. on pp. 33, 34).

[143] John G. Kirkwood. «Statistical Mechanics of Fluid Mixtures». In: *The Journal of Chemical Physics* 3.5 (May 1935), pp. 300–313. ISSN: 0021-9606. DOI: 10.1063/1.1749657 (cit. on p. 33).

[144] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. «Optimization by Simulated Annealing». In: *Science* 220.4598 (May 1983), pp. 671–680. ISSN: 0036-8075. DOI: 10.1126/science.220.4598.671 (cit. on p. 33).

[145] David T. Jones. «Predicting novel protein folds by using FRAGFOLD». In: *Proteins: Structure, Function, and Genetics* 45.S5 (2001), pp. 127–132. ISSN: 0887-3585. DOI: 10.1002/prot.1171 (cit. on p. 33).

[146] Rhiju Das and David Baker. «Macromolecular Modeling with Rosetta». In: *Annual Review of Biochemistry* 77.1 (June 2008), pp. 363–382. ISSN: 0066-4154. DOI: 10.1146/annurev.biochem.77.062906.171838 (cit. on p. 33).

[147] Chengxin Zhang, S. M. Mortuza, Baoji He, Yanting Wang, and Yang Zhang. «Template-based and free modeling of I-TASSER and QUARK pipelines using predicted contact maps in CASP12». In: *Proteins: Structure, Function, and Bioinformatics* 86.S1 (Mar. 2018), pp. 136–151. ISSN: 0887-3585. DOI: 10.1002/prot.25414 (cit. on p. 33).

[148] John Jumper et al. «Highly accurate protein structure prediction with AlphaFold». In: *Nature* 596.7873 (Aug. 2021), pp. 583–589. ISSN: 0028-0836. DOI: 10.1038/s41586-021-03819-2 (cit. on pp. 33, 41, 60–62).

[149] Josh Abramson et al. «Accurate structure prediction of biomolecular interactions with AlphaFold 3». In: *Nature* 630.8016 (June 2024), pp. 493–500. ISSN: 0028-0836. DOI: 10.1038/s41586-024-07487-w (cit. on p. 34).

[150] Google-Deepmind. *AlphaFold3*. Accessed: 2023-10-01. 2024. URL: https://github.com/google-deepmind/alphafold3/blob/main (cit. on pp. 34, 36).

[151] Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. «ColabFold: making protein folding accessible to all». In: *Nature Methods* 19.6 (June 2022), pp. 679–682. ISSN: 1548-7091. DOI: 10.1038/s41592-022-01488-1 (cit. on pp. 34, 35).

[152] Josh Abramson et al. «Accurate structure prediction of biomolecular interactions with AlphaFold 3». In: *Nature* 630.8016 (2024), pp. 493–500. DOI: 10.1038/s41586-024-07487-w (cit. on pp. 36, 39, 43).

[153] Elana Simon Jake Silberg. *The Illustrated Alphafold*. Accessed: 2025-11-01. 2024. URL: https://elanapearl.github.io/blog/2024/the-illustrated-alphafold/#triangle-attention (cit. on pp. 36, 38, 60–62).

[154] Steven A. Benner and Dietlinde Gerloff. «Patterns of divergence in homologous proteins as indicators of secondary and tertiary structure: A prediction of the structure of the catalytic domain of protein kinases». In: *Advances in Enzyme Regulation* 31 (Jan. 1991), pp. 121–181. ISSN: 00652571. DOI: 10.1016/0065-2571(91)90012-B (cit. on p. 37).

[155] B T Korber, R M Farber, D H Wolpert, and A S Lapedes. «Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis.» In: *Proceedings of the National Academy of Sciences* 90.15 (Aug. 1993), pp. 7176–7180. ISSN: 0027-8424. DOI: 10.1073/pnas.90.15.7176 (cit. on p. 37).

[156] Ulrike Göbel, Chris Sander, Reinhard Schneider, and Alfonso Valencia. «Correlated mutations and residue contacts in proteins». In: *Proteins: Structure, Function, and Bioinformatics* 18.4 (Apr. 1994), pp. 309–317. ISSN: 0887-3585. DOI: 10.1002/prot.340180402 (cit. on p. 37).

[157] William R. Taylor and Kerr Hatrick. «Compensating changes in protein multiple sequence alignments». In: *"Protein Engineering, Design and Selection"* 7.3 (1994), pp. 341–348. ISSN: 1741-0126. DOI: 10.1093/protein/7.3.341 (cit. on p. 37).

[158] Prashanth Rawla. «Epidemiology of Prostate Cancer». In: *World Journal of Oncology* 10 (2 2019), pp. 63–89. ISSN: 1920-4531. DOI: 10.14740/wjon1191. URL: http://www.wjon.org/index.php/WJON/article/view/1191 (cit. on p. 57).

[159] Majed Saad Al Fayi, Xiaojun Gou, Shiva S Forootan, Waseem Al-Jameel, Zhengzheng Bao, Philip R Rudland, Philip A Cornford, Syed A Hussain, and Youqiang Ke. «The increased expression of fatty acid-binding protein 9 in prostate cancer and its prognostic significance.» In: *Oncotarget* 7 (50 Dec. 2016), pp. 82783–82797. ISSN: 1949-2553. DOI: 10.18632/oncotarget.12635 (cit. on p. 57).

[160] Irwin D. Kuntz, Jeffrey M. Blaney, Stuart J. Oatley, Robert Langridge, and Thomas E. Ferrin. «A geometric approach to macromolecule-ligand interactions». In: *Journal of Molecular Biology* 161 (2 Oct. 1982), pp. 269–288. ISSN: 00222836. DOI: 10.1016/0022-2836(82)90153-X (cit. on p. 59).

[161] Jerome Eberhardt, Diogo Santos-Martins, Andreas F. Tillack, and Stefano Forli. «AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings». In: *Journal of Chemical Information and Modeling* 61 (8 Aug. 2021), pp. 3891–3898. ISSN: 1549-9596. DOI: 10.1021/acs.jcim.1c00203 (cit. on p. 59).