UNIVERSITY OF FORWARD THINKING WESTMINSTER⌗

WestminsterResearch

http://www.westminster.ac.uk/westminsterresearch

Breast Cancer Progression and Phytoestrogen Interactions with Estrogen Receptors

Balcazar Lopez, C.

UNIVERSITY OF WESTMINSTER
FACULTY OF SCIENCE AND TECHNOLOGY

BREAST CANCER PROGRESSION AND PHYTOESTROGEN INTERACTIONS WITH ESTROGEN RECEPTORS

BY

CARLOS ENRIQUE BALCAZAR LOPEZ

SUPERVISORS

DR MIRIAM DWEK (DIRECTOR OF STUDIES)

DR PAMELA GREENWELL (SECOND SUPERVISOR)

A thesis submitted in partial fulfilment of the requirements of the University of Westminster for the degree of Doctor of Philosophy

August 2017

AUTHOR'S DECLARATION

I declare that the present work was carried out in accordance with the Guidelines and Regulations of the University of Westminster. The work is original except where indicated by special reference in the text.

The submission as a whole or part is not substantially the same as any that I previously or am currently making, whether in published or unpublished form, for a degree, diploma or similar qualification at any university or similar institution.

Until the outcome of the current application to the University of Westminster is known, the work will not be submitted for any such qualification at another university or similar institution.

Any views expressed in this work are those of the author and in no way, represent those of the University of Westminster.

Signed: **Date: 15/08/2017**

Acknowledgement

Firstly, I would like to extend my gratitude to my Director of Studies, Dr Miriam Dwek, and my mentor and Second Supervisor, Dr Pamela Greenwell, for their continuous support, patience, motivation, and knowledge throughout my PhD studies. Without their expertise and guidance, completing this PhD would have not been possible. Their support, enthusiasm and opportune guidance have been vital for my enjoyment and have helped me to get the most from my PhD journey.

Besides my Director of Studies, I would like to thank my research group, Karima Brimah and Dr George Gyamfi-Brobbey for their invaluable help with the lab materials and extensive guidance and encouragement. A particular thanks to Dr Faye Bowker for her invaluable bioinformatics support.

I would like to thank the University of Westminster for awarding me the Cavendish Fee-Waiver Scholarship, thanks to which my PhD was made possible. Thank you also to all the staff and colleagues at the University of Westminster and the Cavendish Campus for their unbelievable support and kindness throughout these years, which made me feel truly at home.

Finally, without the special support, help and comfort of my family and my girlfriend, undertaking this PhD would have never been possible. Thank you.

Abbreviations

aa	Amino acid
AF1	Activation function domain 2
AF2	Activation function domain 1
AP1	Activator protein 1
Å	Ångström
bp	Base pair
BC	Breast cancer
BRCA2	Breast cancer 2 early onset (Human gene)
BRCA1	Breast cancer 1 early onset (Human gene)
CREB	cAMP response element binding protein
COACH	Meta-server protein-ligand binding site prediction
DBD	DNA Binding Domain
E1	Estrone
E2	Estradiol
E3	Estriol
EDCs	Endocrine disrupting chemicals
ER	Estrogen receptor
ER-	Estrogen receptor negative
ER+	Estrogen receptor positive
ERE	Estrogen response element
ERα	Estrogen receptor α
ERβ	Estrogen receptor β
ESR1	Estrogen receptor Alpha
ESR2	Estrogen receptor B
GPER1	G protein coupled estrogen receptor 1
Her-2/neu	Human epidermal growth factor receptor 2
kDa	Kilo Daltons
LBD	Ligand Binding Domain
mM	Millimolar
NF-kB	Nuclear factor kB

NTD	N-terminal Domain
p53	Tumour protein p53
PE	Phytoestrogen
pH	Measure of the acidity or basicity
RUNX1	Runt-related transcription factor 1
SERM	Selective estrogen receptor modulator
SNPs	Single-nucleotide polymorphism
SP1	Specific protein 1
STAT5	Signal transduction and activator of transcription 5
TAM	Tamoxifen
WHO	World Health Organization
WT	Wild Type
µl	Microliters

Abstract

Breast cancer is one of the most common diseases affecting women and approximately 1.3 million females are diagnosed each year with this disease worldwide. Breast cancer is a multi-factorial disease and it is difficult to predict or control the physiopathology, to date one of the major risk factors alongside the patient's genetic background is life time exposure to estrogen. Understanding the estrogen receptor (ER) has been a milestone in elucidating breast cancer biology, leading to advances in disease management. Alongside this, evidence from epidemiological studies suggests that dietary consumption of phytoestrogens may modulate disease progression. This study hypothesises that the interaction between some phytoestrogens (present in the pre-diagnosis diet or in the new diet adopted by breast cancer patients) and specific ER isoforms displayed in breast tumours influences the action of synthetic and endogenous estrogen in breast cancer cells. This study aimed to understand the interaction between estrogen, hormone drugs and phytoestrogens on the ER. In silico modelling of the ER focused on the wild type isoforms ERα and ERβ and different ligands (SWISS MODEL and docked through AutoDock Vina). Subsequently, isoforms of ERα and ERβ and different ligands (E1, E2, E3, PE, Tamoxifen, ICI 182,780) were modelled and tested by docking against the same set of ligands (E1, E2, E3, PE, Tamoxifen, ICI 182,780).

The system described here highlighted the main amino acid residues of the LBD of ERα and ERβ along with ligand interactions for both agonists and antagonists, described in previous X-ray crystallography experiments. All of the phytoestrogens studied using AutoDock Vina interacted with the hormone binding site of both ERα and ERβ, due the phenolic ring of the studied structure which favoured the interaction with the hydrophobic environment of LBD amino acids. All of the dietary phytoestrogens showed lower binding affinity (<9.1 Kcal/mol) compared with estradiol (-10.6 Kcal/mol) in all the isoforms and isotypes studied, suggesting that phytoestrogens should not displace estradiol from the LBD, however it remains unclear if PE can act as an agonist compound in the ER pathways. Also, some phytoestrogens appeared to have greater affinity to the ERα and ERβ than Tamoxifen (antagonist models), but it is uncertain as to whether the resulting structure will interfere with subsequent interactions. Further laboratory experiments will be necessary to understand the impact of the PE in the ERs structure and the respective role in the ER pathway.

The data from this computer modelling approach has provided an insight into the interactions between endogenous estrogens, drugs, phytoestrogens and ER. The in silico studies generated a system that recapitulated data obtained by other research groups (experimentally) and will be of value as a screening tool for further studies of new drugs and exogenous estrogens and their potential role in ER-induced breast cancer pathophysiology.

Table of Contents

Table of figures

Table of tables

Overall aims

The aim of this project was to use *in silico* protein modelling and ligand docking techniques to evaluate the binding of physiological (endogenous) ligands of the ER alongside breast cancer drugs and dietary phytoestrogens.

These aims were addressed as follows:

Chapter 3: Identification and use of wild type human ERα and ERβ ligand binding domains (LBD) as a template for protein homology modelling studies. Evaluation of the binding of estrogens and drugs (Tamoxifen/ Fulvestrant) to the LBD of ERα and ERβ.

Chapter 4: An *in silico* study of the interaction between dietary phytoestrogens and the LBD of wild type human ERα and ERβ. The development of protein homology models of the LBD of variants of ERα and ERβ.

Chapter 5: *In silico* analysis of the interaction between dietary phytoestrogens and variants in the LBD of ERα and ERβ.

CHAPTER ONE

1. Introduction

1.1 Breast cancer epidemiology

Breast cancer is one of the most common diseases in women. Worldwide approximately 1.3 million females are newly diagnosed with breast cancer each year (Ferlay, Parkin, and Steliarova-Foucher 2010). Breast cancer is a heterogenous disease that has been classified into subtypes according to the level of estrogen receptor (ER), progesterone receptor (PR) and the epidermal growth factor receptor-2 protein (Onitilo *et al*. 2009; Weigelt *et al*. 2010). According to estimated projections in the UK, breast cancer prevalence may increase threefold from the incidence of approximately 50,000 patients reported in 2010 to 150,000 by 2040 (Maddams, Utley, and Moller 2012). Worldwide approximately 465,000 patients succumb to breast cancer per annum. Changes in reproductive patterns, sedentarism, obesity and breast cancer screening programmes are partly responsibility for the reported rise in breast cancer incidence (Bray and Møller 2006). Nevertheless, breast cancer incidence and mortality fluctuates worldwide and varies geographically from Western to Eastern countries (IARC, 2012), for example, in Sweden the breast cancer incidence has been reported as 153 cases per 100,000 whilst in China the estimated incidence is 24 cases per 100,000 (Leong *et al*. 2010). Simultaneously, it is worthy of note that even within in the same continent where the population share similar lifestyles and cultural backgrounds differences in incidence rates are reported, for example, in South America; in Argentina the breast cancer incidence rates in 2008 were 74 cases per 100,000, whilst in

Colombia the incidence was less than half this number with 31 cases per 100,000 reported (World Health Organization WHO 2012). Overall, there has been a significant increase in breast cancer incidence globally during the past three decades; in developed nations this may reflect the adoption of screening programs resulting in the detection of asymptomatic (smaller and potentially less-aggressive) tumours (Holford *et al*. 2006), the trend toward decreased parity and age of first child, and in the developing world this reflects the decrease in communicable diseases and improvements in longevity. A breast cancer diagnosis brings vast changes to the lives of patients and individuals related to them and the disease continues to be a significant cause of mortality, hence a better understanding of even a single factor involved in breast cancer aetiology has the potential to make a great impact in both the diagnosis and the treatment of the disease.

1.2   Role of estrogen in breast cancer risk and disease progression

The hormone estrogen, the estrogen receptors (ER) and their downstream molecular components all play an important role in breast cancer development and progression - approximately 70% of all primary breast cancers have been shown to be estrogen receptor positive, ER+ (Hüsing *et al*. 2012). Following the binding of estrogen to the ER a series of intracellular events occur and these lead to cellular proliferation – a key hallmark of cancer(Hanahan and Weinberg 2000, 2011). The greater the rate of cell proliferation, so the increased chance that damage to DNA is inherited by daughter cells as a result of reduced time for DNA repair. The more prolonged exposure an individual or a population of individuals has to estrogen over their lifetime so their risk of breast cancer

3

increases. Such estrogen related risk factors include early menarche and late menopause (Bernstein and Ross 1993), later age of first child, decreased numbers of full term children, breast feeding, nullparity and use of hormonal replacement therapy(Chen and Colditz 2007; King *et al*. 2011).

The incidence of ER+ breast cancer is increasing in urbanised developed areas as well as in developing areas (Leong *et al*. 2010) suggesting that environment and lifestyle play an important role in the potential initiation of breast cancer (Cavallo *et al*. 2011; Giulivo *et al*. 2016). Environmental factors that affect or mimic the action of estrogen may act as carcinogens; substances include endocrine disrupting chemicals (EDCs), xenoestrogens and exogenous estrogen (Fernandez and Russo 2010; Liu *et al*. 2015). Although studies with xenoestrogens have not yet reported consistent associations between exposure and breast cancer incidence (Fernandez and Russo 2010) this may be due to specific effects of the estrogen-like chemical(s) in only a particular subtype of breast cancer, or may be due to synergism with others risk factors which have not been considered in the studies to date.

1.2.1 Estradiol

The endocrine system plays an essential role in human physiology and is key for a wide range of metabolic functions playing a pivotal role in reproduction, growth and development. Amongst the estrogens, estradiol, E2, is the most potent and most closely correlated with hormone-dependant breast cancer (Hall, Couse, and Korach 2001) as well as with other diseases such as cardiovascular diseases (Smiley and Khalil 2009).

4

Estradiol has a molecular volume of 245 Å and as a result of its phenolic structure is relatively lipophilic (Anstead, Carlson, and Katzenellenbogen 1997). Studies with structural models of ERs that have been built in complex with estradiol have helped to provide an understanding of the binding site of these receptors. Knowledge of the structure of the ER as well as the biochemical pathways leading to the *in vivo* synthesis of estrogens have led to the development of drugs targeting the receptors as well as estrogen biosynthesis; many of these have found their way into clinical use as breast cancer treatments (Smith *et al.* 2003).

The ovaries are the main source of estrogen but the conversion of precursor molecules to estradiol can also occur in other tissues. Numerous enzymes involved in the biosynthesis of estradiol (E2) and (the less active estrogen) estrone (E1) have been described, these include the hydroxysteroid dehydrogenase family and the aromatases. The release of estradiol by aromatase enzymes, particularly from adipose cells, is an important source of estrogen in postmenopausal women (Brueggemeier 2001) and provides a rationale for the approximate three-fold increased risk of breast cancer associated with obesity (Ziegler *et al.* 1996).

Estrone is usually stored in the breast in an inactive form as estrone sulphate, this is then reduced by the dehydrogenase enzyme (17β) to the more potent estradiol which interacts with ERs resulting in cell proliferation (Pike *et al.* 1999). The estradiol exerts its effect by docking into the ligand binding domain (LBD) of ER, ligand-dependant transcription is followed by gene activation at estrogen response elements (EREs) located at promoter regions of many estrogen-

responsive genes resulting in DNA replication and leading to cell proliferation (Flötotto *et al.* 2004).

1.3 ER in breast cancer aetiology

Until the 1960s it was believed that estrogens were only involved in the sexual reproductive system through oxidation-reduction pathways but in 1962 Jensen *et al* used tritiated estradiol and reported the uptake of the tritiated compound in uterine and vaginal tissues but not neighbouring normal tissues (Jensen and Jacobson 1962), subsequently Noteboom and Gorski, 1965, proposed that tritiated estradiol interacted with a receptor which was likely to be a protein (Toft and Gorski 1966).

The ER was first described as such in 1985 (Green *et al.* 1986) and at that time was thought to comprise a single protein, now known as ERα, since at that time an additional form of ER (ERβ) was discovered (Bakas *et al.* 2008; Kuiper *et al.* 1996) alongside a number of splice variants and also single nucleotide polymorphisms (SNPs) in the genes for ERα and ERβ themselves.

ERα has been shown to have the greatest affinity to 17β-estradiol but it can also interact with a wide range of compounds; altering normal pathways activated following estrogen binding. The interaction between the different ER isoforms with both endogenous and exogenous estrogen may vary according to the cell/tissue. The complexity of different pathways activated by specific ligands alongside the presence of different heterodimers and SNPs in the ER pose a

major challenge for scientists who aim to gain a fuller understanding of estrogenic pathways.

1.3.1.1 Targeting ER in breast cancer

Recognition of the importance of estrogen as a mediator in breast cancer led initial surgical treatments to reduce levels of circulating estrogens by removal of the ovaries (ovarian ablation/oophorectomy) and although the studies using this approach were small, improved survival after breast cancer was observed in women less than 50 years of age (reviewed in Schiavon & Smith, 2014). These initial studies led to the development of endocrine therapies for breast cancer which have made a lasting impact on the care of patients with breast cancer. The treatment of breast cancer encompasses surgical, radiological, chemotherapeutic, biological as well as hormonal therapies and since the 1970s the 10-year survival rate after a breast cancer diagnosis has more than doubled to approximately 80% (Cancer Research UK 2012).

The first important drug developed targeting the ER was Tamoxifen, a triphenylethylene derivative, initially known as ICI 46,474, Tamoxifen had been developed by ICI as a morning-after pill but studies in this setting proved unsuccessful. Craig Jordan and colleagues used a rodent model of mammary carcinoma and were the first to show the potential of Tamoxifen as a treatment for breast cancer (Jordan 2004). Tamoxifen is metabolised to 4-hydroxy Tamoxifen and 4-hydroxy N-des-methyl Tamoxifen in the liver by enzymes of the cytochrome P450 system (CYP2D6). The active 4-hydroxy forms are selective estrogen receptor modulators (SERM) and act as antagonists of ER in breast cancer cells (Webb *et al*. 1995) Studies with tens of thousands of

patients have shown the importance of Tamoxifen as a first-line therapy for pre-menopausal breast cancer (with ovarian ablation) (reviewed in Schiavon & Smith, 2014).

The development of Tamoxifen paved the way for other selective estrogen receptor modulator compounds (SERMs) such as Raloxifene and the "pure antagonist" Fulvestrant: ICI 182,720 (McDonnell and Wardell 2010). At a molecular level SERMs function by blocking the activation domain (AF2) of the ER by displacing helix 12, rotating it by 110° thereby rendering the binding site for the co-activator protein p160 inaccessible (Kushner *et al.* 2000).

A class of compounds known as aromatase inhibitors (AIs) have been developed for use in post-menopausal women. AIs include anastrazole, letrozole and exemestane, these bind to aromatase and prevent the synthesis of estrogens from parent compounds such as androgens and have proved to be a very successful class of drugs (Dowsett and Howell 2002). Rather than acting directly as SERMs these compounds work "upstream" preventing the synthesis of endogenous estrogens.

It has been estimated that up to 40% of all cases of ER+ breast cancer become resistant to long term treatment with Tamoxifen (Shou *et al.* 2004). The reasons for the development of such resistance is not clear but may include: selection pressures favouring ER negative breast cancer cells; cross-talk between different signalling pathways; the presence of ER negative stem cells and interactions between different isoforms of the ER with endogenous and exogenous estrogens (Chang 2012).

The complexity of different pathways activated by specific ER ligands alongside the presence of different heterodimers and SNPs poses a major challenge for scientists who aim to gain a fuller understanding of estrogenic pathways, resulting in more aspects to investigate and study, and the opportunity to develop approaches for novel breast cancer treatments focusing on, for example, SERMs tailored to patients with different isoforms of the ER and genetic variants (SNPs).

1.3.2 Phytoestrogens and the diet

The term phytoestrogen encompasses a wide range of phenolic compounds called flavonoids, they are low molecular weight compounds with estrogen-like properties due to the presence of the phenolic ring A and the similar distance between the 7 and 4 hydroxyl groups in the isoflavones when compared with the hydroxyl groups of estradiol (Dixon 2004). Phytoestrogens are secondary metabolites and dietary sources include cruciferous vegetables, seeds, fruits, coffee, soy beans, berries and flowers (Middleton, Kandaswami, and Theoharides 2000).

Isoflavonoids are naturally-occurring polyphenolic compounds, mainly found in the *Leguminosae.* Some isoflavanoids have been found to bind to the human ER with an affinity similar to that of estradiol (Davies, Batehup, and Thomas 2011). Soy-containing foodstuffs are enriched in isoflavones and are an important source of dietary phytoestrogens (Velentzis *et al.* 2011). Isoflavone-containing foods are consumed in Asian countries, particularly China and Japan (Shu *et al.* 2009) and this increased consumption of isoflavones: daidzein,

genistein and glycitein (Banerjee *et al.* 2008) may at least partly explain the lower incidence of breast cancer in these countries (Caan *et al.* 2011). Aside from the ingestion of daidzein and genistein the gut microflora of approximately 30% of humans produces metabolites such as enterolactone and equol and these metabolites have been thought to bind to the ER with greater affinity than the parent compound (Setchell and Clerici 2010). It is not known if the association between high consumption levels of dietary phytoestrogens and reduced incidence of breast cancer is linked to the ability to metabolise daidzein into equol (Cederroth and Nef 2009). Daidzein, genistein and equol appear to show greater affinity to the ERβ than ERα(Kuiper *et al.* 1998; Lund *et al.* 2004).

1.3.2.1   *Bioavailability and metabolism*

Phytoestrogens are often classified according to the number of substituents on their ring structures and the most important groups arising from such classification are the flavones (apigenin), flavanones, flavonols and isoflavones (biochanin A, formononetin, genistein, daidzein and glycitein), coumestan (coumestrol), and lignans (secoisolariciresinol, matairesinol, enterolactone, enterodiol) (Romano *et al.* 2013). Dietary phytoestrogens are found either in a free-state or in a more stable state bound to a sugar, leading to relatively low bioavailability after ingestion. The ratio of bound versus glycone free forms of phytoestrogens varies according to the food source itself. Fermented soy-based food stuffs tend to have higher levels of glycone free isoflavones although the conjugated form can be deconjugated by the intestinal microbes after consumption (Axelson *et al.* 1984). Further complexity arises as phytoestrogens such as biochanin A and formononetin can be metabolised to genistein and daidzein respectively by the action of glucosidases of the microbiota (Axelson

et al. 1984). However, to produce the metabolite equol and its alternative O-desmethylangiolensin (O-DMA) from daidzein, specific types of (as yet unidentified) intestinal bacteria are required (Rafii 2015) but these are present in only 30-50% of the population (Setchell and Clerici 2010). Determining the "equol metabolism" status of an individual is of some relevance since equol has been shown to exhibit greater estrogenic potency than daidzein (Setchell and Clerici 2010). Besides their affinity for the ERs phytoestrogens can affect the metabolism of the steroid hormones as they act as natural anti-oxidants and as anti-inflammatory compounds having been reported to affect a wide range of signalling pathways resulting in reduced inflammatory responses in cells (Hwang and Choi 2015).

1.3.2.2 Phytoestrogens and breast cancer

The role of the phytoestrogens in disease prevention has been the subject of over a thousand scientific peer-reviewed publications and their role in breast cancer risk reduction and breast cancer outcomes have been the subject of study by many research groups. Over the past two decades studies have been concerned with both molecular and cellular as well as whole-individual and population studies(Bliedtner *et al*. 2010; Ganry 2005; Harris and Besselink 2005; Rice and Whitehead 2008; Romano *et al*. 2013).

In terms of breast cancer incidence, a lower level of phytoestrogen consumption is found in Western diets (Chun, Chung, and Song 2007) where breast cancer is more prevalent compared with Eastern countries where there are much higher levels of consumption of phytoestrogens reported and lower breast cancer incidence rates (Mortensen *et al*. 2009). After Asian people emigrate to host

11

countries with higher breast cancer incidence levels so the next generation adapts their diet consuming the lower levels of phytoestrogens associated with the behaviour in Westernised countries (Shimizu *et al.* 1991). The serum levels of estradiol are 40% lower in Asian women compared with the levels in Western women (Peeters *et al.* 2003). In addition, women in Western countries tend to consume lignans as well as isoflavones in their diet further adding to the complexities associated with studying this area of biology (Swann *et al.* 2013).

The effect of dietary phytoestrogen consumption on breast cancer survival remains unclear. A large meta-analysis reported by the World Cancer Research Fund in 2014 disappointingly concluded that more research is still required on this topic (Cancer Research UK, 2014). A number of on-going studies including the UK-based DietCompLyf study (Swann *et al.* 2013) and the After Breast Cancer Pooling Project (Nechuta *et al.* 2013) seek to explore this topic in more detail as both dietary and environmental factors have been estimated to play a major role in outcomes after breast cancer.

It remains unclear as to the functional role of phytoestrogens in breast cancer progression since the data produced related to phytoestrogens are inconsistent between *in vitro* and *in vivo* systems. Phytoestrogens such as daidzein and genistein have been suggested to function as SERMs and this may relate to their tissue specific effects, for example, agonistic activity in bone related to osteoporosis, whilst antagonistic activity in breast tissues inhibiting breast cancer. The isoflavones have been shown to act in an agonistic fashion towards human breast cancer cell lines (Simons *et al*, 2011). Genistein functions *in vitro* as an inhibitor of tyrosine kinase action in the epidermal growth factor receptor

12

pathway which is over-expressed in many cancer including breast cancer (Akiyama *et al*. 1987; López, López, and Arias 2015)(Akiyama *et al*. 1987; López *et al*. 2015). Also, it has been reported that genistein acts as an anti-proliferative agent at more than 10 mM concentration but at low doses *in vitro* it acts as a tumour agonist (Wang, Sathyamoorthy, and Phang 1996). Further research should help to unravel the effect of isoflavones on malignant cells and produce data related to the benefits and/or adverse effects that arise from phytoestrogen exposure.

1.3.2.3 DietCompLyf study

The DietCompLyf study is a multi-centre observational study that was run by the Against Breast Cancer Research Unit at the University of Westminster. The study is focussed on determining the role of PEs on BC recurrence in a UK population (Velentzis *et al*. 2011)(Velentzis *et al*., 2011). A key element of the DietCompLyf study is the availability of reported food intake alongside biological samples enabling analysis of PE intake, biomarker levels and correlation with clinical treatments and patient recurrence rates. The findings of the *in silico* and *in vitro* experiments will be compared with data and samples collected as part of the DietCompLyf study. The database of the DietCompLyf study counts comprises around 3,169 women. A separate nested case control cohort of around 200 patient matched with all the clinical criteria is available. All the subjects have provided a food frequency questionnaire, and a 7-day food diary allowing PE levels to be estimated in addition their urinary PE levels which have been quantified by liquid chromatography/tandem mass spectrometry(LC–MS/MS) (Grace *et al*. 2007).

1.4 ERα and ERβ

Of the two ER subtypes, ERα is mainly expressed in the human breast, uterus, cervix and vagina, while ERβ is mainly found in the spleen, lung, hypothalamus, thymus, immune and central nervous systems, urogenital tract (ovary, prostate, and testis), bone and kidneys (Couse *et al.* 1997). The selective expression of the two main forms of ER leads to specific ligand action in each of these tissue types. Both of the major subtypes of ERs have been identified in breast cancer (Fixemer, Remberger, and Bonkhoff 2003). Despite being encoded by different genes (ERα: *ESR1* gene and ERβ: *ESR2* gene) and on different chromosomes, 6 and 14 respectively (Barnes 2010), both receptors have structural similarities and a high degree of homology as shown in Figure 1.
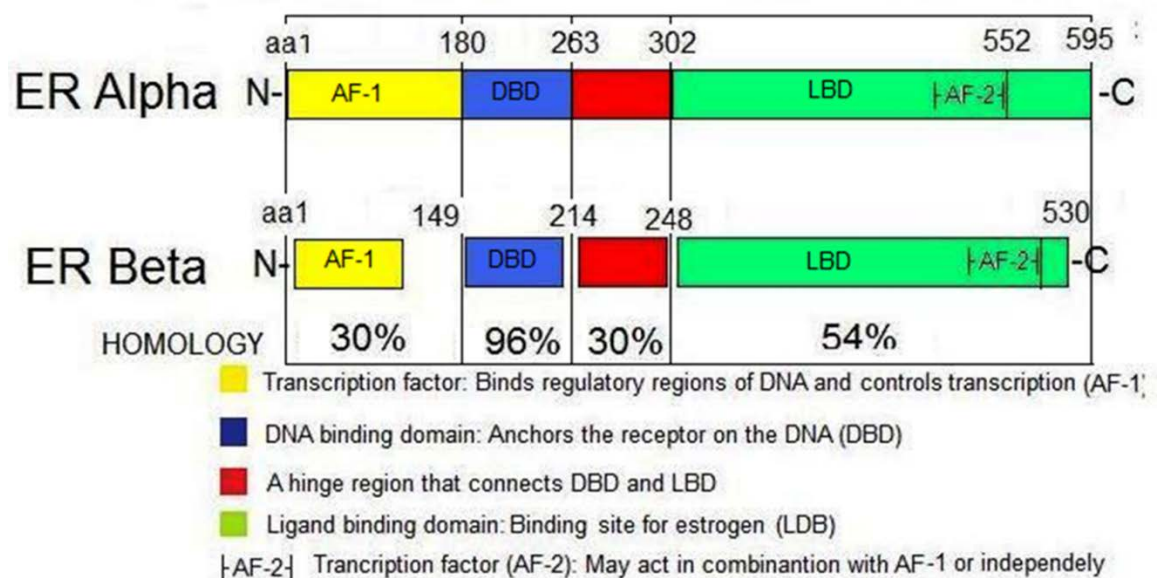
Figure 1. Structure of estrogen receptor alpha (ERα) and estrogen receptor beta (ERβ). Amino acid positions for the structural motifs are displayed above each protein. The percent amino acid homology is given under ERα and ERβ.

The DNA-binding domains (DBD) of human ERα and ERβ show 96% amino acid homology. Whilst ERα and ERβ share a homology of 54% in the ligand binding domain (LBD) amino acid sequence, the affinity for estrogen is similar

in both subtypes (Huang, Chandra, and Rastinejad 2010) but each receptor regulates specific genes and displays a different profile of gene networks following activation (Eyster 2016). The major difference between the sequences for ERα and ERβ is in the activation factor (AF-1) domain which relates to hormone independent transcription activity. ERα and ERβ have been reported to be co-expressed, acting as homo- or hetero-dimers although the impact on cellular pathways has not yet been unveiled.

ERs are members of the nuclear receptor superfamily sharing common regions (A/B, C, D, E and F regions) which form independent but interacting functional domains; N-terminal domain (NTD); DBD; LBD and two activation function domains AF1 and AF2, located in NTD and LBD, respectively. The AF1 domain recruits co-factors independent of ligand binding and AF2 recruits co-factors but is dependent on ligand binding (Kumar *et al*. 2011). AF1 and AF2 together are responsible for the regulation of receptor activity during transcription. A number of important co-activators have been described for ER including: SCR-1 / NCoA1 and SCR-2 / TIF2 / GRIP1 / NCoA-2, these bind a hydrophobic groove in the receptor termed the LXXLL motif, the co-activator/ER complex then leads to gene transcription by virtue of recognition of estrogen response elements (EREs) (Gruber *et al*. 2004). Another mode of action is via interaction between ER and transcription factors for example: AP1, SP1, NF-kB, CREB, RUNX1, p53 and STAT5, these bind the ER following ligand-ER complex formation (Kushner *et al*. 2000). This recruitment of transcription factors leads to gene transcription in genes that lack EREs. ERs exhibit a high degree of plasticity interacting with different ligands ranging from natural endogenous estrogens to other substances which mimic these hormones, binding to the receptor and

leading to a conformational change in the LBD domain and enabling co-activator interactions with ER (Shiau *et al.* 1998). Understanding the molecular interactions between the range of ligands in nature, their effect on the molecule structure of ER and pathways activated down-stream of ligand binding will be an important tool in understanding breast cancer development and may offer potential for the development of breast cancer treatments.

1.4.1 ER ligand binding domain (LBD)

The LBD of ERα and ERβ are only 54% similar at amino acid residues 302 to 552 and 255 to 504 in ERβ (Kuiper *et al.* 1996) leading to different outcomes (agonist or antagonist) from the two forms of this receptor despite interaction with the same ligand. When the 3D structure is considered, despite the low degree of sequence homology, the nuclear receptor superfamily display a highly degree of similarity in terms of conformation of the LBD. In ERs the LBD is formed by eleven helices (H1 to H11) contained within the ligand binding site, a further helix (H12), a dimerization interface and co-regulator interaction is sited within a globular structure (Moras and Gronemeyer 1998).

The helices found within the ER LBD are arranged in three layers in agonist conformation with helices H4, H5, H6, H8 and H9 lining up on one side, H1 and H3 and on the opposite side helices H7, H10 and H11. H12 sits on the binding site cavity against H3, H5/6 and H11 without making contact with the ligand and forming a lid-like structure in the binding pocket and contributing to the hydrophobic microenvironment within the cavity, becoming an essential component for binding to partner proteins for later transcriptional activation (Brzozowski *et al.* 1997). In the ligand-free state the ERs are found in a complex

with chaperon proteins (Heat shock proteins 70 and 90) mainly interacting with the LBD of the protein (Pratt *et al.* 2004) upon ligand binding, the ER structure displays structural modifications leading to homo- (ERαα and ERββ) and hetero- (ERαβ) dimerization and nucleus translocation.

1.4.2 Mechanism of action ER

Several genes are regulated by ER and these can be identified by the presence of EREs or by the binding sites for a range of transcription factors which have been recognised to bind to ER. The classical pathway has a more limited range of targets when compared to ER activity at alternative response sites (Bjornstrom and Sjoberg 2005)(Bjornstrom and Sjoberg 2005). The effects of estradiol occur through at least four pathways, Figure 2.
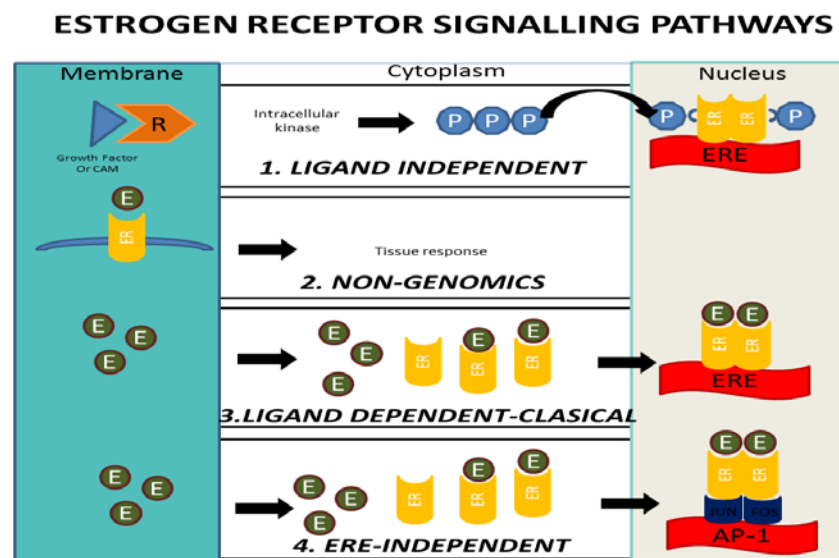
Figure 2. Mechanism of estrogen receptor signalling. 1. Ligand-independent: intracellular kinase pathways are activated by growth factors and cyclic adenosine monophosphate (CAM) phosphorylation (P) leading to the activation of ER and ERE binding in a ligand-independent fashion. 2. Non-genomic signalling: Estrogen activates a form of ER linked to intracellular signal transduction generating a tissue response. 3. Ligand-dependent: E2-ER binds to EREs leading to up- or down-regulation. 4. ERE-independent: E2-ER associates with DNA transcription factors FOS/JUN tethering ER to alternative response elements.

17

In the ligand-independent pathway ERs may be activated by epidermal growth factor (EGF), insulin-like growth factor-1 (IGF-1) and 8-bromo-cyclic adenosine monophosphate, activating intracellular kinases leading to protein phosphorylation, which then increases gene expression by ER activation (Hall *et al.,* 2001; Thomas and Gustaffson, 2011). In a further pathway the E2-ER complex can induce gene expression even in the absence of direct DNA binding (Kushner *et al.,* 2000), furthermore cell-surface signalling may occur. ERs are associated with signalling pathways functioning via cell membrane receptors, for example, CPCRs, ion channels and protein tyrosine kinase pathways (Simoncini and Genazzani 2003).

After ligand binding, the receptor undergoes a conformational modification whereby the chaperon protein heat shock protein 90 (Hsp-90) is released resulting in the nuclear translocation of ER dimers, binding with high affinity to specific nucleotide sequences in palindromic motifs (EREs) and recruitment of co-activator or co-repressor molecules resulting in gene expression or preventing transcription according to the nature of the ligand, agonist or antagonist, respectively (Dao and Hanson 2012). The alpha helix 12 at the LBD plays an important role at this stage allowing or preventing the interaction of the transcription complex. In the genomic pathway ERs can regulate transcription without binding to DNA, instead tethering receptors through protein to protein interactions to a transcription factor contacting the DNA. A large number of estrogen-responsive genes not containing ERE sites are modulated by this mechanism, and this is generally termed transcriptional cross-talk (Bjornstrom

and Sjoberg 2005). Subtypes of ER are able to mediate the non-genomic effects of estrogens within less than 15 minutes after estrogen treatment through ion movement through the membrane (Marino *et al.,* 2006; Kampa *et al.,* 2013). It appears, however, that the ligand binding domain of ERα targeted to the plasma membrane is sufficient for the regulation of non-genomic activity of estrogens (Adlanmerini *et al.* 2014).

1.4.3 ERα isoforms

Despite being encoded on different chromosomes by different genes: ERα, ESR1 and ERβ, ESR2, ERα on chromosome 6 and ERβ on chromosome 14 (Barnes, 2010), the receptors show structural similarities and a high degree of homology as described above.

1.4.3.1 Estrogen receptor alpha isoforms

Wild type ERα66 is the most prevalent form in BC cells, other isoforms have been reported as involved in cancer pathology (ERα46, ERα36, ERαΔ5/6/7, ERαΔ3/4/5, ERαV) (Flouriot *et al.* 2000; Marino, Galluzzo, and Ascenzi 2006; Rao *et al.* 2011) ERα36 has been reported to interfere with ERα66 activity and the estrogen treatment in BC. (Wang *et al.* 2005). ERα36 is an isoform of ER-α66 cloned in 2005 by Wang *et al.* (2005) which lacks both activation domains while retaining a DNA binding domain, ligand binding domain and partial dimerization, and is found in both the cytoplasm and cellular membrane. This might indicate its involvement in both genomic and non-genomic pathways (Bjornstrom and Sjoberg 2005). There is evidence that ERα36 possesses a broader activity spectrum than ERα66, initiating membrane-linked signalling in response to E2-α, E2-β, E3 and E4. It competes with ERα66 and ERβ for the

DNA-binding elements, and its expression appears to be related inversely with ERα66 expression. At UNIPROT Database ERα isoform 1 or wild type has been chosen as a canonical form due to its similarity to orthologous sequences found in other species, by lengths and amino acid composition. ERα isoform 1, also known as a ERα66 or hER-alpha66, was isolated and sequenced in MCF-7 covering the entire mRNA and expressed in HeLa cells (Greene *et al.* 1986). ERα isoform 2 (known as a short ERα) and comparing with the canonical form this isoform is missing the amino acid residues 255 to 366 and has a molecular weight of 53 KDa in comparison with the wild type which is 66 KDa. ERα isoform 3, also known as hER-alpha46 or ER46 due to its molecular weight of 46 KDa, is missing the N-terminal 173 amino acids encoded by a transcription factor which lacks the exon 1. ERα isoform 3 is a potent inhibitor of ERα66 isoform 1 when transactivation AF1 is predominant over AF2 hence heterodimerization may play an important role in this context or simply binding competition for ERE (Flouriot *et al.* 2000). The ERα isoform 4 also known as hER-alpha36 or ER36 lacks of both transcriptional activation (AF1 and AF2) domains and a unique sequence of 27 amino acid residues replaces the last 138 amino acids of the canonical sequence (ERα isoform 1) (Wang *et al.* 2005). ERα isoform 4 activates non-genomic pathways hence inducing cell proliferation. Also, this isoform has been detected in ERα positive and ERα negative BC patients. Knocking down its expression in BC cancer cells was reported to decrease in cancer cell migration, invasion and increase the response of BC cells to endocrine therapy (Zhang *et al.* 2012).

1.4.3.2 Estrogen receptor beta isoforms

ERβ acts as an antagonist of ERα in some estrogen stimulated cells. ERβ was described in 1996 and in 1998 five isoforms of ERβ were identified (ERβ1-5, where ERβ1 was the wild type and ERβ5 was already known as a ERβcx) (Moore *et al.* 1998). ERβ1 is the only ligand dependant transcription isoform though all the isoforms can participate in hetero-dimerization with ERα and ERβ subtypes, negatively impacting transcriptional activity (Shanle and Xu 2010). ERβ1 is linked to better survival in triple-negative BC as well as better response to Tamoxifen monotherapy, while nuclear ERβ2 levels have been shown to be associated with metastasis and vascular invasion, also, cytoplasmic ERβ2 is linked to a poorer outcome in BC (Nilsson, Koehler, and Gustafsson 2011). Studies of ERβ remain a focus of clinical research, isoforms ERβ1 and ERβΔ5 are co-expressed in healthy mammary glands (Mandušić, Nikolić-Vukosavljević, Tanić, Ksenija, *et al.* 2007); in particular, ERβΔ5 may has inhibitory activity on both ERα and ERβ, and its higher expression is linked to indicators of low aggressiveness of the tumour. Levels of ERα and ERβ decline with progression of disease in a subtype specific manner. Adverse effects of estrogens upon expression of ERα is observed in most cancers. On the other hand, the activation of WT ERβ has been linked with beneficial estrogenic activity, supporting the rationale of ERβ agonists for treating this disease. ERβ isoform 1 known as Wild type or ERβ1 has a molecular weight of 60 KDa and it is considered the canonical sequence of the ERβ. ERβ isoform 2 is called ERbeta-2 or CX which is truncated at the N-terminal and differs from the canonical sequences at the amino acid residues 469 to 530 and was described along with ERβ isoform 4, 5 and 6 (also known as beta 3, beta 4 and beta 5 respectively). The ERβ isoform 2 sequence is almost identical to that of the ERβ isoform 1

though the last 61 amino acid residues at the C-terminal are replaced but a unique sequence of 26 amino acids. It has been reported that ERβ2 inhibits E2-Transcription activation when it is co-expressed with the ERα isoform 1, possibly by competition in the ERE binding site or transcriptional silencing (affecting co-activators or ER-related basal transcriptional factors), without discarding the possibility that ERβ isoform 2 acts through a novel pathway, ligand binding effect or specific ERE (Ogawa *et al*. 1998). ERβ isoform 1 prefers to dimerise with its own isoform overall ERβ isoform 5 and ERβ isoform 6 leading to a higher transactivation activity than the normal ERβ isoform 1 homo-dimerization (Nettles *et al*. 2008).The presence of more isoforms of ER and changes in the ratio ERα:ERβ isotypes and isoforms following by subsequent hetero or homo-dimerization alongside varying levels of expression in different cells/tissues provides an opportunity for a better understanding and development of different approaches for improving prognosis and treatment of cancer.

1.5 Modelling three dimensional protein structure and structure-based studies

The study of bio-active molecules through several fields including drug discovery has contributed to the understanding of how they can influence physiological pathways. One target could align several molecules as a ligand, which could play different roles such as agonist, antagonist or to have multiples effects on disease progression (Feixas *et al*. 2014). A description and characterization of the diverse ligand-protein interactions will be important to understand biological function and downstream effects. Protein three-dimensional structures have played a major role in the functional annotations, determinations yet the gap between number of experimentally elucidated

structures and known protein sequences is increasing. Although, the amino acid sequence of a protein determines its three-dimensional structure, the mechanism of protein folding is not completely unravelled. The structures of protein families members are more conserved than the sequences (Chothia and Lesk 1986) hence sequence similarity means structural similarity as well. Several computational methods have been used to predict 3D protein models (Template-based or homology methods and *de novo* modelling). *De novo* methodologies are based on the folding process simulation using the protein amino acid sequence as a starting point followed by the structure prediction with the lowest free energy structure without template from first principle (Bradley, Misura, and Baker 2005). Despite the challenge *de novo* approaches to understanding the protein folding has displayed recent success in structure prediction (Khoury *et al.* 2014) yet there is no method that can accurately predict protein structures without a template. On the other hand, homology modelling is considered the most common applied method due to its high accuracy in comparison with *de novo* methods.

1.5.1 Homology modelling

Homology modelling relies on experimental structures of homologous proteins to provide protein models that ideally should match the accuracy of the experimental model. Homology modelling has helped researches to narrow the space between structural space and protein sequences leading to production of models of proteins that have been difficult to crystallize or determine by NMR spectroscopy. Starting from the fact that protein sequences that are evolutionary connected will show similar 3D structures(Chothia and Lesk 1986) known experimental protein structures will lead to protein models of unknown

experimental structures. Homology models can be used for a wide range of *in silico* experiments from drug screening to support experimental data in crystallography (Geromichalos 2007; Uchoa *et al.* 2004). In fact, homology modelling requires several systematic steps to reach more accurate and reliable models: template identification, sequence alignment, back bone generation, loop modelling, model optimization, validation and iteration.

Validation of a model is highly relevant when models are used for structural-based studies and screening. Models accuracy tends to be lower that the used template or parent structure yet low quality models still can be used for *in silico* experiments such as mutagenesis experiments. However, to perform structural-based virtual screening (SBVS) high accuracy models are needed (Fan *et al.* 2009), when a high-quality models are built it can be comparable with medium resolution X-ray crystallography structures.

1.5.1.1 *SWISS-MODEL*

Several *in silico* tools and user-friendly web servers have been developed to facilitate researchers to either allow the modeller to go step by step in the building process of the models ranging from template selection to model validation or automated/semi-automated web-servers that aim for the model through specialised algorithms to provide and resolve critical choices to help an inexpert modeller.

SWISS-MODEL (Biasini *et al.* 2014) is a user-friendly automated web server interface that can generate homology protein models through the integration of a suite of web servers that provide protein homology modelling and model

assessment. The main steps for homology modelling are template identification, sequence alignment, modelling and model validation. Template identification can be performed based on the target sequence using BLAST (Altschul *et al.* 1997) E-value as a limit, sequence alignment can be reliable only at >50% identical amino acid residues (Rost 1999) but alongside BLAST a search for reliable templates is performed against a library of Hidden Markov Model (Söding 2004) method to detect distance relationships between target and possible template even if part of the target sequence has not been uncovered yet. E-value score is used to rank the results of the template searching and after finding the template with significant sequence similarity and respective sequence alignment template-target, the model coordinates are built based on the protein structure conserved regions between target and template by ProMod3 (Biasini *et al.* 2014) and energy minimized using Gromos force field (Christen *et al.* 2005). For each model built its respective quality estimation will be provided with local and global estimates, QMEAN and QMEAN Z-score respectively.

1.5.1.2   Model quality assessment: QMEAN and QMEAN Z-score

Local and global quality of the models is assessed by the quality model energy analysis (QMEAN) scoring function, this assesses important aspects of the protein structure to identify the native structures among the generated models, which cover three statistical potential terms and the agreement of the predicted secondary structure and   solvent accessibility (Benkert, Tosatto, and Schomburg 2008) including:

- The torsion angle potential: extended torsion potential over three residues, describing the local geometry.

- Solvation potentials: Cβ atoms, the potential reflecting the degree of solvent exposure (in a sphere of around 9 Å) for a given amino acid residue.

- Distance dependent pairwise potential: pairwise α/β, residue-specific pairwise distance-dependent potential using C α and β as the interaction centre; secondary structure specific implementation of the residue-level pairwise potential for long range interactions.

- Secondary structure and solvent accessibility agreement: secondary structure implementation for derivation and application of the potential agreement of the target-sequence calculated secondary structure using PSIPRED (Jones, 1999) and agreement of the predicted binary burial status provided by ACCpro (Cheng *et al.*, 2005) to calculate the solvent accessibility cut off at 25%.

In order to assess protein models independent of size it is needed to estimate the absolute model quality by relating the features of the predicted model to experimental structures of similar size based on QMEAN score leading to a QMEAN Z-score (estimate of degree of nativeness or deviation from native behaviour). Scores of single body potentials are normalised by the number of amino acids residues and the score of the interaction potentials (atoms and Cβ potentials) compared with the number of interactions in order to correct the size dependence of the potential scores (Benkert, Biasini, and Schwede 2011). QMEAN is an important tool for model quality assessment in homology modelling increasing the performance of the score through the combination of

different terms and QMEAN Z-scored provides data about the absolute quality of the predicted homology models.

1.5.2 Molecular Docking

Molecular docking is an approach that is able to compute stable and energetically favourable poses of a specific ligand in complex with a protein structure. Sampling or search algorithms are responsible for exploration of the motion space and is able to score and generate energy binding data of the complex, ligand conformation and orientation (Elokely and Doerksen 2013). In order to predict the conformation of the receptor-ligand complex a computer simulation procedure is needed. Each docking software needs one or more search algorithms to predict possible conformations. These methodologies can be divided in three groups; systematic methods (Matching algorisms and incremental construction), stochastic methods (Monte Carlo and genetic algorithm) and simulation methods (molecular dynamics and energy minimization) (Kitchen *et al*. 2004). Matching algorithms are used where the ligands are treated as pharmacophores and hydrogen bonds are taken into account for the calculations (DOCK). Incremental construction methodology is where the largest fragment is used as an anchor and the remaining fragments are added incrementally (FlexX). LUDI fragments are fitted in the binding site to evaluate interaction sites to finally connect all the single fragments into a final molecule. Stochastic methods randomly change a ligand conformation to explore the conformational space. Monte Carlo and genetic algorithms belong to this method group. Monte Carlo methodology is based on the production of poses through bond rotations or rigid body and conformation are assessed with an energy-based selection. Genetic algorithm methods stem from Darwin's

theory of evolution (Meng *et al.* 2011). In simulation methods, molecular dynamics is the most used approach though it is computationally exigent and time consuming and may not be able to fit the ligand into the target and usually simulation methods are more used to complement other algorithms and methodologies. After docking the ligand in the protein it is necessary to rank and evaluate the potential binding modes. After estimating prediction of the binding conformation, is necessary to discern between correct and false poses. This calculations are mainly divided in three groups: force field-based (quantification of ligand-target complex energy and internal ligand energy), empirical (regression analysis using experimental pre-calculated binding data) and knowledge-based scoring function (pairwise atoms potential to evaluate ligand binding interaction.

1.5.2.1 *AutoDock Vina*

AutoDock Vina is an open-source software that has been widely used in research and academia. Vina can be used for a docking single ligand at a time or for virtual screening with the support of the MGL tools (Forli *et al.* 2016). AutoDock is a suite of automated docking tools. AutoDock combines an empirical force field with Lamarckian genetic algorithms (Morris *et al.* 1998). Providing fast prediction for bound poses or conformations alongside free energies score. Through the use of grid-based method autodock that allow an efficient evaluation of the motion space of the protein binding site and the binding energy of the conformations. In this method the protein is embedded in a grid, in which the protein is embedded in a grid to search the entire conformational space around a protein available to a ligand. At this point, a probe atom placed at each grid point calculates the energy from the interaction

between probe and target and the value is stored and remains available to refer to during the docking simulation. Although the use of the grid treats the target protein as rigid, the ligand is flexible and only torsional degree of freedom are used keeping bond lengths and bonds angles constant (Cosconati *et al.* 2010). The Lamarckian genetic algorithm is the principal method for conformational searching, where a population of "trial" conformations is generated and mutates during successive generations, exchanging conformational parameters, and competing in a similar way to biological evolution. Finally, the individuals with the lowest binding energy are selected (Morris *et al.* 1998). The minima are then passed to later generations.

AutoDock4 possesses also a simulated annealing search method as well as a traditional algorithm search. Conversely, the empirical free energy force field is based on a thermodynamic model that includes the intramolecular energies into the predicted binding energy. This force field evaluates both bound and unbound energy states and is used to predict binding free energies of molecules. It makes use of a charge based desolvation method which is calibrated using a set of 188 protein-ligand complexes for all of which the structure and binding energy is known. For this set, studies have shown a standard error of 2-3 Kcal/mol in the prediction of binding free energy (Morris *et al.* 2009). Compared to AutoDock4, AutoDock Vina shares some concepts and approaches, such as the pre-calculation of grid maps - in Vina this happens internally - the idea of docking as a stochastic optimisation of the scoring function, the structure format (PDBQT) and various implementation characteristics, such as pre-calculating the interaction between every atom pair at each distance. AutoDock Vina has an improved search routine (empirical scoring function and Iterated Local Search

global optimizer) to provide docking conformations and is able to take advantage of multicore setups (Trott and Olson 2010b).

CHAPTER TWO

2. Methodology

2.1 Bioinformatics docking studies, molecular and homology modelling

2.1.1 Homology modelling

The ligand binding domains (LBD) of human ERα and ERβ have been well described. These were utilised as templates for homology modelling experiments using X-ray crystallography data retrieved from the protein data bank (RCSB) and using the on-line SWISS-MODEL server (Biasini *et al.* 2014). The PDB files for the experimental crystal structures (coordinates for the templates) were uploaded in the oligomeric form without ligands and in the absence of non-amino acid groups. The input data used was the LBD sequence for ERα and ERβ, accession numbers P03372 and Q92731 respectively and submitted in FASTA format.

Homology models of the 3D-structure of ERs were constructed using the amino acid sequences for different human ER subtypes and variants retrieved from UNIPROT (Appendix 1, table 4), in conjunction with the multiple sequence alignment program Cluster Omega (The European Bioinformatics Institute EMBL-EBI, 2017) to identify the LBDs (Appendix 1, table 4). The quality of the structures built were evaluated using the workspace at SWISS-MODEL and were assessed using the quaternary structure coordinates of the template QMEAN. The model structures were further validated and assessed using the SWISS-MODEL server tools – the global and per residue model quality was assessed using the Qmean scoring function (Benkert *et al.* 2011), packaging

quality was evaluated using Anolea (Melo and Feytmans 1998) and Ramachandran plots (Uchoa *et al.* 2004) to analyse the protein structure and the quality of the model (Laskowski *et al.* 1993).

2.1.2 Ligand selection

A range of potential/known ER binding ligands were evaluated in this study. These included naturally-occurring (endogenous), synthetic (drug) and exogenous (dietary) compounds with agonist/antagonist ER binding properties. The SMILES codes for the ligands were downloaded from the phenol database (http://phenol-explorer.eu) and input structures were generated by Open Babel (O'Boyle *et al.* 2011) and Cactus software (http://cactus.nci.nih.gov/translate/) and from Zinc docking database alongside those generated using Open Babel. The structure of the ligands estradiol (ER agonist) and 4-hydroxyTamoxifen (ER antagonist) were obtained from the X-ray crystallography data available at the PDB databank described above. These were compared with the ligand structures produced and retrieved ligand structures from downloaded *in silico* experiments. The ligand structures were not subjected to energy minimisation as the docking software used in downstream experiments accommodates this.

2.1.3 Locating ligand binding site(s)

The ligand binding sites were predicted using the meta-server approach to protein-ligand binding site prediction COACH (Consensus approach TM-SITE and S-SITE with other structure-based programs) (Yang, Roy, and Zhang 2013) which utilises two comparative methods: specific substructure binding comparisons (TM-SITE) and sequence profile alignments (S-SITE), finally

32

combining with other structure based software to generate a consensus of the possible binding sites for a given ligand and enabling the optimal coordinates to be localised for the grid box for the docking experiments.

2.1.4 Protein, ligand input file preparation and grid parameters

Docking experiments utilised PDBQT files for both ligand and receptor. The ligand file encodes a torsion tree and both files display hydrogen atoms, partial charges and non-polar hydrogen merged.

For the macromolecule preparation; firstly, all hydrogen atoms were removed, then polar hydrogen atoms were added and restored from the ERs PDB files. The AutoDock tool (ADT) function was used to check for the net charge of the macromolecule and Gasteiger charges were added to each atom, the files were then saved as PDBQT files. For ligand preparation, hydrogen atoms (polar and non-polar) were added, ADT was used to compute the Gasteiger charges, calculate and define the rotatable bonds and to estimate the number of torsional degrees of freedom, again, the files were saved in the PDBQT format.

For AutoDock Vina docking it was necessary to define the centre coordinates x, y, z and the sizes of the molecules of the grid location in the PDB file (File containing atomic coordinates. The coordinates were obtained using the "Grid Widget" in ADT, a three dimensional box was created whereby the binding pocket was located, obtained by literature review and by PBD ER-LBD files and COACH server consensus. Grid volumes were generated encompassing the

space of the ligand binding pocket but ensuring sufficient room for free-movement of the ligand. For convenience some command line options were written in a configuration file.

2.1.5 Computer simulated ligand binding

AutoDock Vina was used as the docking software (Trott and Olson 2010a) and was run from the command prompt in a default directory; all the files (receptor.pdbqt, ligand.pdbqt, configuration.txt) were saved in the same folder. The command was "\vina.exe --config configuration.txt –log ligand.txt, and two additional files were generated one with the docking coordinates and a txt file with output. Data generated by the software was analysed based on the binding affinity predictions with a maximum of 10 binding modes generated per analysis.

CHAPTER THREE

3. ERα and ERβ in breast cancer

3.1 Introduction

Receptors for 17β-estradiol (estradiol) are present in many cell types and are found on the epithelial cell membranes of the breast (in the form of G-protein coupled receptors) and in the intracellular space where they function as nuclear receptors (Kurokawa *et al*. 1995). In breast cancer (BC) most of our understanding of ER relates to the function of the intracellular receptors ERα and ERβ. The downstream molecular components activated following the dimerization of ERα and ERβ play an important role in BC development and disease progression (Palmieri *et al*. 2002). Elucidation of the biological role of ER has led to the development of important classes of drugs that prevent the normal function of estradiol, blocking BC cell proliferation through the ER. Drugs acting as estrogen antagonists such as Tamoxifen have led to significant improvements in patient survival, for example as shown in the large studies undertaken by the Early Breast Cancer Trialists' Collaborative Group (EBCTCG, 2017). The success of anti-estrogen drugs illustrates the importance of strategies aimed at targeting the ER as both anti-BC (preventative) as well as treatment modalities (Furr and Jordan 1984; Howell 2006; Jiang, Zheng, and Wang 2013).

As described in Chapter 1, ERs are involved in complex pathways as a result of the binding of endogenously produced estradiol. In addition, exogenous substances that mimic or interfere with the action of estradiol can also lead to a

wide range of outcomes in terms of the ER related signalling pathways and transcription events that are affected.

ERα and ERβ (Fixemer *et al.* 2003) are encoded on different chromosomes by the *ESR1* and *ESR2* genes respectively (Barnes, 2010). ERα and ERβ show structural similarities and their affinity to estradiol is similar for both the wild type and subtypes of the receptors. However, each ER type displays a different profile in terms of the pathways that are activated (Marino *et al.* 2006). Further complexity arises as ERα and ERβ can be co-expressed and may act as homodimers or heterodimers. Variants in both ERα and ERβ have been described.

ERβ is an important mediator of estradiol action in BC and has been shown to be co-expressed with ERα (Green *et al.* 1986; Mosselman, Polman, and Dijkema 1996) but currently it is only ERα that is used for the classification of ER status in BC. There remains the potential to classify BC according to both ERα and ERβ status and the ERβ receptor may also be a novel drug target in BC. Further knowledge of ERβ biology may result in an improved understanding of BC biology, prognosis and potentially lead to new approaches to BC treatment.

Computational methods for protein structure prediction and ligand binding interaction are widely used in drug discovery programmes. Modelling 3-D protein structures for applications in structure based studies is an important

prerequisite for the study of protein dynamics and interactions with ligands and offers the potential for screening large numbers of compounds *in silico* (Schmidt, Bergner, and Schwede 2014). Information related to the spatial arrangement of amino acids and consequent protein structure allows the design of further experiments around ligand-receptor interactions, for example ligand docking. Methods for prediction of protein structure such as homology modelling are frequently used as they overcome the need for direct work with proteins and the associated obstacles that occur related to gene cloning, protein expression and purification and also circumvent issues around protein crystallisation. Homology modelling does, however, rely on high quality crystal structures of proteins from the same family and with similar amino acids sequences, similar structures and folds (Schmidt *et al*. 2014) to use as templates in order to generate reliable models (Vyas *et al*. 2012).

3.2 Aims

The aim of the work described in this chapter was to identify the ERα and ERβ ligand binding domain (LBD) and assess the use of this as a template for homology modelling studies with ER binding drugs Tamoxifen and Fulvestrant. In order to achieve these aims, the following objectives were considered:

3.2.1 Objectives

- Selection of human ERα and ERβ crystal structures from the PDB database, determining the most representative structure for agonist and antagonist conformations.

- Building a 3-D homology model and validation of the LBD for wild type ERα and ERβ.

- *In silico* docking of wild type ERα and ERβ with different ligands (E1, E2, E3, Tamoxifen, Fulvestrant).

3.3 Results

3.3.1 Homology modelling

The homology modelling was based on structures that had been identified using the sequences of the canonical forms of ERs retrieved from Universal Protein Resource (Uniprot) in FASTA format (ERα accession number P03372, ERβ accession number Q92731). This study investigated the ligand binding domain (LBD) of the ERα and ERβ receptors. The following criteria were used for the selection:

- Structures where the ER had been evaluated as a complex with a range of exogenous and endogenous ligands.

- Structures with the highest possible resolution - to ensure high quality X-ray crystallography based structures.

- Avoidance of crystal structures showing mutations - in order to avoid possible aberrant conformations.

301 structures of the human estrogen receptor are available in the Research Collaboratory for Structural Bioinformatics (RCSB), only 94 displayed a reasonable good resolution (under 2.5 Å). The top ten structures with the best resolution and only in complex with one of the ligands (estradiol and Tamoxifen) were selected. After the structures with mutations were excluded as well, 3D structures were retrieved from the PDB database, RCSB protein database (URL www.rcsb.org) (Berman *et al*. 2000) as shown in below:

ERα agonist   PDB ID: 1GWR; 2.4 Å (Warnmark *et al*. 2002)
ERα antagonist   PDB ID: 3ERT; 1.9 Å (Shiau *et al*. 1998)
ERβ agonist   PDB ID: 2J7X; 2.1 Å (Pike *et al*, unpublished)
ERβ antagonist   PDB ID: 2FSZ; 2.2 Å (Wang *et al*. 2006)

3.3.1.1   Wild type ERα

The homology model for wild type human ERα produced using PYMOL and

based on the PDB crystal structure PDB ID 1GWR is shown in Figure 3.
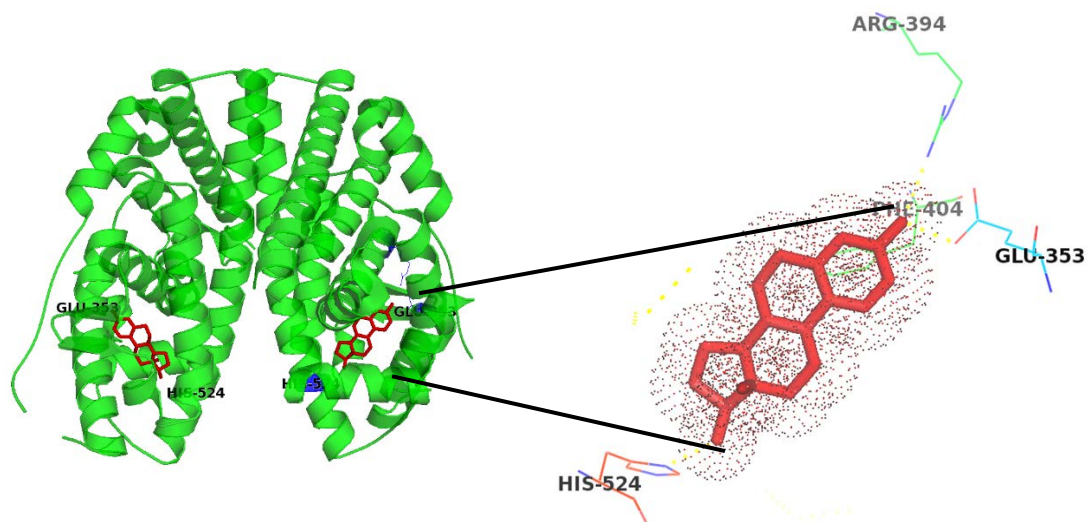
Figure 3. ERα (ESR1) ligand binding domain. On the left, structure of human
wild type ERα ligand-binding domain homodimer complexed with estradiol taken
from the PDB database (based on PyMOL rendering of PDB 1GWR). On the
right, zoom view of the ERα/estradiol complex: showing key amino acid residues
(GLU-353, ARG-394, PHE-404 and HIS-524) in the binding pocket that interact
with estradiol when complexed with ERα.

Previous studies have described two key domains in the structures of the ER

proteins, the ligand binding and the DNA binding domains. In this study the LBD

was chosen for the structure-based studies and used as a template for the

homology modelling process. The key interactions between the amino acid

backbone and the ligand(s) were analysed using Pymol version 1.7.4

(Schrodinger LLC 2015) and PMV tools (Weigelt *et al.* 2010) (Figures 3 and 4).

For ERα wild type 3D structures (PDB IDs: 1GWR estradiol/ERα, 1X7R Genistein/ERα and 3ERT 4-OHT/ERα) and for ERβ wild type (PDB IDs: 4J24 estradiol/ERβ, 1QKM Genistein/ERβ and 2FSZ OHT/ERβ) were retrieved from the protein database. The PDB files were used in the different stages of the structure-based study. Multimeric complexes were reduced to the single amino acid chain structures from the corresponding PDB files using the MGL tool as the interface and editor.
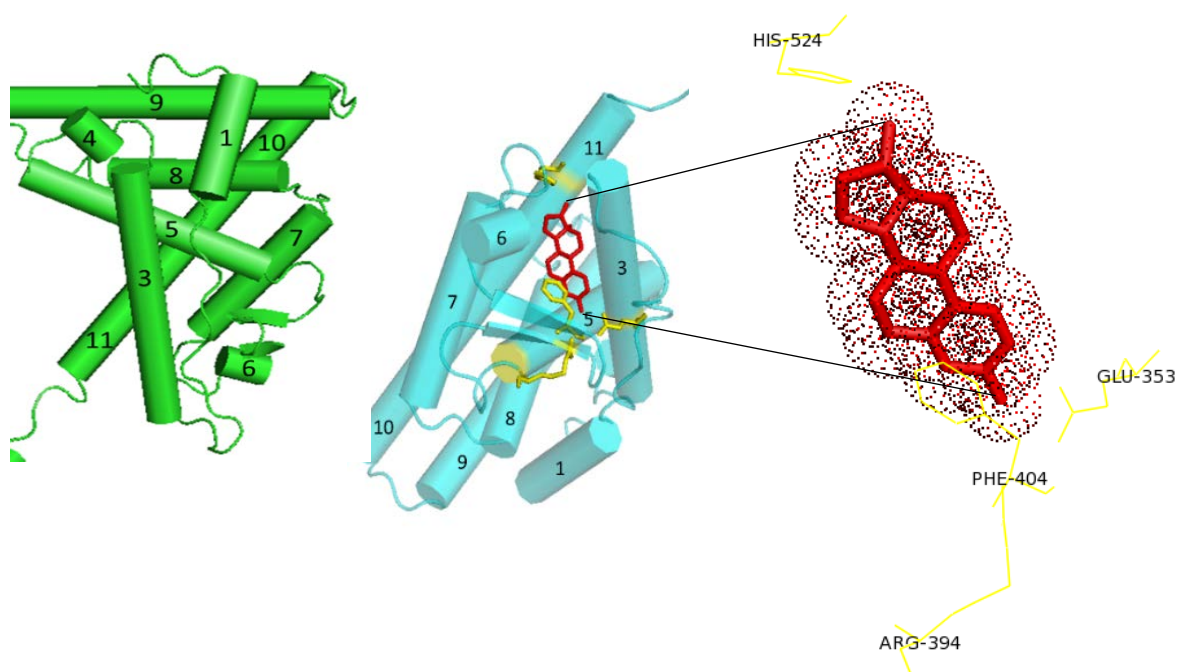
Figure 4. ERα (ESR1) ligand binding domain (LBD) homology model (SWISS-MODEL) showing the amino acid residues that interact with estradiol, recapitulating the information shown in figure 3. On left: cartoon cylindrical helix representation (helices 1 to 12).In the centre, LBD shown in cartoon cylindrical helix (cyan) representation and estradiol (red stick) and on the right a zoom view showing the key amino acids (yellow) interacting with estradiol.
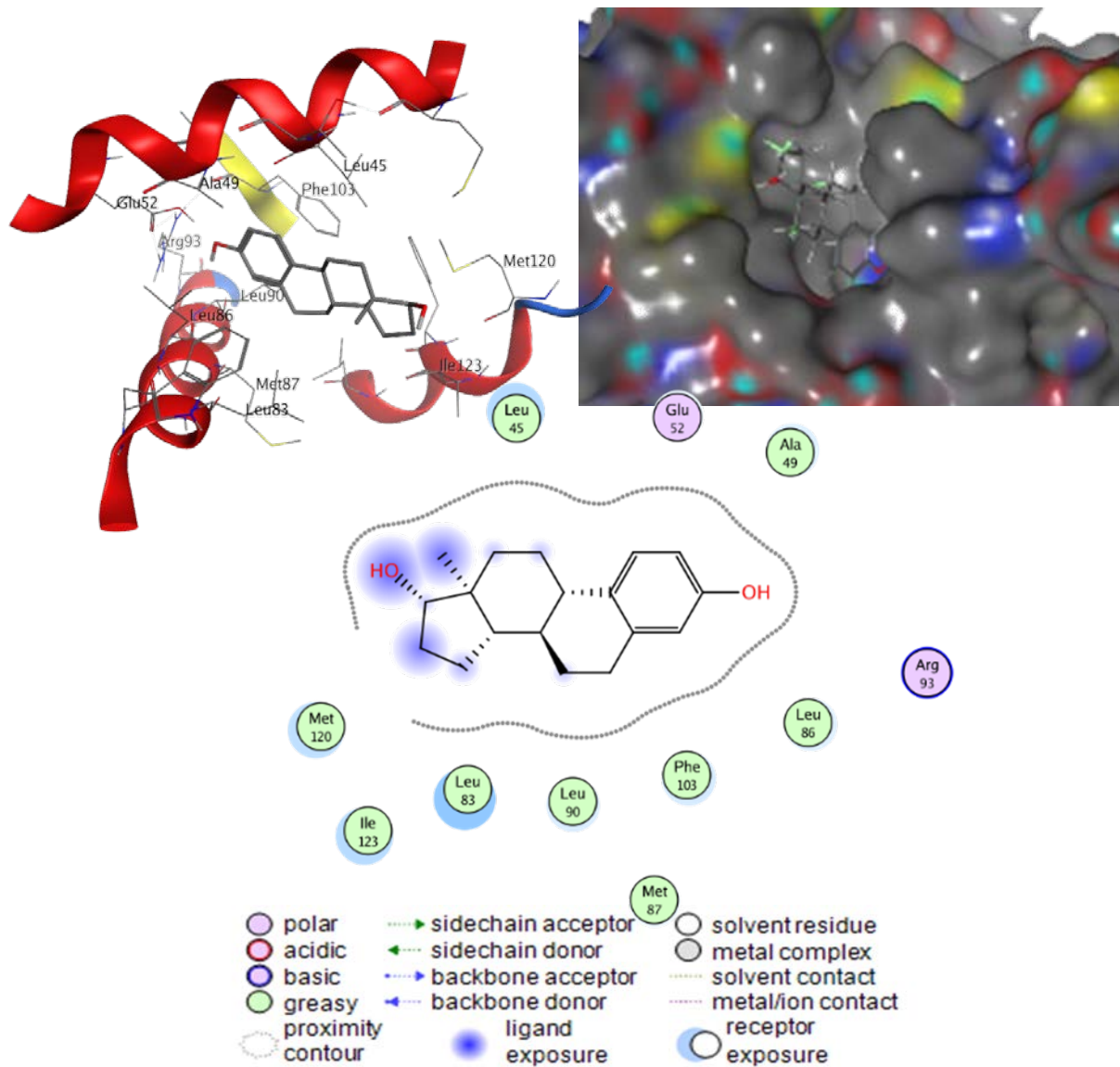
Figure 5. ERα (ESR1) ligand binding domain homology model (SWISS-MODEL). Top left hand, predicted ligand binding pocket in complex with estradiol based on analysis using the RaptorX binding prediction server, the highest scoring residues predicted to lie within the binding site for sequence segment [1, 252]:1): M42 L45 A49 E52 L83 L86 M87 L90 R93 F103 M120 I123 G220 H223 L224 Possible ligands: EST, OHT. Top right hand, surface representation of wild type ERα LBD in complex with estradiol (red). Lower panel: 2D layout of the key amino acids in the ligand binding pocket that retain estradiol.

Identification of the LBD is an important aim for structure-based studies and docking experiments but *in silico* approaches may be inaccurate in this regard.

Methods have been developed to increase the accuracy of predictions for

example, by adopting specific structure comparisons and sequence alignment through the use of the meta-server COACH (Yang *et al.* 2013).

In this approach, coordinates for the amino acids described by the binding site prediction meta-server were identified enabling generation of an accurate grid box for use in the AutoDock Vina system (Trott and Olson 2010b). The LBD was identified by multiple sequence alignment using CLUSTAL-OMEGA (EMBL-EBI 2013) and quality of the models assessed by the SWISS-MODEL server Qmean Z-score providing data about the absolute quality of the predicted homology models (Figure 5).

3.3.1.2 Wild type ERβ

Wild type human ERβ is smaller than ERα but they show a high homology in the DNA binding domain (96 %). However, the structures are only 57% similar at the LBD (ERα amino acid residues 302 to 552 and 255 to 504 in ERβ (Kuiper *et al.* 1996). Whilst many of the studies undertaken with ER in BC have focussed on ERα, the ERβ has been reported to be an important partner acting as a homodimer and heterodimer with ERα (Cowley *et al.* 1997).

In this study, wild type ERβ was evaluated using PYMOL, basing the PYMOL rendering on the PDB crystal structure 4J24 for ERβ as shown in Figure 6.
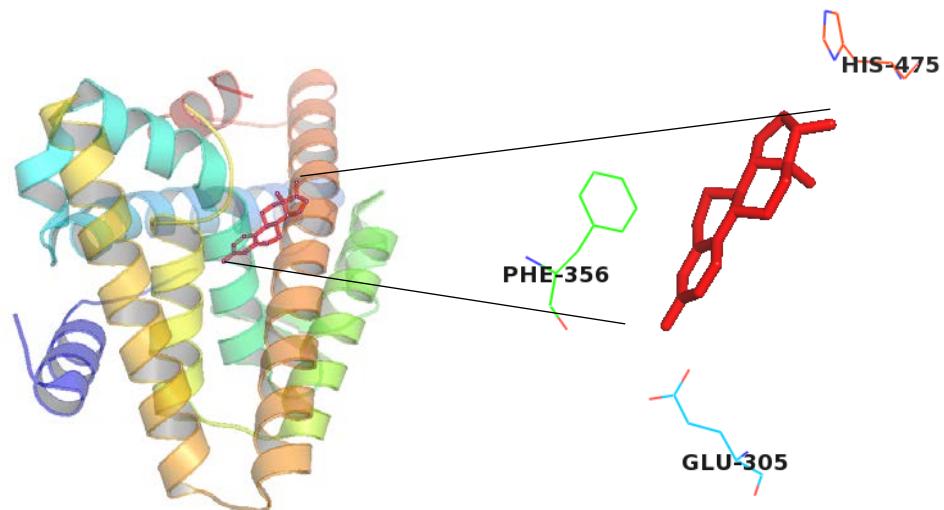

Figure 6. ERβ (ESR2) ligand binding domain. On the left, structure of the human wild type ERβ ligand binding domain in complex with estradiol taken from the PDB database (based on PyMOL rendering of PDB 4J24). On the right, zoom view of ERβ/estradiol complex showing the binding pocket for estradiol.

Although the crystal structure of wild type ERs rendered using PyMol (as shown above) are often used as targets in the next step of the work flow for *in silico* experiments, the approach taken here was to use the SWISS-MODEL server and produce a new model enabling the PyMol rendered and the SWISS-MODEL structures to be compared (Figure 7).
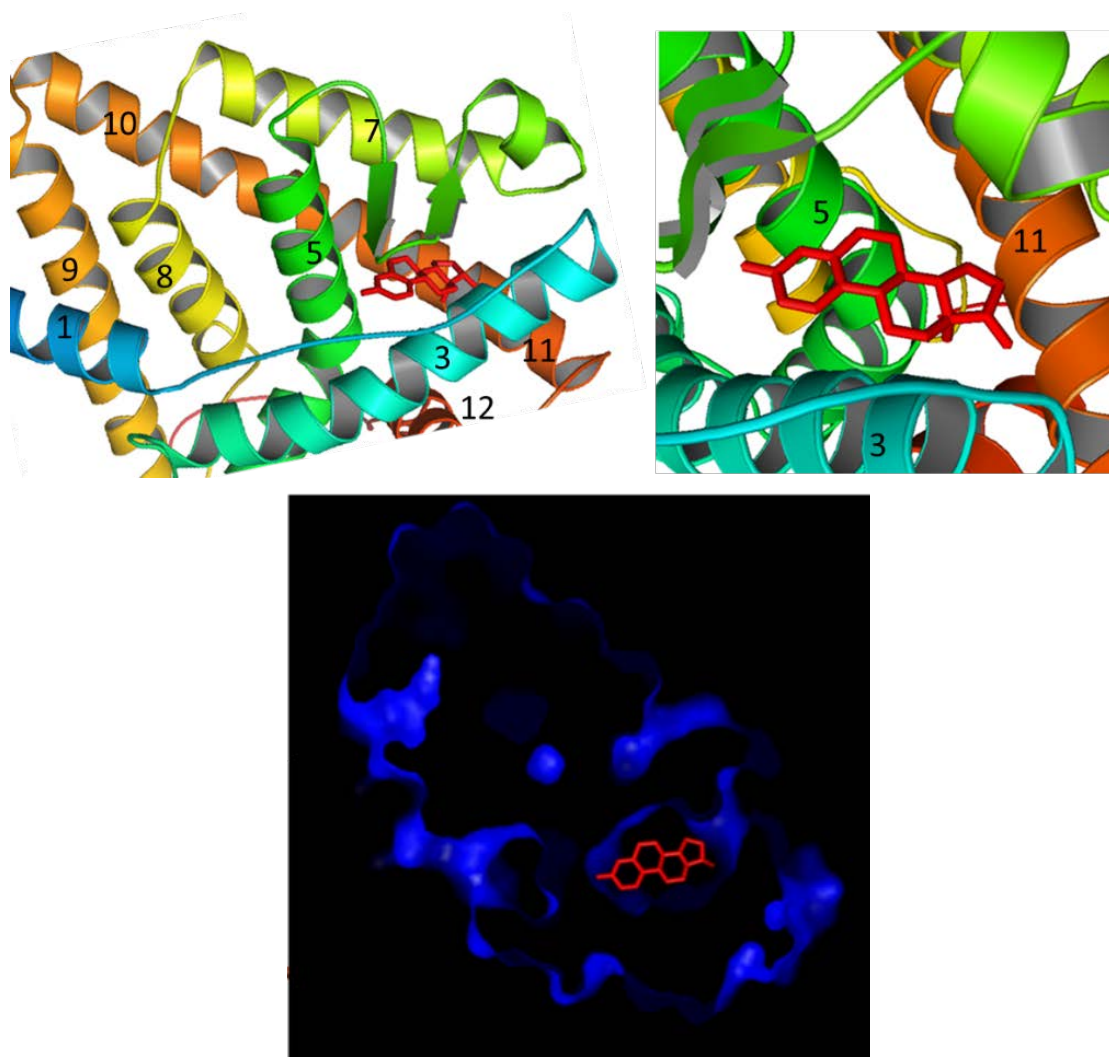
Figure 7. ERβ (ESR2) ligand binding domain (LBD) homology model (SWISS-MODEL). Upper left: cartoon showing cylindrical helix representation (helices 1 to 12). Upper right: zoom view of upper left showing the LBD in cartoon with cylindrical helices and estradiol (red stick representation) and helix number indicated (black). Lower panel: ERβ surface structure (blue) and estradiol (red) in the ligand binding site. After, re-docking the ligand (E2) AutoDock Vina was able to emulate a near-native docking pose.

45

The ERs have been widely studied and although the binding pocket has been described for many endogenous ligands and drugs it has been less well studied in terms of the binding of other compounds for example those with either agonist or antagonist binding properties found in the diet or the environment. The *in silico* methods described and utilised here enabled investigation of the amino acids in the binding pocket of the ER and the availability of these coordinates allowed the ER to be investigated using other *in silico* structure based studies (Figure 8).
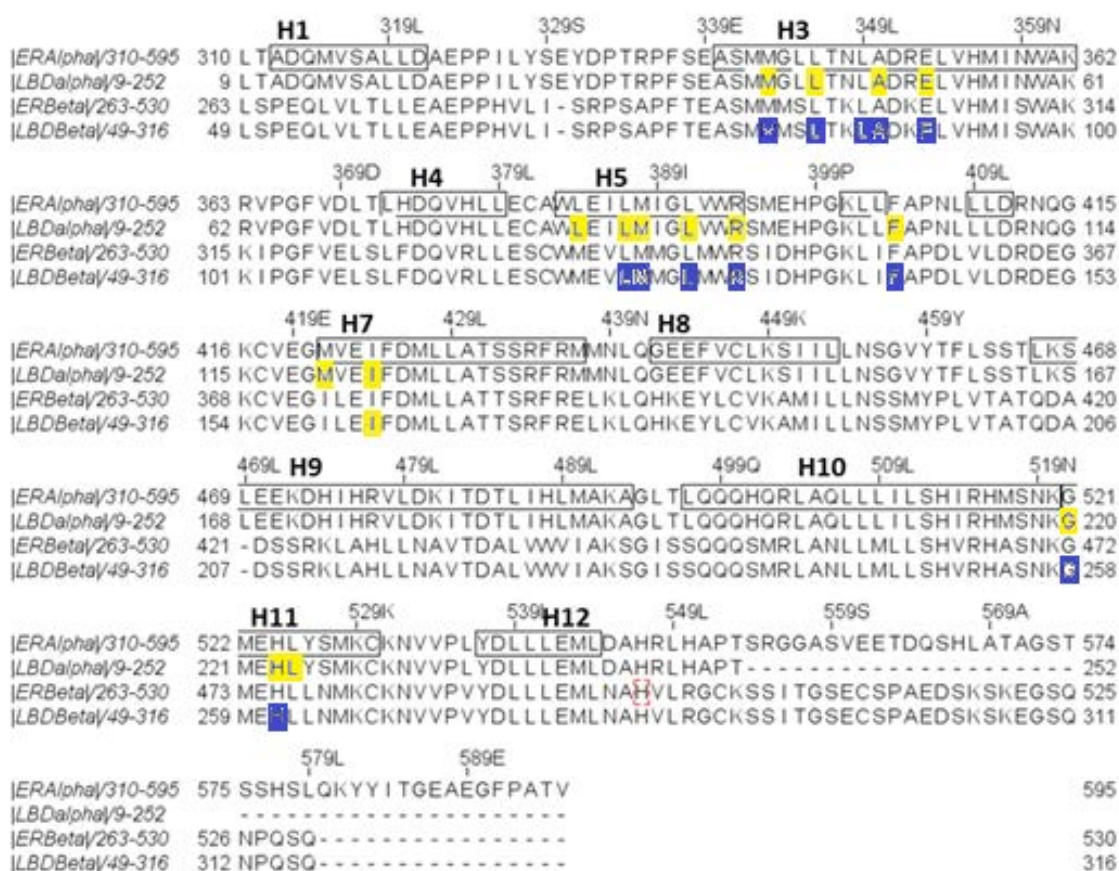
Figure 8. Sequence alignment ERα and ERβ. Residues that belong to helices in the ERα are framed by a black rectangle (Number of the helices are provided in the top of each one in instance H1, H2 and so on) in yellow boxes are the key estradiol binding residues in ERα (M42 L45 A49 E52 L83 L86 M87 L90 R93 F103 M120 I123 G220 H223 L224) and in blue for ERβ (M81 L84 L87 A88 E91 L125 M126 L129 R132 F142 I162 G258 H261 L262). Identifying these residues provides an understanding of the "grid" or "space" needed for investigation of other ligands in later docking studies. Each receptor was run twice in the multiple alignment one with the respective residue number in the full length protein structure and the second one with the residue number corresponding in the LBD.
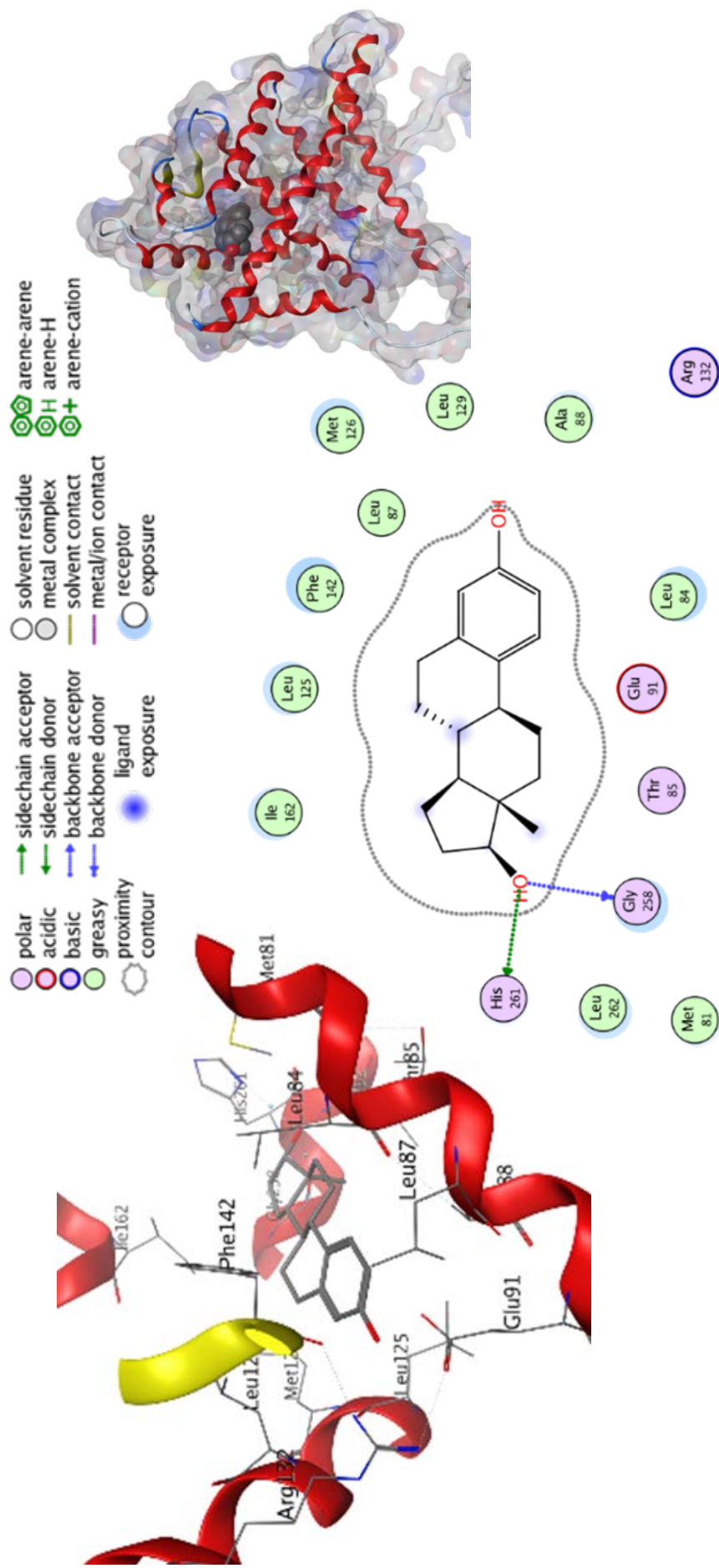
Figure 9. ERβ (ESR2) ligand binding domain homology model (SWISS-MODEL). Top left hand, predicted ligand binding pocket in complex with estradiol based on analysis using the RaptorX binding prediction server, the highest scoring residues predicted to lie within the binding site for sequence segment [1, 252]:1) M81 L84 L87 A88 E91 L125 M126 L129 R132 F142 I162 G258 H261 L262.. Possible ligand: estradiol. Top right hand, surface representation of wild type ERβ LBD in complex with estradiol (grey). Bottom centre, 2D layout of the key amino acids in the ligand binding pocket that retain estradiol.

3.3.2 Physiological and antagonist ligands for ERα and ERβ
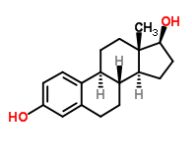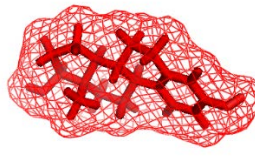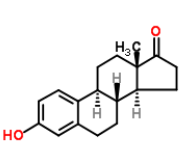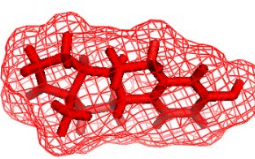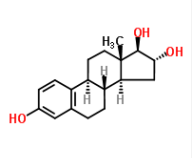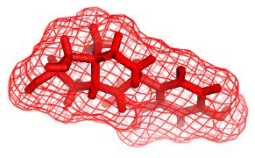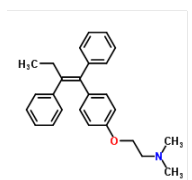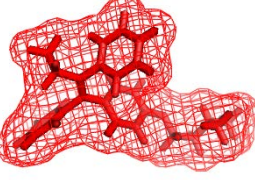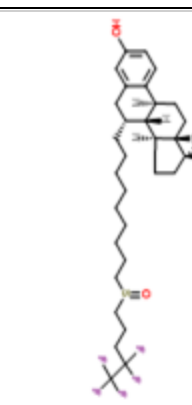
This study was concerned with the comparative docking of three endogenous estrogens and two anti-breast cancer drugs.

To undertake this study it was necessary to evaluate the ER structures built *in silico* and their predicted ER binding sites, calculated spatial position of the ligand and amino acid environment needed for successful docking of these agonist and antagonist structures.

The ligand structures were downloaded from the zinc docking database (http://zinc.docking.org) and input structures were generated using Open Babel (O'Boyle *et al.* 2011). Energy minimization was performed for each of the ligand structures in turn as the next (docking) step using AutoDock Vina was known to apply partial atomic charges.

As can be seen in Table 1, one of the key differences between the ER agonists (estradiol, estrone, and estriol) and antagonists (Tamoxifen, Fulvestrant) is their molecular size. The agonists were of a lower molecular weight compared with the antagonists which exhibited a more bulky structure accompanied by larger side chains attached to the core steroid molecules.

Table 1. ER ligands. In the Simplified Molecular-Input Line-Entry system (SMILE) code strings. 2D and 3D structures were produced using the Open Babel software(O'Boyle *et al* 2011); the molecular mass was obtained from the ChemSpider database (http://www.chemspider.com).

Ligand	SMILE Code	Mass (Da)	2D Structure	3D structure
Estradiol	C[C@]12CC[C@@H]3c4cc(cc4CC[C@H]3[C@@H]1CC[C@@H]2O)O	272.4		
Estrone	C[C@]12CC[C@H]3c4ccc(cc4CC[C@@H]3[C@@H]1CCC2=O)O	270.4		
Estriol	C[C@]12CC[C@H]3c4ccc(cc4CC[C@@H]3[C@@H]1C[C@H]([C@H]2O)O)O	288.4		
Tamoxifen	CC/C(=C(\c1ccccc1)/c2ccc(cc2)OCC[NH+](C)C)/c3ccccc3	371.5		
Fulvestrant	C[C@]12CC[C@@H]3c4cc(cc4C[C@H]([C@@H]3[C@@H]1CC[C@@H]2O)CCCCCCCC[S@](=O)CCC(C(F)(F)F)(F)F)O	606.8		
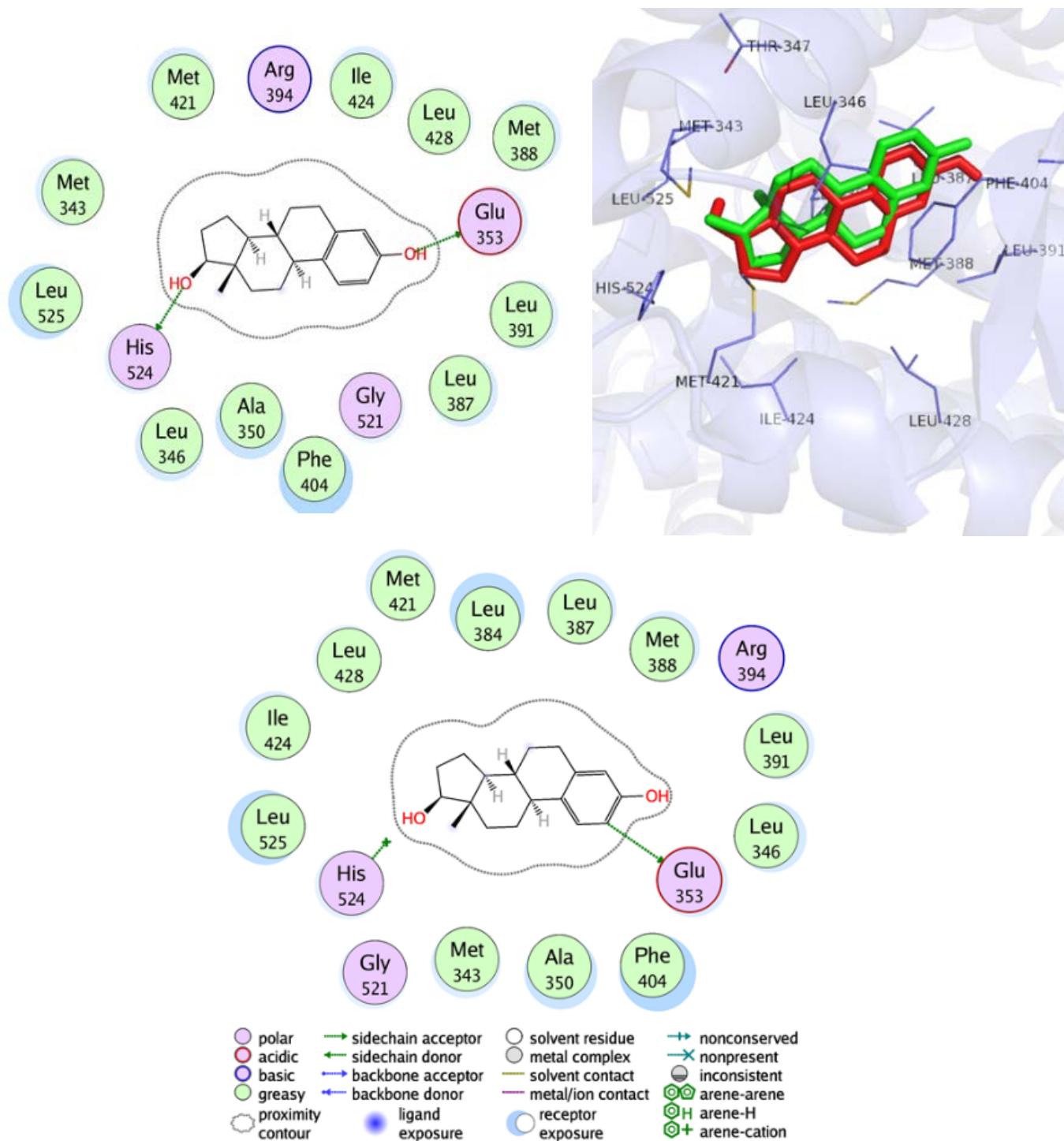
Figure 10. Comparison of the results of docking estradiol to ERα (ESR1) with the crystal structure of ERα: estradiol complex. Upper left: 2D representation of binding pocket with estradiol based on the crystal structure (taken from the PDB database PDB ID: 1GWR). Upper right: 2D representation of binding pocket and estradiol based on the AutoDock Vina pose. In green the ligand in the binding pocket from the crystallography, in red the ligand in the binding pocket from the AutoDock Vina. Lower panel: 2D representation of the

binding pocket and ligand based on the pose from the AutoDock Vina showing the lowest binding energy (-10.9 Kcal/mol).

The model developed for estradiol binding to ERα using the docking approach with AutoDock Vina, above, closely recapitulated the X-ray crystallography experiments. *In silico* experiments were run with AutoDock Vina three times to compare consistency of the data, each time the exhaustiveness or number of runs were set up by default (10) models were also developed for estrone and estriol binding and these are shown in Appendix 2 Figure 3 and 4 .

Ligand-binding interactions in ERα-estradiol complex show GLU353 aa and HIS524 forming hydrogen bonds in both sides of the ligand with the hydroxyl group and ARG394 support the location of GLU353 and PHE404 (Tanenbaum *et al*. 1998) . It was clear from this analysis that molecules of larger molecular sizes would not be accommodated in the ligand binding site (as agonists). Accordingly, a second template model was produced to enable the interaction between antagonists and the ER to be evaluated. In this approach, again, the crystal structure of ERα in complex with 4-hydroxyTamoxifen was taken and used as the template for the homology modelling as shown in Figure 11.

4-hydroxyTamoxifen (Figure 11) and Fulvestrant interact with ERα binding to a larger groove compared with the interaction between the agonists, for example estradiol (Figure 7), estrone and estriol (Appendix 2, figure 3 and 4). The following amino acids of the ERα receptor backbone anchored both 4-Hydroxytamoxifen and Fulvestrant in the same place as estradiol; GLU353 and ARG394 play an important role and also ASP351. The experiment was

repeated using ERβ as shown in Appendix 2 (figure 20), ARG346 and GLU305 play the same role as in its homologous part in the ERα (ARG394, GLUE353 and ARG394) forming hydrogen bonds with the Tamoxifen.
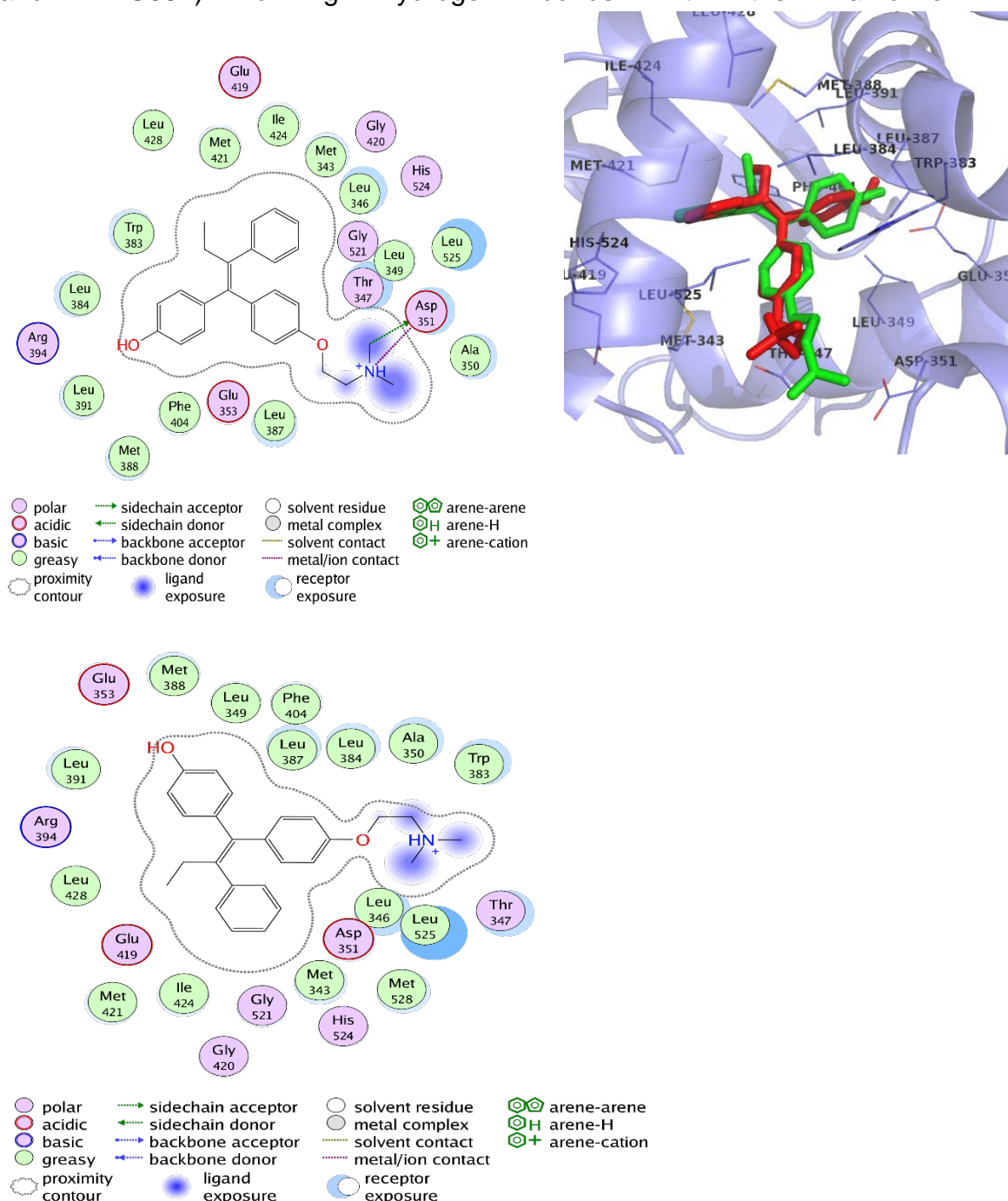
Figure 11. 4-hydroxyTamoxifen binding to ERα. On the left: 2D representation of binding pocket and ligand reported from crystallography experiments (taken from PDB database 3ERT). On the right: 2D representation of binding pocket and ligand based on the AutoDock Vina pose. Lower panel: cartoon representation of the amino acid interactions, in green the ligand from the crystallography, in the red the ligand in the pose produced by AutoDock Vina at the lowest binding energy (-9.3 Kcal/mol).

3.4 Discussion

Over the last three decades different computational approaches have been adopted as prediction methods for *in silico* evaluation of receptor-ligand interactions (Schmidt *et al.* 2014). The methods developed have become increasingly complex necessitating a greater number of permutations and combinations of variables to be assessed in the computational space pre-calculating the interaction of each atom type Vina uses an united-atom scoring function (only heavy atoms) also Vina does not take into account partial charges since internally it uses electrostatic interactions (Trott and Olson 2010b). Computer processing power has vastly improved and modelling/docking approaches may now be undertaken by applications that utilise either single desktop computers or by access to clusters or computer-grids (Forli *et al.* 2016). As a result of these improved methods we have now reached the point whereby reliable models can be produced and used for *in silico* experiments of receptor-ligand interactions (Trott and Olson 2010b), such as has been used here for the wild type forms of monomeric ERα and ERβ. The fact that we pre-calculated the binding site helped to reduce the grid size making it easier for the docking program to explore it.

In this study the *in silico* modelling of wild type ERα and ERβ was first attempted using the full length protein (data not shown), this proved unsuccessful due to the length and complexity of the full length structure and the absence of an X-ray crystallography-derived template. Homology modelling was then performed using monomeric forms of the X-ray crystallography derived coordinates of the

LBD (human wild type ERα and ERβ in *E. coli* based expression systems); the arrangement of the α-helices was very similar both in terms of the number of helices and their position/orientation. Importantly α helix 12 which has been reported to extend away from the monomer and on ligand-binding has been shown to interact with the partner monomer to stabilise homodimeric and heterodimeric forms of ER and was found to adopt a similar configuration in the homology models produced here.

When the docking of agonists was considered using AutoDock Vina the key amino acids were found orientating and positioning estradiol in the binding pocket. Tanenbaum *et al. (*1998) have described the crystal structure of the LBD of human ERα and highlighted the importance of the following amino acid residues: Glu-353 for anchoring the 3-hydroxyl group of estradiol and its analogues (in contrast to the 3-keto group found in other steroid hormones); the "bracing" of the side chain of Arg-394 by a hydrogen bond to the carbonyl residue of the amino acid preceding Phe-404; the presence of hydrogen bonding between the 17-OH of estradiol and His-524. These four key amino acids were all observed at the correct positions with respect to estradiol in the binding pocket following the AutoDock Vina experiments though in this case the binding site can be pre-calculated due to the extensive research into the receptor with AutoDock Vina which has displayed good performance with a bigger grid increasing the number of runs per experiments (exhaustiveness).

The binding of the anti-breast cancer drug 4-Hydroxytamoxifen to human ER has also been evaluated. This important antagonist functions by preventing estradiol occupying the LBD of ER and giving rise to a conformational change on the surface of the ER hindering the interaction of co-activator peptides, such as the NR-box II peptide (Shiau *et al*. 1998). The orientation and anchoring of 4-Hydroxytamoxifen following the docking experiment with AutoDock Vina is shown as an overlay in Figure 11 illustrates the success of this approach.

As it was not possible to model the full length receptor, potential interference by other domains of the ER, e.g. the DNA binding domain, activation factor domain, etc., with respect to agonist and antagonist binding this remains an area for further future exploration. It is important to note, however, that this consideration also applies to data from the X-ray crystallography based experiments with these receptors as well. Despite the inherent shortcomings associated with studying the human ERα and ERβ LBD alone, the system described here recapitulated the important amino acid residues of the LBD with ligand poses and interactions commensurate with those reported for both the agonists and the antagonists of ERα and ERβ in the X-ray crystallography experiments.

The next phase of this work sought to use these homology models to study the interaction between dietary phytoestrogens and ERα/ERβ.

CHAPTER FOUR

4. *In silico* docking experiments with phytoestrogens and wild type human ERα and ERβ LBD

4.1 Introduction

The variation in breast cancer incidence and survival across the globe has led to an interest in the potential protective effect of dietary phytoestrogens, particularly as countries where the diet is enriched in phytoestrogens have lower breast cancer incidence and greater survival rates. Phytoestrogens are found in the greatest abundance in soya products and other legumes (Mazur 1998), seeds, whole-grains, fruit, vegetables and beverages (Kuhnle *et al.* 2008; Kuhnle *et al.* 2009) and are classified into two main groups: isoflavones and lignans. Meta-analyses have reported a decreased breast cancer risk with increased lignan and isoflavone exposure in Asian but not in Western populations (Buck *et al.* 2010; Dong and Qin 2011; Wu *et al.* 2008).

Phytoestrogens elicit a multitude of effects *in vivo* acting as naturally-occurring selective ER modulators via their ER binding properties (Mueller *et al.* 2004). The pharmacokinetic properties of these compounds have led to concerns that increased exposure to dietary phytoestrogens might promote breast cancer recurrence and/or affect the efficacy of hormone therapies. At present, studies concerned with unravelling the effect of phytoestrogens have shown that they can reduce tyrosine kinase activity (Akiyama *et al.* 1987), inhibit signalling pathways (Kim, Peterson, and Barnes 1998), cell proliferation (Abarzua *et al.*

2012), angiogenesis (Farina *et al.* 2006), cellular invasion and metastasis (Shao *et al.* 1998). However, a placebo-controlled study of breast cancer patients (n=114) randomised to soy-protein consumption (over a 7-30 day period) followed by tumour tissue genome-wide expression analysis reported an increase in the expression level of mRNA for genes associated with cellular proliferation (Shike *et al.* 2014), exposure of the MCF7 breast cancer cell line to the isoflavone daidzein increased the expression levels of transcription factors (Koo *et al.* 2015) and in a murine model increased isoflavone exposure resulted in an increase in the formation of lung metastases (Yang *et al.* 2015).

In the last two decades, a number of databases have been established containing considerable information about dietary polyphenols (Pasinetti *et al.* 2017)(Schweartz *et al*, 2009). One of the largest databases is Phenol-Explorer (http://phenol-explorer.eu/) (Rothwell *et al.* 2012). This resource contains information related to the polyphenol content in foods with >35,000 content values of 500 different polyphenols in >400 foods, derived from >1300 peer reviewed publications. Within the database are polyphenols with estrogen-like structures, they have been classified into either the isoflavone or lignan groupings.

One of the main concerns around phytoestrogen consumption by breast cancer patients is the potential for interference with the binding of Tamoxifen (and other drugs) to ER. A further consideration is whether known splice variants in the ER are associated with different binding properties of drugs for example Tamoxifen,

or endogenous estrogens such as dietary phytoestrogens or endogenous ligands such as estradiol to the ER.

The DietCompLyf study is a multicentre prospective cohort study investigating diet, lifestyle and complementary therapy use of UK breast cancer patients. It is the most comprehensive UK-based study concerned with diet and breast cancer progression to date and includes 3144 breast cancer patients with >5 year follow up data. The study is being coordinated within the Department of Biomedical Sciences and is concerned with effects of the isoflavones genistein, daidzein, glycitein, biochanin A and formononetin, and the lignans secoisolariciresinol, matairesinol, shonanin, and their metabolites enterodiol, enterolactone and equol on breast cancer progression. Formononetin and biochanin A are methyl forms of isoflavones that are often found in dietary supplements and these can be metabolised to daidzein and genistein, Daidzein itself can also be metabolised to O-desmethylangolensin (ODMA) and to equol by the intestinal microbiome.

To evaluate the potential role of these dietary chemicals with respect to breast cancer aetiology, the binding properties of these phytoestrogens to the LBD of ERα and ERβ was assessed using the homology models developed as described in Chapter 3.

4.2 Aim

The aim of the work described in this chapter was to develop and analyse, *in silico*, the interaction between dietary phytoestrogens (assessed in the DietCompLyf study) and the LBD of wild type human ERα and ERβ. A secondary aim was to develop homology models of the LBD of variants in ERα and ERβ for later studies with these structures.

4.2.1 Objectives

- To rank the binding properties of dietary phytoestrogens to the binding of Tamoxifen and endogenous ligands to the LBD of ERα and ERβ.
- To develop homology models of different ER isoforms (variants).

4.3 Results

4.3.1 Phytoestrogen ligands for ERα and ERβ

This study was concerned with the comparative docking of dietary phytoestrogens, endogenous estrogens and two anti-breast cancer drugs (Tamoxifen and Fulvestrant). The ER structures built *in silico* (Chapter 3) were used. The SMILES codes for the phytoestrogens investigated in this study were downloaded from the phenol-explorer database described above (http://phenol-explorer.eu). Each of the input structures were generated by the use of the Open Babel (O'Boyle *et al.* 2011) and Cactus software (http://cactus.nci.nih.gov/translate/) as before. The phytoestrogens considered in this study are shown in Table 2.

Table 2. Phytoestrogens under investigation in this study. ODMA = O-desmethylangolensin.
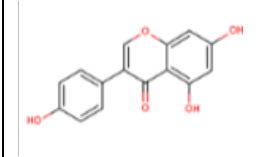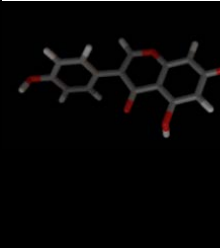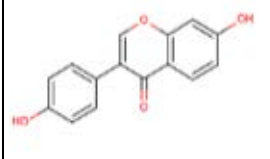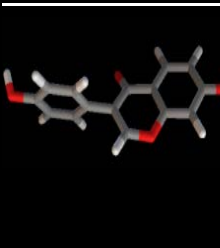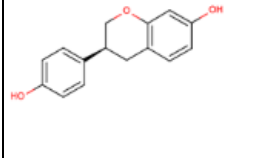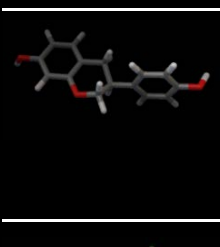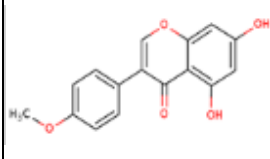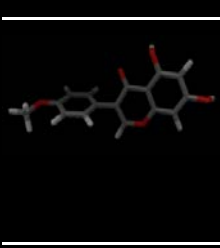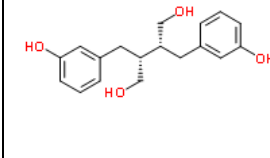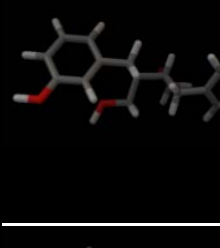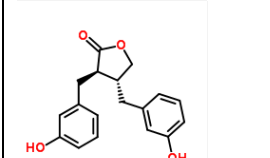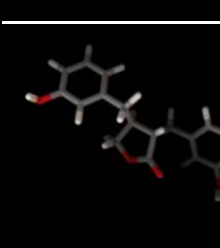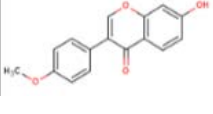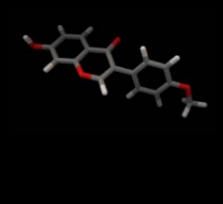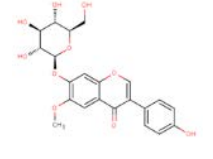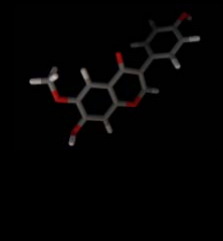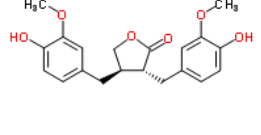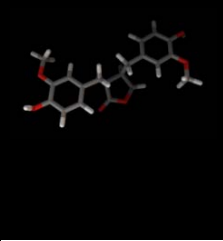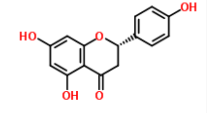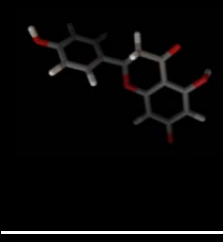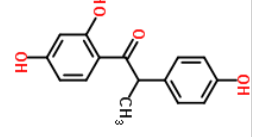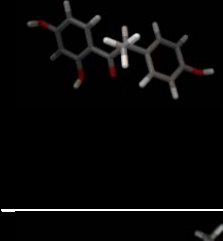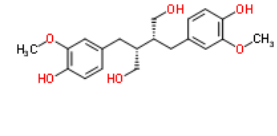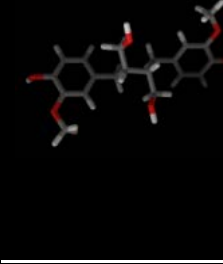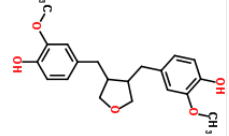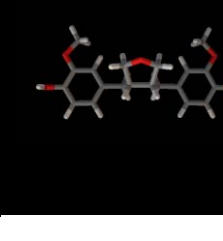
Ligand	SMILE Code	Mass (Da)	2D structure	3D structure
Genistein	c1cc(ccc1c2coc3cc(cc(c3c2=O)O)O)O	270.2		
Daidzein	c1cc(ccc1c2coc3cc(ccc3c2=O)O)O	254.2		
Equol	c1cc(ccc1[C@@H]2Cc3ccc(cc3OC2)O)O	242.2		
Biochanin A	COc1ccc(cc1)c2coc3cc(cc(c3c2=O)O)O	284.2		
Enterodiol	c1cc(cc(c1)O)C[C@H](CO)[C@@H](Cc2cccc(c2)O)CO	302.3		
Enterolactone	c1cc(cc(c1)O)C[C@H]2COC(=O)[C@@H]2Cc3cccc(c3)O	298.3		

Formononetin	COc1ccc(cc1)c2coc3c c(ccc3c2=O)O	268.2		
Glycetin	COc1cc2c(cc1O)occ(c 2=O)c3ccc(cc3)O	446.5		
Matairesinol	COc1cc(ccc1O)C[C@ @H]2COC(=O)[C@@ H]2Cc3ccc(c(c3)OC)O	358.3		
Narigenin	c1cc(ccc1[C@@H]2C C(=O)c3c(cc(cc3O2)O) O)O	272.2		
ODMA	CC(C1=CC=C(C=C1) O)C(=O)C2=C(C=C(C =C2)O)O	184.2		
Seco-isolariciresinol	COc1cc(ccc1O)C[C@ @H](CO)[C@@H](Cc 2ccc(c(c2)OC)O)CO	362.4		
Shonanin	COC1=C(C=CC(=C1) CC2COCC2CC3=CC(=C(C=C3)O)OC)O	344.4		

Phytoestrogens are a group of heterogeneous plant-derived compounds many of which are thought to be endocrine disruptors. This diverse group includes chemicals with structural similarity to estradiol, Chapter 3, Table 1, but additionally comprises chemicals containing both di- and tri-phenolic rings, Table 2.

The only structures deposited in the PDB database pertaining to the human ER and dietary phytoestrogens were complexes of ER-genistein. Indeed, of the approximately 250 X-ray crystallography/NMR structures of human ER deposited in the PDB database only six are concerned with phytoestrogens and these focus on the interaction between ER and genistein. This is rather unsurprising as genistein is found in the greatest abundance of all the dietary phytoestrogens and is found in particular in soy products and is an important dietary constituent for populations in the Far East (Lu and Anderson 1998). The crystal structure for genistein in complex with ERβ LBD, PDB ID: 1QKM (Pike *et al*. 1999) was been solved at 1.8 Å. In addition, the ER-genistein complex in conjunction with the NR box peptide has also been reported (ERα PDB ID: 1X7R, 2.0 Å; ERβ PDB ID: 1X7J, 2.3 Å) (Manas *et al*. 2004). The X-ray crystallography work described genistein as interacting with the hormone binding pocket of ERβ altering the conformation of helix-12 (Chapter 3). Upon genistein binding to ERβ, helix-12 no longer adopts the typical position observed when an agonist (for example estradiol) is bound but rather helix-12 becomes orientated in the manner observed during the binding of an antagonist (such as Tamoxifen).

In this study the interaction between genistein and the LBD of ERα was determined using the approach described in Chapter 3. The template used was ERα PBD ID: IGWR. Subsequently, the phytoestrogens shown in Table 2 were evaluated with respect to the docking into the LBD. For genistein, and as described in the X-ray crystallography (Figure 12), in this model system, the OH- of the phenolic ring was found to interact with the side chains of the amino acids Glu-353 and Arg-394, both of these residues were correctly positioned following the AutoDock Vina experiments. Similarly, the –OH of E2 showed the correct conformation - consistent with the formation of a hydrogen bond between the –OH and His-524 at the "distal end" of the binding pocket.
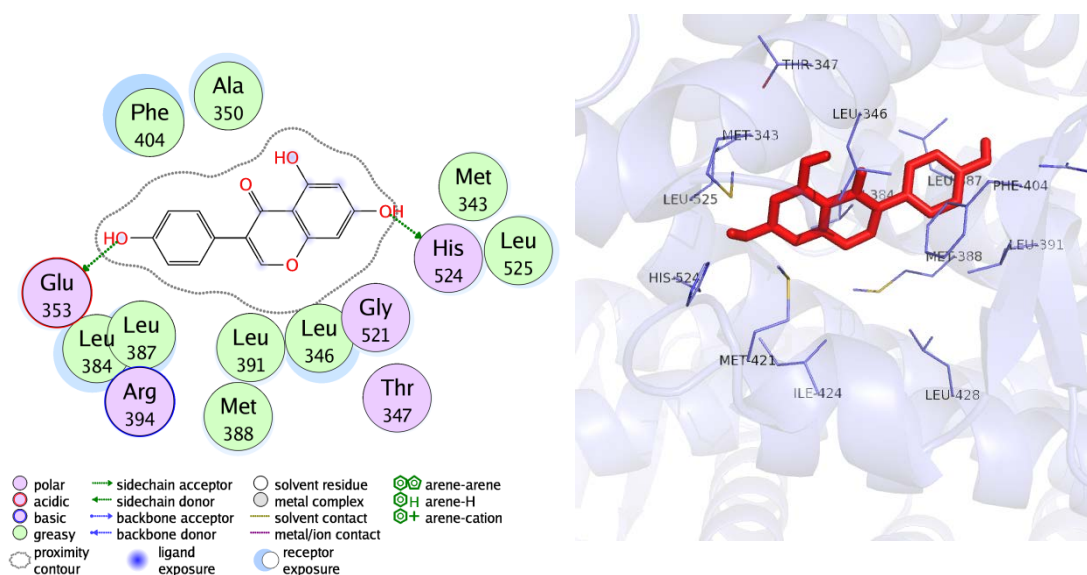
Figure 12. Genistein binding to ERα, on the left 2D representation of binding pocket and ligand. On the right cartoon representation of the amino acid interactions. Model template based on agonist conformation (PDB ID: 1GWR Chain A) and the red ligand is the pose produced by AutoDock Vina at the lowest binding energy.
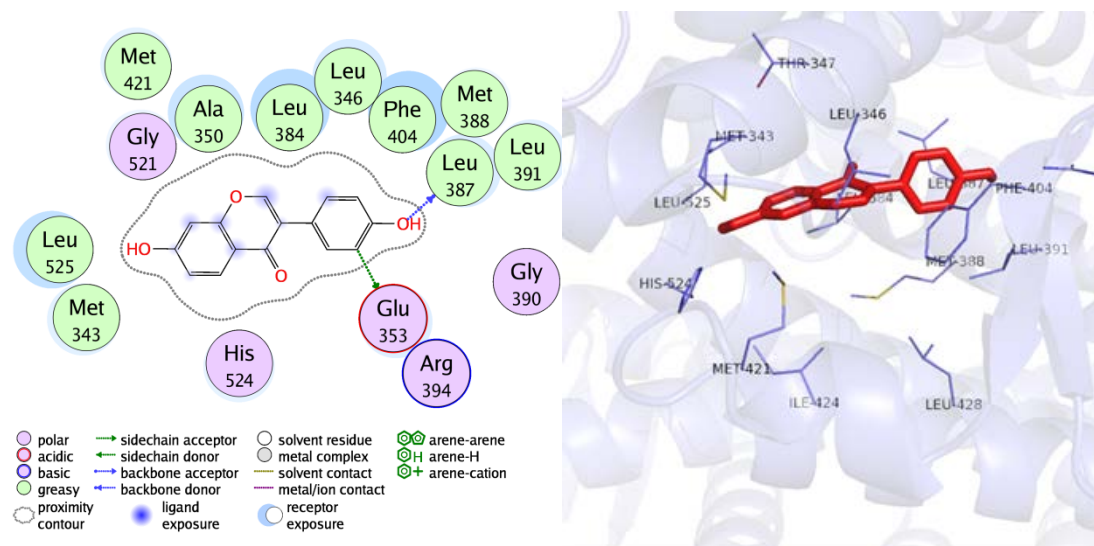
Figure 13. Daidzein binding ERα. On the left 2D representation of binding pocket and ligand. On the right cartoon representation of the amino acid interactions. Model template based on agonist conformation (PDB ID: 1GWR Chain A) and the red ligand is the pose produced by AutoDock Vina at the lowest binding energy.
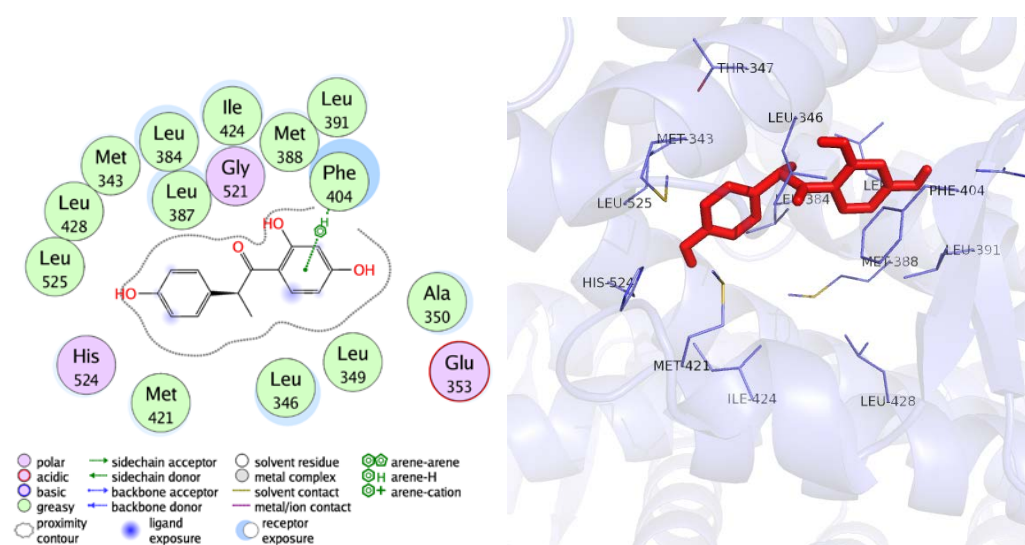
Figure 14. O-desmethylangolensin (ODMA) binding ERα. On the left 2D representation of binding pocket and ligand. On the right cartoon representation of the amino acid interactions. Model template based on agonist conformation (PDB ID: 1GWR Chain A) and the red ligand is the pose produced by AutoDock Vina of the lowest binding energy.

65
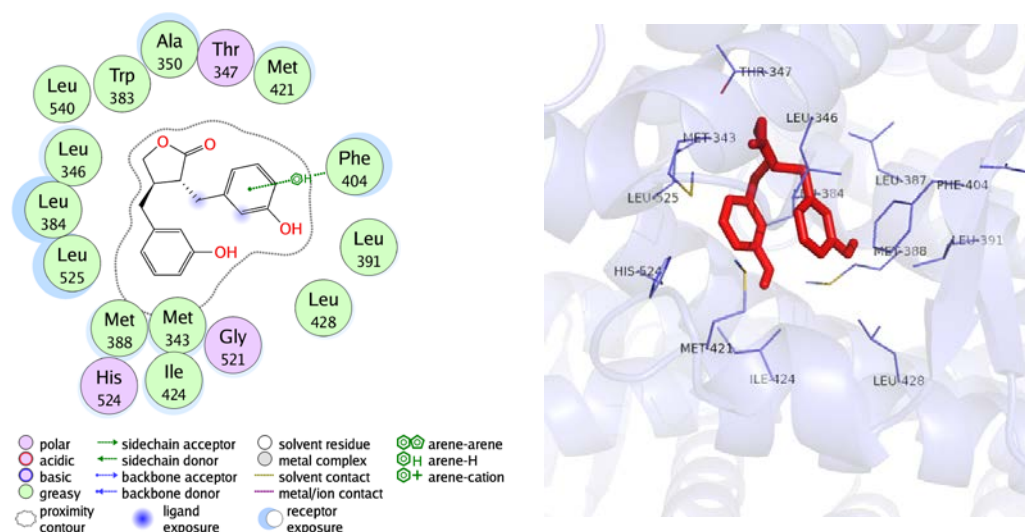
Figure 15. Enterolactone binding ERα. On the left 2D representation of binding pocket and ligand. On the right cartoon representation of the amino acid interactions. Model template based on agonist conformation (PDB ID: 1GWR Chain A) and the red ligand is the pose produced by AutoDock Vina at the lowest binding energy.

All of the phytoestrogens studied using AutoDock Vina interacted with the hormone binding site of ERα as shown in Figures 12 to 15 and Table 3. The LBD is characterised by the presence of hydrophobic amino acids, favouring the interaction with the (hydrophobic) bi- and tri-phenolic ring structures that are a feature of the phytoestrogens studied. Genistein, daidzein and ODMA show considerable structural similarity and, as would be predicted, behaved similarly: interacting with Glu-353 and Arg-354 in the LBD. His-524 was also consistently found within the binding sites anchoring the phytoestrogens as shown in the previous figure 12 and 14. As the LBD for wild type ERα and ERβ also show structural similarities (Chapter 3) (Moras and Gronemeyer 1998) the observations regarding phytoestrogen interactions with ERα were expected to be very similar for ERβ - evaluation of the free energy of binding, a measure of the affinity of the interactions, shows that this was the case (section 4.3.2 below).

66

4.3.2 The affinity of endogenous estrogens, drugs and dietary phytoestrogens to the LBD of ERα and ERβ

AutoDock Vina is a highly sophisticated computational system (Chapter 1) with in-built scoring functions, for example, the root-mean square deviation (RMSD) score and estimations for the change in free energy (ΔG) and predicted free energies of binding (kcal/mol) for ligand-receptor interactions. The AutoDock Vina free energy measurement tool was used to rank the affinity of interaction of estrogens, drugs and phytoestrogens to ERα and ERβ as shown in Table 3.

Table 3. Affinity of estrogen and phytoestrogens to wild type ERα and ERβ LBD. The computational docking program AutoDock Vina provided free energy values. The compounds have been listed in order of ranking with highest affinity interactions shown at the top of the list for each class of compounds. Seco = secoisolariciresinol. ODMA = O-desmethylangolensin.

Description	Ligand	ERα (ESR1)	ERβ (ESR2)
		wild type (variant 1)	
		Kcal/mol	
Estrogens	Estradiol	-10.6	-10.7
	Estrone	-10.2	-10.4
	Estriol	-7.4	-10
Drugs	Fulvestrant	-7.1	-6.4
	Tamoxifen	-6.3	-6.3
Isoflavones	Genistein	-8.7	-8.3
	Daidzein	-8.5	-8.4
	Naringenin	-8.1	-8.2
	Biochanin	-7.9	-6.6
	Formononetin	-7.6	-7.1
	Glycitein	-7.3	-6.5
Lignans	Shonanin	-7.9	-6.7
	Seco	-7.2	-5.9
	Matairesinol	-6.9	-6.8
Phytoestrogen metabolites	Enterolactone	-9.1	-8.6
	ODMA	-8.3	-8.5
	Equol	-7.2	-8.5
	Enterodiol	-5.4	-5.4

In systems such as AutoDock Vina, the free-energy binding calculation is used to rank the ligands with the greatest affinity to the receptor. This value will be influenced by a range of factors including the resolution of the X-ray crystal structure used as the template. The highest affinity interactions with the LBD of ERα and ERβ calculated by the AutoDock Vina system were for the estrogens (estradiol, estrone, and estriol). In addition, the isoflavones genistein, daidzein, and metabolites enterolactone, equol and ODMA all showed greater affinity to the receptor than the lignans secoisolariciresinol, matairesinol and shonanin with binding energies for these phytoestrogens similar or greater than for Tamoxifen. Further work was undertaken to determine whether the phytoestrogens described above interact with the ER as antagonists (Chapter 5) and to assess the potential binding of phytoestrogens to known variants of the ER (Chapter 1). As a pre-requisite for this work, homology models of the variants reported for ERα and ERβ were prepared. A multi-sequence alignment was undertaken using Clustal-Omega with the amino acid sequences for the LBDs for ERα and ERβ as shown in Figure 16. Four variants have been described for ERα (including the wild type, canonical sequence, isoform 1). ERα isoform 2 has a large part of the LBD deleted from amino acids 255-366 (shown in Figure 16 as residues 1-65); isoform 3 has precisely the same LBD as the wild type sequence; isoform 4 comprises a shorter LBD of only 184 amino acids exhibiting significant changes to the LBD alterations/deletions in amino acids 458-595 (shown in Figure 16 as residues 157-252). In addition, isoforms 3 and 4 are missing amino acids 1-173 as a result of alternative promoter usage (data not shown).
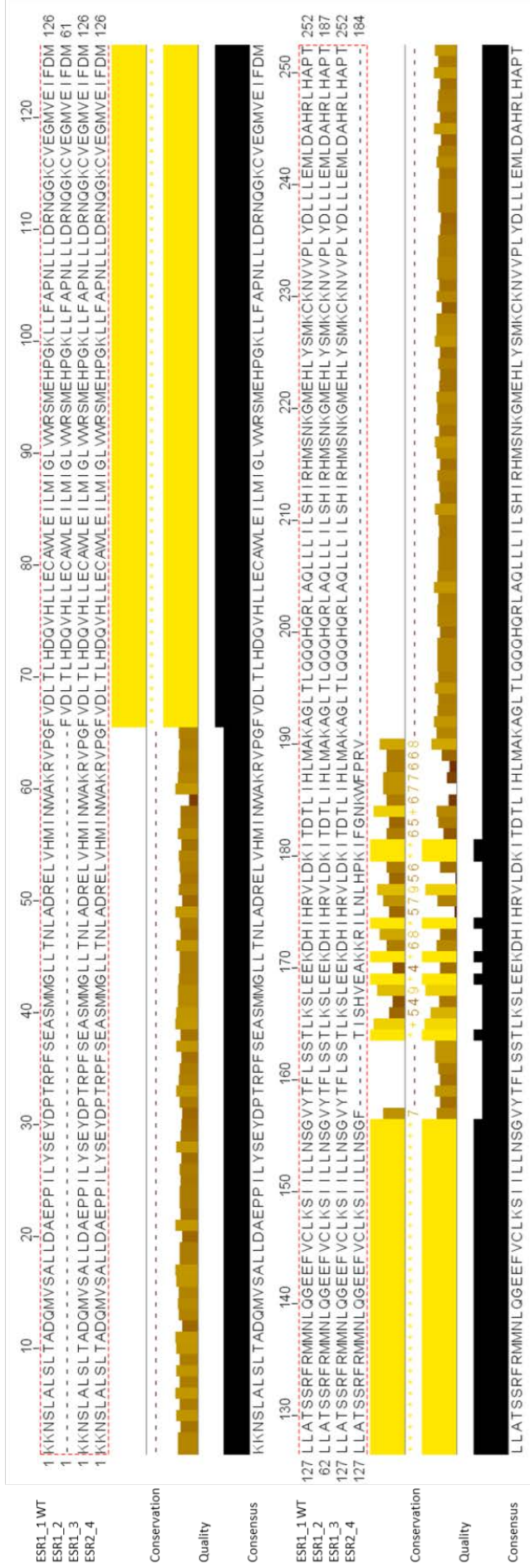
Figure 16. Cluster alignment human ERα (ESR1) LBD, output from Clustal-Omega analysis. The wild type receptor ESR1_1 wt (accession number P03372-1) and isoforms 2, 3 and 4 shown as 1_2; 1_3; 1_4 (accession numbers P03372-2; P03372-3 and P03372-4 respectively). Note: isoform 2 (ESR1_2) is missing residues from the LBD; isoform 3 (ESR1_3) possesses the same LBD as the wild type (ESR1_1) and isoform 4 (ESR1_4) possesses the segment: FTISHVEAKKRILNLHPKIFGNKWFPRV in lieu of amino acids VYTFL...FPATV.

69

In contrast to ERα, nine variants have been described for ERβ (including the wild type, canonical sequence, isoform 1), Figure 17. ERβ isoform 2 has a large part of the LBD deleted from amino acids 469-530 (shown in Figure 17 as residues 173-204); isoform 3 has premature stop codon in the mRNA the result of which is it is unlikely to be translated into a mature protein, isoforms 4, 5 and 6 show similarity in the LBD until changes which occur in amino acids 469-530 (shown in Figure 17 as residues 174 onwards). These changes in isoforms 5 and 6 mean that they will not be able to form homodimers (Bairoch *et al.* 2005).

Isoform 7 is missing amino acids 319-409 (shown in Figure 17 as residues 2-94); isoform 8 is missing amino acids 376-530 (shown in Figure 17 as residues 77 onwards) and displays 365-375: DEGKCVEGILE→YVPSGHSDPGC; finally; isoform 9 is missing amino acids 475-530 (shown in Figure 17 as residues 174 onwards) and displays 469-474: SNKGME→RSCVYK.

4.3.3 Homology modelling of the ERα and ERβ isoforms

The amino acid sequences for isoforms of ERα and ERβ presented in Figures 16 and 17 were used in homology modelling experiments performed in the SWISS-MODEL workspace, Table 4. The quality of the models produced was assessed using the "global quality of the model" (GQME QMEAN4) scoring system provided as a tool in the SWISS-MODEL server (Benkert *et al.* 2011).
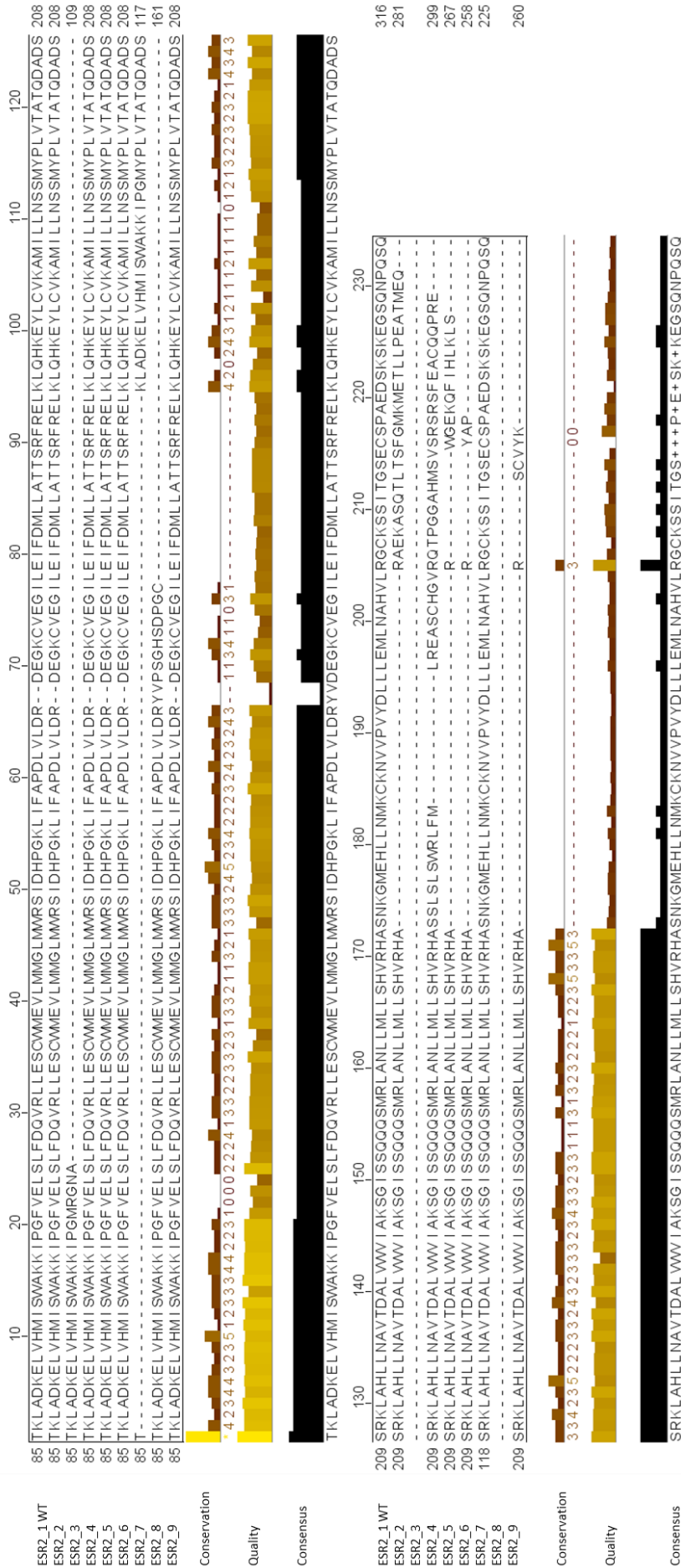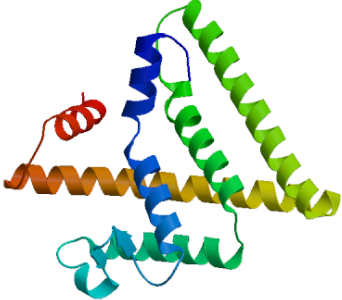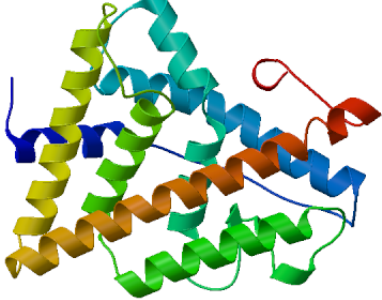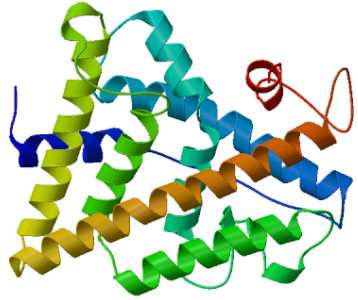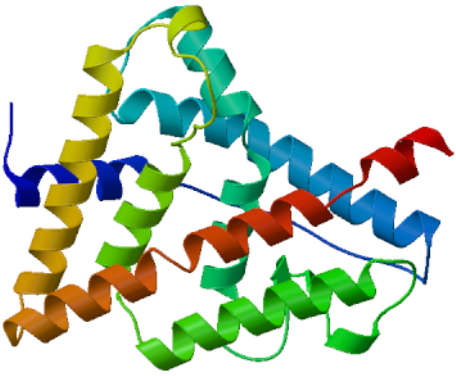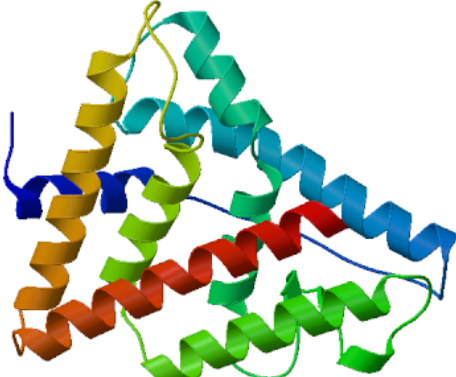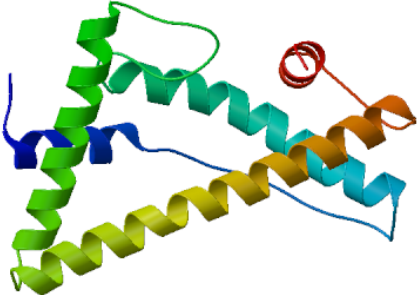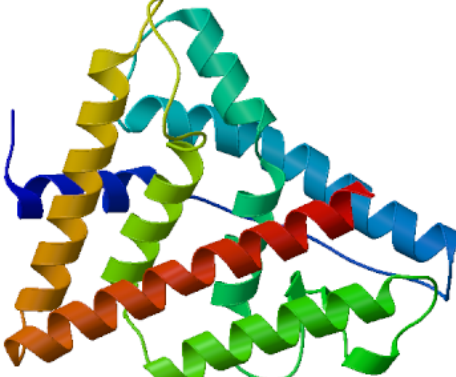
Figure 17. Cluster alignment human ERβ (ESR2) LBD, output from Clustal-Omega analysis. The wild type receptor ESR2_1 wt (accession number Q92731-1) and isoforms 2-9 shown as 2_2→2_9; (accession numbers Q92731-1 to Q92731-9 respectively). Note: the differences in these isoforms are described in the text.

Table 4. Homology models of different isoforms of ERα/ERβ LBD. The homology modelling and model quality evaluation was performed in the SWISS-MODEL workspace. Global quality of the model was scored using the GQME-QMEAN4 system. GQME estimates the accuracy of the model according to the

PROTEIN NAME	MODEL
ERα (ESR1) Isoform 2 **Template PDB ID** 1GWR, Chain A **GQME** 0.99 **Qmean4** -1.33 **Template Resolution** 2.40 Å	
ERα (ESR1) Isoform 4 **Template PDB ID** 1GWR, Chain A **GQME** 0.96 **Qmean4** -1.35 **Template Resolution** 2.40 Å	
ERβ (ESR2) Isoform 2 **Template PDB ID** 4J24 Chain A **GQME** 0.81 **Qmean4** -1.55 **Template Resolution** 2.10 Å	
ERβ (ESR2) Isoform 4 **Template PDB ID** 4J24 Chain A **GQME** 0.79 **Qmean4** -1.12 **Template Resolution** 2.10 Å	

target template alignment, while QMEAN scores local geometry and distance-dependent potentials to assess possible ligand-receptor interactions.

Table 4a continued. Homology models of different isoforms of ERα/ERβ LBD.

PROTEIN NAME	MODEL
ERβ (ESR2) Isoform 5 **Template PDB ID** 4J24 Chain A **GQME** 0.83 **Qmean4** -1.10 **Template Resolution** 2.10 Å	
ERβ (ESR2) Isoform 6 **Template PDB ID** 4J24 Chain A **GQME** 0.83 **Qmean4** -1.19 **Template Resolution** 2.10 Å	
ERβ (ESR2) Isoform 7 **Template PDB ID** 4J24 Chain A **GQME** 0.76 **Qmean4** -0.22 **Template Resolution** 2.10 Å	
ERβ (ESR2) Isoform 9 **Template PDB ID** 4J24 Chain A **GQME** 0.83 **Qmean4** -1.08 **Template Resolution** 2.10 Å	

The structures presented in Table 4, represent all the isoforms that were successfully modelled using the SWISS-MODEL system with estradiol in the LBD. Note, ERα isoform 3 is not shown as this has exactly the same sequence as the wild type receptor. ERα isoform 2 and 4 were both found to bind estradiol despite the many missing amino acid residues from the LBD isoform 2 which is missing the first three helices physically opening the binding pocket in the structure, hence the number of residues that it can interact with are less; L42, L45, A49, E52 are not present in this structure hence lowering the predicted energy binding. Of the nine isoforms that have been described for ERβ isoforms 2 and 5 were able to form homodimers. Conversely, all the other isoforms that were modelled using this system remained in monomeric form using the same PDB file retrieved to model the wild type (Chapter 3).

4.4 Discussion

The improvements in breast cancer survival rates over the past two decades have led to an increased interest in the role of diet and lifestyle on breast cancer patient outcomes. Observational studies have reported a reduction in breast cancer mortality in post- but not in pre-menopausal women consuming the highest intakes of lignans (McCann *et al.* 2012) and reduced mortality overall, but not breast cancer-specific, for the highest quintile of the phytoestrogen metabolites, enterolactone and enterodiol (Buck *et al.* 2011). Similarly, a recent meta-analysis described the association between enterolactone levels and reduced breast cancer specific mortality (Seibold *et al.* 2014). Despite these reports there is very limited information around the potential interaction between dietary phytoestrogens and their metabolites and the ER.

The work presented in this chapter was concerned with the potential estrogenic activity of dietary phytoestrogens. The work extended the current understanding beyond the interaction between genistein (the isoflavone abundant in soy products) with the ER and showed the binding properties of both lignans as well as phytoestrogen metabolites. Most research on phytoestrogen enriched diets has been concerned with the effect of soy-based isoflavone consumption, however, lignans are the more widely consumed phytoestrogens of Western diets (Valsta *et al.* 2003). The DietCompLyf study has reported that lignans are the main source of phytoestrogens in the diet of UK breast cancer patients and therefore it is of interest to consider the interaction between lignans and the ER (Swann *et al.* 2013).

Analysis of the docking data from the homology models produced using AutoDock Vina showed that estrogen and estrone bound to the LBD of ERα with the highest affinities (-10.6 and -10.7 Kcal/mol respectively) followed by ERβ binding (-7.4 and -10 Kcal/mol respectively). Tamoxifen, as expected, bound weakly (ERα -7.1 Kcal/mol and ERβ -6.4 Kcal/mol) in contrast the phytoestrogen metabolite enterolactone had an affinity for the ER slightly less than estradiol but greater than Tamoxifen (ERα -9.1 Kcal/mol and ERβ -8.6 Kcal/mol). In terms of the "ranking" of binding affinities, all of the dietary phytoestrogens showed lower affinity binding compared with estradiol this suggests that phytoestrogens should not displace estradiol from the LBD and therefore should not act as stimulatory compounds. It remains the case, however, that this observation will need to be tested in laboratory studies but this does support the findings of the epidemiological observational studies (Buck et al. 2011; McCann et al. 2012; Seibold et al. 2014) described above and suggesting that there is no disadvantage in terms of survival outcomes for breast cancer patients consuming phytoestrogens.

The increasing interest in pharmacogenomics led to the preparation of *in silico* homology models of the isoforms for ER. It may be the case that for some of the variants of the ER, phytoestrogen binding may be of a higher affinity and therefore may be potentially deleterious. Some of the isoforms were observed to retain estradiol binding properties and were able to form homodimers. The next phase of this work, Chapter 5, is concerned with the binding of phytoestrogens to these isoforms and potential antagonistic properties of this important class of dietary chemicals.

CHAPTER FIVE

5. Phytoestrogen interactions with variants of ERα/ERβ

5.1 Introduction

Variants and spliced forms of the ER have been reported in both normal breast tissue and in breast cancer; this has led to the hypothesis that outcomes in ER positive breast cancer might relate to the presence of wild type ER or variants in the ER. The wild type ERα (also known as ERα66) is the most prevalent form of ER found in breast cancer. However, other ER isoforms have been reported (isoform 3, 4 and splicing variant ERαΔ5/6/7, ERαΔ3/4/5, ERαV) (Dixon and Oshiro 1995; Flouriot *et al.* 2000; Marino *et al.* 2006; Ohshiro *et al.* 2010; Rao *et al.* 2011). Isoform 4 (ERα36) has been reported to interfere with ERα66 activity and negatively affect estrogen treatments of breast cancer (Wang *et al.* 2005). Conversely, ERβ1 (Isoform 1) is associated with better survival in triple-negative breast cancer as well as better response to Tamoxifen monotherapy, while nuclear ERβ2 (Isoform 2) levels have been shown to be associated with metastasis and vascular invasion and cytoplasmic ERβ2 linked to a poorer outcome(Thomas and Gustafsson 2011).

The study of ERβ remains an important focus of current clinical research. ERβ1 and ERΔ5 isoforms have been reported to be co-expressed in healthy mammary glands (Mandušić, Nikolić-Vukosavljević, Tanić, Kanjer, *et al.* 2007) and ERΔ5 may have inhibitory activity acting on both ERα and ERβ, and the high expression of this isoform has been linked to indicators of less aggressive breast cancer (Speirs *et al.* 2000).

The presence of different ER isoforms and the potential for dimerisation amongst the isoforms, alongside different expression levels in different cells/tissues provides an opportunity for a better understanding and development of different approaches for improving prognosis and treatment of cancer. The optimisation and validation of a docking approach to assess potential ligand interactions with the ER ligand binding domain (LBD) might help to better unravel the biological function of the natural ligands of the ER isoforms. It is expected that such an approach will also help to improve our understanding of the significance of phytoestrogen consumption in breast cancer patients exhibiting different ER variants/isoforms.

5.2 Aim

The aim of the work presented in this chapter was to analyse, *in silico*, the interaction between phytoestrogen ligands and variants in the LBD of ERα and ERβ using the homology models generated in Chapter 4.

5.3 Results

The ligand binding was assessed using the agonist and antagonist models generated in Chapter 4 and compared with the wild type LBD of ERα and ERβ and the LBD of different ERs subtypes. Models were built by homology modelling using agonist and antagonist conformations with template crystal structures extracted from the protein database as before the ligands (Estradiol (EST) and Tamoxifen (TAM)) were selected (ERα agonist PDB ID: 1GWR; 2.4 Å, ERα antagonist PDB ID: 3ERT; 1.9 Å, ERβ agonist PDB ID: 2J7X; 2.1 Å, ERβ antagonist PDB ID: 2FSZ).

Validation of the models was performed using well-described estrogen receptor ligands (estradiol, Tamoxifen and genistein).

5.3.1 The affinity of endogenous estrogens, drugs and dietary phytoestrogens to the LBD of ERα and ERβ isoforms using agonist and antagonist model templates.

5.3.1.1 LBD ERα

The AutoDock Vina free energy measurement tool was used to rank the affinity

of interaction of estrogens, drugs and phytoestrogens to the homology models

of the LBD of wild type and ERα isoforms in both the agonist Tables 5, Figure

18 and antagonist Table 6 and Figure 19.

Table 5. Agonist model of the affinity of estrogen and phytoestrogens to the LBD of wild type ERα and isoforms of ERα as shown. The computational docking program AutoDock Vina provided free energy values. The compounds have been listed alphabetically. Seco = secoisolariciresinol. ODMA=O-desmethylangolensin

Description	Ligand	AGONIST MODEL ER α LBD Kcal/mol			
		ESR1_1 wild type variant 1	ESR1_2 variant 2	ESR1_3 variant 3	ESR1_3 variant 4
Physiological Estrogen	Estradiol	-10.6	-8.4	-10.6	-8.8
	Estriol	-7.4	-8.1	-7	-8.2
	Estrone	-10.2	-8.5	-10.6	-8.9
Drugs	Fulvestrant	-7.1	-7.6	-6.7	-8.5
	Tamoxifen	-6.3	-6.4	-6.3	-8.1
Isoflavones	Biochanin	-7.9	-7.1	-7.9	-7.8
	Daidzein	-8.5	-7.1	-7.6	-7.7
	Formononetin	-7.6	-7.2	-7.6	-7.7
	Genistein	-8.7	-6.9	-8.6	-7.9
	Glycitein	-7.3	-7.1	-7.3	-7.3
	Narigenin	-8.1	-7	-8	-7.3
Lignans	Matairesinol	-6.9	-6.8	-7.2	-7.1
	Seco	-7.2	-6.5	-6.5	-6.7
	Shonanin	-7.9	-6.9	-8.2	-7
Lignan metabolites	Enterodiol	-5.4	-5	-5.3	-5.1
	Enterolactone	-9.1	-7	-9.1	-6.8
	Equol	-7.2	-7.3	-8.8	-7.2
	ODMA	-8.3	-6.8	-8.3	-7.4

.
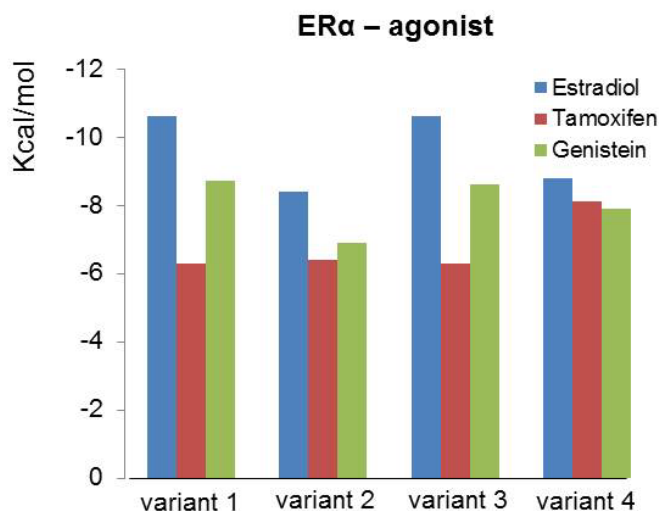
Figure 18. ERα LBD affinity to endogenous and exogenous ligands using the agonist template model in AutoDock Vina docking experiments. The lowest energy conformation per compound is shown. Wild type receptor: variant 1 = ESR1_1; variant 2= ESR1_2, variant 3= ESR1_3, variant 4 = ESR1_4.

As shown in Table 5 and Figure 18, above, the highest affinity interactions with the LBD of wild type ERα calculated by the AutoDock Vina system was for the estrogens (estradiol, estrone, estriol), also described in Chapter 4. Further analysis of binding to variant 3, which has an identical LBD to the wild type LBD of ERα, served as an internal control and the results obtained were almost the same as for wild type receptor. In addition, the isoflavones genistein, daidzein, and metabolites enterolactone, equol and ODMA all showed greater affinity to the wild type and variant 3 of the LBD of ERα than the lignans secoisolariciresinol, matairesinol and shonanin with binding energies for these phytoestrogens similar or greater than for Tamoxifen. The AutoDock Vina free-energy binding values suggested that the LBD of ERα variants 2 and 4 bind more weakly to both Tamoxifen and phytoestrogen ligands than the wild type receptor and variant 3.

Table 6. Antagonist model of the affinity of estrogen and phytoestrogens to the LBD of wild type ERα and isoforms of ERα as shown. The computational docking program AutoDock Vina provided free energy values. The compounds have been listed alphabetically. Seco = secoisolariciresinol. ODMA=O-desmethylangolensin.

Description	Ligand	ANTAGONIST MODEL ERα LBD Kcal/mol			
		ESR1_1 wild type variant 1	ESR1_2 variant 2	ESR1_3 variant 3	ESR1_4 variant 4
Physiological Estrogen	Estradiol	-8.8	-7.6	-7.8	-8.4
	Estriol	-8.1	-7.3	-9.1	-8.4
	Estrone	-10.2	-7.8	-8.1	-8.7
Drugs	Fulvestrant	-10.1	-7.1	-9.8	-6.9
	Tamoxifen	-7.6	-6.8	-7.5	-7.9
Isoflavones	Biochanin	-6.8	-7	-6.9	-7.2
	Daidzein	-9	-7.3	-7.7	-7.5
	Formononetin	-7.7	-6.7	-7.7	-7.3
	Genistein	-8.9	-7.2	-7.2	-8
	Glycitein	-6.5	-7.2	-6.8	-6.9
	Narigenin	-8.5	-7.1	-8.5	-7.5
Lignans	Matairesinol	-6.9	-6.7	-6.8	-6.8
	Seco	-5.9	-5.6	-6	-6.2
	Shonanin	-6.7	-6.8	-7.4	-6.9
Lignan metabolites	Enterodiol	-5.1	-5.7	-5.1	-5.2
	Enterolactone	-6.6	-6.3	-8.2	-7.4
	Equol	-8.9	-7.4	-8.3	-7.5
	ODMA	-7.4	-6.9	-8.5	-7.3
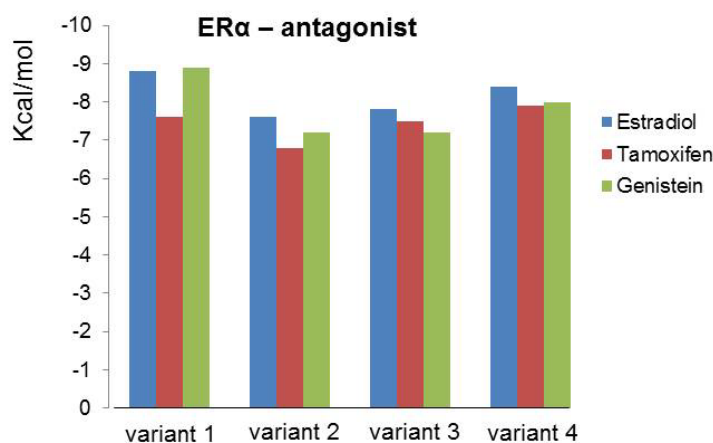
Figure 19.ERα LBD affinity to endogenous and exogenous ligands using the antagonist template model in AutoDock Vina docking experiments. The lowest energy conformation per compound shown. Wild type receptor: variant 1 = ESR1_1; variant 2= ESR1_2, variant 3= ESR1_3, variant 4 = ESR1_4.

When the LBD of ERα was investigated using AutoDock Vina using the antagonist model as the template as shown in Table 6 and Figure 19, the ranking of the binding partners in terms of affinity was much altered compared with the agonist model. In this scenario, estrone and Fulvestrant were found to be the most important binding partners (-10.2 and -10.1 Kcal/mol respectively), followed by daidzein, genistein and the metabolite equol (-9, -8.9 and -8.9 Kcal/mol respectively).

Tamoxifen was ranked 10th in terms of affinity of interaction with the LBD of ERα with lower affinity values than the phytochemicals above. The results from this analysis with the LBD of ERα suggest that daidzein, genistein and equol might all act as more potent antagonists of the LBD of wild type ERα than Tamoxifen, this suggests that the consumption and metabolism of isoflavones may result in a conformational change in the LBD of the receptor and warrants further investigation. Variants 2 and 4 both bound to the physiological estrogens, the

drug and the exogenous dietary phytoestrogen ligands with lower affinity than the wild type receptor and variant 3.

Table 7. Agonist model of the affinity of estrogen and phytoestrogens to the LBD of wild type ERβ and isoforms of ERβ as shown. The computational docking program AutoDock Vina provided free energy values. The compounds have been listed alphabetically. Seco = secoisolariciresinol. ODMA=O-desmethylangolensin. Shaded yellow: physiologically occurring estrogens; pink: drugs; green: isoflavones; blue: lignans, lilac: lignan metabolites.

Ligand	AGONIST MODEL ERβ LBD Kcal/mol								
	ESR2_1 wild type (variant 1)	ESR2_2 variant 2	ESR2_3 variant 3	ESR2_4 variant 4	ESR2_5 variant 5	ESR2_6 variant 6	ESR2_7 variant 7	ESR2_8 variant 8	ESR2_9 variant 9
Estradiol	-10.7	-7.1	-6.2	-7.2	-6.5	-9.9	-7.3	-8	-9.1
Estriol	-10	-6.9	-5.7	-6.9	-6.6	-8	-7.5	-7.6	-7
Estrone	-10.4	-6.8	-6	-6.8	-6.9	-9.4	-7.6	-7.8	-9.6
Fulvestrant	-6.4	-6.5	-4.7	-6.1	-6.5	-8.5	-6.4	-7.2	-6.4
Tamoxifen	-6.3	-6.4	-5.9	-6.6	-6.7	-5.9	-6.2	-7.1	-5.8
Biochanin	-6.6	-6.4	-5.8	-6.4	-7.1	-8.3	-6.6	-7	-8
Daidzein	-8.4	-7	-5.9	-6.6	-6.6	-8.3	-6.7	-7	-8
Formononetin	-7.1	-7.1	-5.5	-7	-6.9	-8.2	-6.7	-6.9	-8
Genistein	-8.3	-7.2	-5.9	-6.5	-6.1	-8.3	-6.8	-7	-8
Glycitein	-6.5	-6.1	-6.1	-6.2	-6.4	-6.9	-6.5	-6.9	-7.9
Narigenin	-8.2	-6.7	-5.6	-6.4	-6.3	-7.9	-6.9	-7.7	-7.8
Matairesinol	-6.8	-6.4	-6.2	-6.5	-6.6	-6.2	-6.8	-7.2	-7.5
Seco	-5.9	-6.2	-4.9	-5.9	-5.1	-4.8	-6	-6.3	-5.8
Shonanin	-6.7	-6.5	-6.2	-6.8	-6.5	-6.8	-6.7	-7.1	-7.1
Enterodiol	-5.4	-5.4	-5.2	-5.1	-5.1	-5.6	-4.6	-4.3	-4.9
Enterolactone	-8.6	-6.2	-5.4	-6	-6	-7.3	-7.5	-7.6	-7.7
Equol	-8.5	-7.4	-5.7	-6.6	-6.3	-8.2	-6.7	-7	-8.5
ODMA	-8.5	-5.9	-6.1	-6.1	-5.7	-7.9	-6.7	-7	-7.4

In the case of ERβ there have been nine variant (including the wild type) forms of the receptor described. The AutoDock Vina free energy measurement tool was used to rank the affinity of interaction of estrogens, drugs and phytoestrogens to the LBD of ERβ in the agonist model as shown in Table 7.

Ligand	ANTAGONIST MODEL ERβ LBD Kcal/mol								
	ESR2_1 wild type (variant 1)	ESR2_2 variant 2	ESR2_3 variant 3	ESR2_4 variant 4	ESR2_5 variant 5	ESR2_6 variant 6	ESR2_7 variant 7	ESR2_8 variant 8	ESR2_9 variant 9
Estradiol	-7.6	-6.6	-6.5	-7.1	-6.5	-9.3	-6.8	-8.3	-7
Estriol	-8.9	-6.5	-6.4	-7.4	-6.9	-7.7	-6.6	-8.2	-7.5
Estrone	-9.9	-7.1	-6.4	-7.3	-6.9	-9	-6.9	-8.1	-9.2
Fulvestrant	-9.4	-7.7	-6.6	-5.6	-8	-8.4	-6.2	-7.9	-8.7
Tamoxifen	-7	-6.4	-5.7	-6.6	-6.3	-8.2	-7.3	-7.2	-6.4
Biochanin	-8.6	-8.5	-5.8	-8.4	-8.3	-8.3	-6.2	-7.4	-8.1
Daidzein	-8.9	-8.1	-5.7	-8.1	-8	-8	-6.1	-7.3	-7.9
Formononetin	-8.1	-8.2	-5.7	-8.1	-8.1	-8.1	-6.2	-7.3	-7.9
Genistein	-8.9	-8.4	-5.8	-8.3	-8	-8.2	-6.2	-7.4	-8.2
Glycitein	-7.7	-7.8	-5.4	-7.6	-7.1	-7.5	-6.2	-7.4	-7.2
Narigenin	-8.4	-8.1	-6.1	-8.5	-8.1	-8.1	-6.4	-7.7	-8.1
Matairesinol	-7.2	-7.3	-5.5	-7.1	-7.4	-7.4	-6.5	-7.8	-7.4
Seco	-6.9	-5.5	-5.2	-5.8	-6.6	-6.6	-6.2	-7.4	-6.4
Shonanin	-7.1	-6.8	-5.8	-7.4	-7.6	-7.3	-6.6	-7.6	-7.4
Enterodiol	-4.8	-5	-3.8	-6.4	-5.8	-4.8	-5.8	-4.6	-4.4
Enterolactone	-7.6	-7.5	-5.8	-7	-8.6	-7.4	-6.3	-8.1	-7.1
Equol	-8.6	-8	-5.7	-8	-8	-8	-6.4	-7.3	-8.3
ODMA	-8.2	-7.6	-5.8	-7.9	-7.7	-7.6	-6.1	-7.6	-7.8

Table 8. Antagonist model of the affinity of estrogen and phytoestrogens to the LBD of wild type ERβ and isoforms of ERβ as shown. The computational docking program AutoDock Vina provided free energy values. The compounds have been listed alphabetically. Seco = secoisolariciresinol. ODMA=O-desmethylangolensin. Shaded yellow: physiologically occurring estrogens; pink: drugs; green: isoflavones; blue: lignans, lilac: lignan metabolites

As shown in Table 8, the highest affinity interactions with the LBD of wild type ERβ calculated by the AutoDock Vina system was for the estrogens (estradiol, estrone, estriol), also described in Chapter 4. In addition, a number of the isoflavones including genistein and daidzein, and the metabolites enterolactone, equol and ODMA all showed greater affinity to the wild type LBD of ERβ compared with tamoxifen, mirroring the observations that were made for ERα previously. Again, the lignans secoisolariciresinol (-5.9 Kcal/mol), matairesinol (-6.8 Kcal/mol) and shonanin (-6.7 Kcal/mol) showed free-binding energies similar to Tamoxifen (-6.3 Kcal/mol). With the exception of variants 6 and 9, the AutoDock Vina free-energy binding values suggested that the LBD of the ERβ variants bind more weakly to both estrogens, Tamoxifen and phytoestrogen ligands than the wild type receptor.

When the LBD of ERβ was investigated using AutoDock Vina using the antagonist model as the template as shown in Table 8 the ranking of the binding partners in terms of affinity was much altered compared with the agonist model. In this scenario, as for the LBD of ERα, estrone and Fulvestrant were found to be the most important binding partners (-9.9 and -9.4 Kcal/mol respectively), followed by daidzein, genistein and the metabolite equol (-8.9, -8.9 and -8.6 Kcal/mol respectively). Tamoxifen was ranked 16th in terms of affinity of interaction with the LBD of ERβ with lower affinity values than the phytochemicals above.

The results from this analysis with the LBD of ERβ mirror those obtained with ERα: suggesting that daidzein, genistein and equol might all act as more potent antagonists of the LBD of wild type ERβ than Tamoxifen, this suggests that the consumption and metabolism of isoflavones may result in a conformational change in the LBD of the receptor and warrants further investigation.

5.3.2 Overview of affinity of estrogens to the LBD of ERα and ERβ

The data obtained in the experimental approaches described above were compiled as shown in Figures 20 and 21 below. These give an overview of the binding properties of endogenous estrogens, drugs and phytoestrogens to the wild type forms of the LBD of ERα and ERβ in both the agonist and antagonist models used in this study. As can be appreciated, endogenous estrogens, particularly estradiol are the most potent agonists of both ERα and ERβ. In terms of antagonists, somewhat surprisingly, Tamoxifen did not emerge as the most potent antagonist, rather estrone, some of the dietary phytoestrogens, including genistein, daidzein, equol and narigenin all appeared to show greater affinity in the antagonist form to ERα. The results from both ERα and ERβ were very similar in terms of the ranking of the chemicals studied and their overall binding affinities.

Figure 20. The affinity of endogenous and exogenous estrogens including dietary phytoestrogens to ERα. Blue: agonist model, orange: antagonist model. Output of docking experiments in terms of binding affinity (Kcal/mol); the lowest energy conformation per compound. Computational docking allowed the affinity values to be calculated by AutoDock Vina.
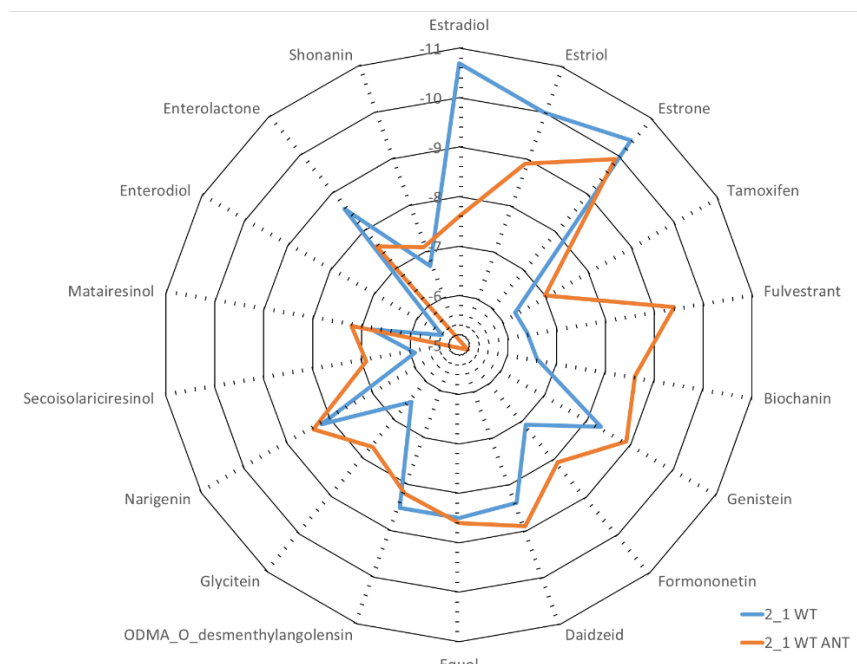
Figure 21. The affinity of endogenous and exogenous estrogens including dietary phytoestrogens to ERβ. Blue: agonist model, orange: antagonist model. Output of docking experiments in terms of binding affinity (Kcal/mol); the lowest energy conformation per compound.

91

5.4 Discussion

The use of ER status as a companion diagnostic for determining the likely response of breast cancer patients to anti-estrogen treatments has been recognised for over three decades (Alldred DC, Modern Pathology (2010) 23, S52–S59). At the time of its adoption in the late 1970s and 1980s the ER status (of ERα) was assessed using a ligand binding assay with radiolabelled estrogen and protein extracts from fresh tumour tissue extracts. This approach offered the advantage of being relatively physiologically relevant and also being directly quantifiable. However, the ligand binding assay was both expensive and technically challenging and since the early 1990s the ER status has been determined using immunohistochemistry in an approach approved by the College of American Pathologists and the American Society for Clinical Oncology (reviewed in Alldred DC 2010). However, there remain difficulties in the standardisation of the immunohistochemistry methodology and on-going issues with intra- and inter-laboratory variation in results as a result of differences in tissue fixation, antigen retrieval and so on. The classification of ER status in breast cancer is, therefore, currently subject to inaccuracies and guidelines were recently published in recognition of this issue with aim of improving consistency across laboratories (Phillips *et al.* 2007).

The discovery of ERβ in 1996 (Kuiper *et al.* 1996) and subsequent gene sequencing and identification of isoforms of both ERα and ERβ has resulted in further complexity around this class of receptors in breast cancer. In this study, the interaction between endogenous and exogenous estrogens and the LBD of ERα and ERβ and isoforms of these receptors was evaluated. ERα and ERβ

were observed to behave similarly, with the free-energy binding values (affinity) ranked in the same order for both ERα and ERβ. ERα variant 3 behaved the same as the wild type receptor.

One of the main concerns around the consumption of dietary phytoestrogens by breast cancer patients is that phytoestrogens have the potential to interfere with anti-estrogen treatments such as Tamoxifen. Conflicting data has continued to emerge around this topic and this led to the independent panel of scientists at the World Cancer Research Fund's Continuous Update Project conclusion that there may be a link between improved survival and the consumption of foods containing soy but that further research was needed into this topic (Aune *et al.* 2012).

This study was concerned with the potential estrogenic effect of dietary phytoestrogens as well as their possible interference with Tamoxifen therapy. The AutoDock Vina system was used to provide estimations for the free-energy of binding and this can be compared with data that has been obtained experimentally, estradiol along the other estrogen prove that they have stronger and more stable molecular interaction with the ERα β. At the same time genistein, daidzein and equol prove to be able to bind the ER in a lower affinity as have been previously reported in the literature.

The results obtained in the experiments above suggest that the dietary phytoestrogens would not bind to either ERα or ERβ or their isoforms with

greater affinity than estradiol. Some of the phytoestrogens appeared to have greater affinity to the ERα and ERβ than Tamoxifen (in the antagonist model) but it is not clear as to whether the resulting structure occludes the binding of helix-12 (chapter 3) to the neighbouring monomer and therefore, this would require further analysis in the form of laboratory experiments.

CHAPTER SIX

6. Discussion

6.1 Development of homology models of ERα and ERβ

The aim of this research project was to assess the ability to apply structure-based studies and computational tools to predict and describe the interactions of the ERα and ERβ with phytoestrogens and synthetic ligands. The first step (Chapter 3) was to develop protein (homology) models to use as targets to later assess the interaction with phytoestrogens (Chapters 4 and 5). Homology modelling has helped researchers to narrow the space between structural modelling from experimental work and protein sequences leading to the production of models of proteins that have been difficult to crystallise or unravel by NMR spectroscopy. Since more than 300 3D-structures of wild type ER are available in the Protein database (URL www.rcsb.org; Berman *et al.,* 2000) the criteria used to select possible templates were based on the resolution of the crystal or NMR structure, ligand binding interactions and the 3D conformations. The performance of the models to re-dock the respective ligands derived from their own templates and to emulate the poses of the original ligands enabled an assessment of the utility of the model. These models offer the opportunity for further studies for example mutagenesis studies using *in situ* or amino acid substitution and docking with a wide range of possible ligands.

There are several tools to assess and validate the models including quality estimation approaches which assess the specific features of the model such as the sterochemical quality of the interactions (Morris *et al*. 1992), structure descriptions (Kabsch and Sander 1983) and quality scores to evaluate certain

potentially unreliable parts of the protein model structures (Wiederstein and Sippl 2007) or web-servers that can complement and provide extra information about the models (Willard *et al.* 2003). The best assessment approaches rely on the consensus provided by a whole range of *in silico* validation tools. QMEAN is considered an important tool for the quality assessments of homology models and increasing the performance of the score through the combination of different terms and the QMEAN Z-score provides further data about the absolute quality of the predicted homology models.

Homology modelling was first applied to model the wild type isoform of each receptor (α and β). AutoDock Vina was used for the molecular docking and also provided a binding energy score. Afterwards, the isoforms for the ligand binding domain of both ERs were built based on the same template with two models built per isoform. The first approach was to produce a model on an agonist complex and the second in antagonist conformation. This is particularly important when considering potential ER interaction partners as several hormonal treatments for breast cancer (including Tamoxifen) act as antagonists, whilst receptor stimulation occurs via the agonist properties of estradiol and other estrogens.

6.2 Docking studies with estrogens, drugs and phytoestrogens

Molecular docking is a valuable approach to understanding receptor-ligand interactions and work in many research laboratories has illustrated the value of this approach (Yuriev and Ramsland 2013). The creation of new, and improvements to, existing algorithms has resulted in improvements in performance of molecular docking systems (Xu and Lill 2013). Furthermore, this field has greatly benefited from the research advances in fields such as first principle thermodynamics and computer developments, that have been applied to help refine the programmes as well as to aid understanding of the results (Dahlqvist, Meshkian, and Rosen 2017). At the same time the use and understanding of bioinformatics in research laboratories has evolved such that it is now a mandatory field of knowledge for those active in many areas of molecular and biomedical sciences. These diverse developments have resulted in the creation of somewhat more user-friendly tools for biology in order to allow the researcher to focus on interpretation of data collected rather than forcing them to develop and understand complex equations and algorithms (Trott and Olson 2010b).

Docking analysis of estradiol and phytoestrogens into the LBD of ERα and ERβ showed the intrinsic quality of these natural polyphenolic compounds to emulate the estradiol-like docking poses and possibly alter the conformational internal structure of the ligand binding pocket (and potentially the conformation of whole protein structure). At the same time, as a function of its complex algorithm, AutoDock Vina had the capacity and accuracy to predict the binding conformation such that estradiol was able to emulate the poses registered

through the experimental data and described in the literature. Also, there was a correlation between the binding affinities reported in the literature with the values of the free-energy binding that were obtained using the docking system. As expected estradiol showed the greatest affinity (lowest free-energy binding) of the structures assessed, consistent with reports that estradiol is the most potent ligand of all the estrogens to ERα (Hall *et al.* 2001).

6.3 Significance of the results obtained

We hypothesised that natural ligands with estrogen-like structures might drive ER-related biological functions and therefore might have the ability to affect breast cancer outcomes. These potential effects would depend upon the molecular profile of the breast cancer and may be relevant to patients with both ER+ and ER- breast cancer, affecting the efficacy of clinical treatments that target the ER.

Natural compounds such as dietary phytoestrogens vary according to the plant strains from which they derive, the agricultural conditions (including harvest time), food preparation and food composition alongside their bioavailability. Furthermore, based on scientific evidence obtained from epidemiological reports, case studies, *in vitro* and *in silico* experiments it is clear that differences in exposure to phytoestrogens will occur in an intra- and inter-individual manner. This may be due, for example, to differences in the proportions of different phytoestrogens consumed, the presence of aglycol-derivatives, compound concentration and metabolism (Dixon 2004; Hwang and Choi 2015; Swann *et al.* 2013). It is necessary, therefore, to acknowledge the complexity of this

system and recognise that the mechanisms by which this highly heterogeneous group of natural compounds interact with the ER might have potential positive but as yet unmeasured impact on breast cancer biology. Identifying and understanding the molecular interactions of such compounds with the ER is the first step in understanding their biological impact in breast cancer. The impact of three estradiol analogs on intracellular signalling pathways has been reported and described the ability of estrogen-like molecules to alter gene expression levels depending on the ligand binding interactions (Yarger *et al*. 2013). Whilst it will be necessary to validate the work of the study performed here using *in vitro* experiments to determine if the binding affinities and free-energy binding values truly correlate with estrogenicity of the compounds described, the system described has much potential as a screening tool for the assessment of chemicals such as xenoestrogens (and other exogenous compounds) which might exert biological effects through the ER.

6.4 Validation

An important consideration when evaluating *in silico* modelling is whether the docking results obtained mimic the behaviour of the ER *in vitro.* Since the main goal of this research was to investigate the interactions between the ER and phytoestrogens, ideally, full length ER needs to be produced. Mammalian cells and a baculovirus system have been shown to be of utility in this respect but these are costly, time-consuming systems which are difficult to reproduce. An alternative approach is to express only the LBD of the ER. This has been feasible with the LBD of wild type ERα and ERβ isoforms successfully cloned, expressed and purified in a good yield using an expression system which adds a small ubiquitin-related modifier (pET-SUMO) that helps to stabilise the recombinant protein. So far it has not been possible to purify each variant of each ER isoform but they have been expressed and optimisation steps for the purification are on-going (data not shown here). The purified ER proteins will be used in a fluorescence polarisation ligand binding assay and the plasmid will be used for mutagenesis studies.

6.5 Future work

The use of *in silico* work with AutoDock Vina and other open source bioinformatics tools have provided an insight data into the possible interaction between a wide range of phytoestrogen compounds and the LBD of ER and its variants. Further work in terms of analysis of these interactions is needed with an experimental approach using laboratory based screening; this will constitute an important approach for comparing the *in silico* and the *in vitro* data allowing for the prediction and design of further *in vitro* and *in vivo* experiments. Another important approach will be to assess the effect of these compounds in terms of promoting breast cancer cell proliferation. This might be evaluated by addition of different concentrations of estrogens (E1, E2, E3) (Kumar *et al.,* 2011), phytoestrogens: genistein, daidzein, coumestrol, equol (Rice *et al.,* 2008; Setchell *et al.,* 2010) and synthetic agonists and antagonists (TAM, ICI 182,780) (Pearce and Jordan 2004). Ligand-free controls (for example only ethanol) will need to be produced with each ligand assessed on its own and in combination with different concentrations of other ligands presented in the study. As the LBD of the ER is not directly related to its transcriptional activity the selectivity of phytoestrogens for ER subtypes and isoforms may be influenced by ligand concentration. Furthermore, in order to measure transcriptional activity cell-based assays need to be used based on reporter systems such as the ERE-luciferase reporter methodology.

The DietCompLyf study is a multi-centre observational study that is coordinated by the Cancer Research Unit at the University of Westminster (Principal Investigator: Dr Dwek). The DietCompLyf study is focussed on determining the role of phytoestrogens on breast cancer recurrence rates in a UK population

(Velentzis *et al.,* 2011). A key element of the DietCompLyf study is the availability of questionnaires giving a snap shot of food intakes alongside biological samples enabling analysis of phytoestrogen intake, biomarker levels and correlation with clinical treatments and patient recurrence rates. The database of the DietCompLyf study counts comprises around 3,144 women. A separate nested case control cohort of around 200 patient matched with all the clinical criteria is available. All the subjects have provided a food frequency questionnaire, and a 7-day food diary allowing the phytoestrogen levels to be estimated in addition their urinary phytoestrogen levels which have been quantified by liquid chromatography/tandem mass spectrometry(LC–MS/MS) (Grace et al. 2007). The findings of the *in silico* and *in vitro* experiments from the work described in this thesis will be compared with data and samples collected as part of the DietCompLyf study.

6.6   Conclusion

The work described in this thesis has shown the potential of homology modelling and ligand docking for unravelling interactions between physiological estrogens, hormone therapies and dietary phytoestrogens and the ER. The models have been interrogated in terms of the residues recognised as well as the free-energy binding values and have been found to correlate with experimentally derived values. The system developed offers considerable potential for the screening of chemicals which might be important in driving breast cancer progression through the ER, and also evaluating potential therapeutic candidates as antagonists.

References

(EBCTCG), E.B.C.T.C.G., 2017. Relevance of breast cancer hormone receptors and other factors to the efficacy of adjuvant tamoxifen: patient-level meta-analysis of randomised trials. Lancet 378, 771–784. doi:10.1016/S0140-6736(11)60993-8

Abarzua, S., Serikawa, T., Szewczyk, M., Richter, D.-U., Piechulla, B., Briese, V., 2012. Antiproliferative activity of lignans against the breast carcinoma cell lines MCF 7 and BT 20. Arch. Gynecol. Obstet. 285, 1145–1151. doi:10.1007/s00404-011-2120-6

Adlanmerini, M., Solinhac, R., Abot, A., Fabre, A., Raymond-Letron, I., Guihot, A.-L., Boudou, F., Sautier, L., Vessieres, E., Kim, S.H., Liere, P., Fontaine, C., Krust, A., Chambon, P., Katzenellenbogen, J.A., Gourdy, P., Shaul, P.W., Henrion, D., Arnal, J.-F., Lenfant, F., 2014. Mutation of the palmitoylation site of estrogen receptor alpha in vivo reveals tissue-specific roles for membrane versus nuclear actions. Proc. Natl. Acad. Sci. U. S. A. 111, E283-90. doi:10.1073/pnas.1322057111

Akiyama, T., Ishida, J., Nakagawa, S., Ogawara, H., Watanabe, S., Itoh, N., Shibuya, M., Fukami, Y., 1987. Genistein, a specific inhibitor of tyrosine-specific protein kinases. J. Biol. Chem. 262, 5592–5595.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Anstead, G.M., Carlson, K.E., Katzenellenbogen, J.A., 1997. The estradiol pharmacophore: Ligand structure-estrogen receptor binding affinity

relationships and a model for the receptor binding site. Steroids 62, 268–303. doi:10.1016/S0039-128X(96)00242-5

Aune, D., Chan, D.S.M., Vieira, A.R., Rosenblatt, D.A.N., Vieira, R., Greenwood, D.C., Norat, T., 2012. Fruits, vegetables and breast cancer risk: a systematic review and meta-analysis of prospective studies. Breast Cancer Res. Treat. 134, 479–493. doi:10.1007/s10549-012-2118-1

Axelson, M., Sjovall, J., Gustafsson, B.E., Setchell, K.D., 1984. Soya--a dietary source of the non-steroidal oestrogen equol in man and animals. J. Endocrinol. 102, 49–56.

Bairoch, A., Apweiler, R., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N., Yeh, L.-S.L., 2005. The Universal Protein Resource (UniProt). Nucleic Acids Res. 33, D154-9. doi:10.1093/nar/gki070

Bakas, P., Liapis, A., Vlahopoulos, S., Giner, M., Logotheti, S., Creatsas, G., Meligova, A.K., Alexis, M.N., Zoumpourlis, V., 2008. Estrogen receptor alpha and beta in uterine fibroids: a basis for altered estrogen responsiveness. Fertil. Steril. 90, 1878–85. doi:10.1016/j.fertnstert.2007.09.019

Banerjee, S., Li, Y., Wang, Z., Sarkar, F., 2008. Multi-targeted therapy of cancer by genistein. Cancer Lett. 269, 226–242. doi:10.1016/j.canlet.2008.03.052.MULTI-TARGETED

Barnes, S., 2010. The biochemistry, chemistry and physiology of the isoflavones in soybeans and their food products. Lymphat. Res. Biol. 8,

89–98. doi:10.1089/lrb.2009.0030

Benkert, P., Biasini, M., Schwede, T., 2011. Toward the estimation of the absolute quality of individual protein structure models. Bioinformatics 27, 343–350. doi:10.1093/bioinformatics/btq662

Benkert, P., Tosatto, S.C.E., Schomburg, D., 2008. QMEAN: A comprehensive scoring function for model quality assessment. Proteins 71, 261–277. doi:10.1002/prot.21715

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The Protein Data Bank. Nucleic Acids Res. 28, 235–242.

Bernstein, L., Ross, R.K., 1993. Endogenous Hormones and Breast Cancer Risk. Epidemiol Rev 15, 48–65.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., Schwede, T., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42, W252-8. doi:10.1093/nar/gku340

Bjornstrom, L., Sjoberg, M., 2005. Mechanisms of estrogen receptor signaling: convergence of genomic and nongenomic actions on target genes. Mol. Endocrinol. 19, 833–842. doi:10.1210/me.2004-0486

Bliedtner, A., Zierau, O., Albrecht, S., Liebhaber, S., Vollmer, G., 2010. Effects of genistein and estrogen receptor subtype-specific agonists in ArKO mice following different administration routes. Mol. Cell. Endocrinol. 314, 41–52. doi:10.1016/j.mce.2009.07.032

Bradley, P., Misura, K.M.S., Baker, D., 2005. Toward High-Resolution de Novo Structure Prediction for Small Proteins. Science (80-.). 309, 1868 LP-1871.

Bray, F., Møller, B., 2006. Predicting the future burden of cancer. Nat. Rev. Cancer 6, 63–74. doi:10.1038/nrc1781

Brueggemeier, R.W., 2001. Aromatase, aromatase inhibitors, and breast cancer. Am. J. Ther. 8, 333–344.

Brzozowski, A.M., Pike, A.C.W., Dauter, Z., Hubbard, R.E., Bonn, T., Engstrom, O., Ohman, L., Greene, G.L., Gustafsson, J.-A., Carlquist, M., 1997. Molecular basis of agonism and antagonism in the oestrogen receptor. Nature 389, 753–758.

Buck, K., Zaineddin, A.K., Vrieling, A., Heinz, J., Linseisen, J., Flesch-Janys, D., Chang-Claude, J., 2011. Estimated enterolignans, lignan-rich foods, and fibre in relation to survival after postmenopausal breast cancer. Br. J. Cancer 105, 1151–1157. doi:10.1038/bjc.2011.374

Buck, K., Zaineddin, A.K., Vrieling, A., Linseisen, J., Chang-Claude, J., 2010. Meta-analyses of lignans and enterolignans in relation to breast cancer risk. Am. J. Clin. Nutr. 92, 141–153. doi:10.3945/ajcn.2009.28573

Caan, B.J., Natarajan, L., Parker, B., Gold, E.B., Thomson, C., Newman, V., Rock, C.L., Pu, M., Al-Delaimy, W., Pierce, J.P., 2011. Soy food consumption and breast cancer prognosis. Cancer Epidemiol. Biomarkers Prev. 20, 854–8. doi:10.1158/1055-9965.EPI-10-1041

Cancer Research UK, 2012. Statistics and outlook for breast cancer [WWW Document]. URL

http://www.cancerresearchuk.org/?gclid=CNTfgvvbiK8CFcgntAodPjAlAQ
(accessed 3.12.12).

Cavallo, F., De Giovanni, C., Nanni, P., Forni, G., Lollini, P.-L., 2011. 2011: the
immune hallmarks of cancer. Cell 60, 319–326.

Cederroth, C.R., Nef, S., 2009. Soy, phytoestrogens and metabolism: A
review. Mol. Cell. Endocrinol. 304, 30–42. doi:10.1016/j.mce.2009.02.027

Chang, M., 2012. Tamoxifen Resistance in Breast Cancer. Biomol. Ther.
(Seoul). 20, 256–267. doi:10.4062/biomolther.2012.20.3.256

Chen, W., Colditz, G., 2007. Risk factors and hormone-receptor status:
epidemiology, risk-prediction models and treatment implications for breast
cancero. Nat. Rev. Clin. Oncol. 4, 415–423. doi:10.1038/ncponc0851

Cheng, J., Randall, A.Z., Sweredoski, M.J., Baldi, P., 2005. SCRATCH: a
protein structure and structural feature prediction server. Nucleic Acids
Res. 33, W72-6. doi:10.1093/nar/gki396

Chothia, C., Lesk, A.M., 1986. The relation between the divergence of
sequence and structure in proteins. EMBO J. 5, 823–826.

Christen, M., Hünenberger, P.H., Bakowies, D., Baron, R., Bürgi, R., Geerke,
D.P., Heinz, T.N., Kastenholz, M.A., Kräutler, V., Oostenbrink, C., Peter,
C., Trzesniak, D., van Gunsteren, W.F., 2005. The GROMOS software for
biomolecular simulation: GROMOS05. J. Comput. Chem. 26, 1719–51.
doi:10.1002/jcc.20303

Chun, O.K., Chung, S.J., Song, W.O., 2007. Estimated dietary flavonoid intake
and major food sources of U.S. adults. J. Nutr. 137, 1244–1252.

Cosconati, S., Forli, S., Perryman, A.L., Harris, R., Goodsell, D.S., Olson, A.J.,
2010. Virtual Screening with AutoDock: Theory and Practice. Expert Opin.
Drug Discov. 5, 597–607. doi:10.1517/17460441.2010.484460

Couse, J.F., Lindzey, J., Grandien, K., Gustafsson, J.A., Korach, K.S., 1997.
Tissue distribution and quantitative analysis of estrogen receptor-alpha
(ERalpha) and estrogen receptor-beta (ERbeta) messenger ribonucleic
acid in the wild-type and ERalpha-knockout mouse. Endocrinology 138,
4613–4621. doi:10.1210/endo.138.11.5496

Cowley, S.M., Hoare, S., Mosselman, S., Parker, M.G., 1997. Estrogen
receptors alpha and beta form heterodimers on DNA. J. Biol. Chem. 272,
19858–62.

Dahlqvist, M., Meshkian, R., Rosen, J., 2017. Dataset on the structure and
thermodynamic and dynamic stability of Mo2ScAlC2 from experiments
and first-principles calculations. Data Br. 10, 576–582.
doi:10.1016/j.dib.2016.12.046

Dao, K.D., Hanson, R.N., 2012. Targeting the Estrogen Receptor using
Steroid-Therapeutic Drug Conjugates. Bioconjug. Chem.

Davies, N.J., Batehup, L., Thomas, R., 2011. The role of diet and physical
activity in breast, colorectal, and prostate cancer survivorship: a review of
the literature. Br. J. Cancer 105 Suppl, S52-73. doi:10.1038/bjc.2011.423

Dixon, J.S., Oshiro, C.M., 1995. Flexible ligand docking using a genetic
algorithm. J. Comput. Aided Mol. Des. 9, 113–130.

Dixon, R.A., 2004. Phytoestrogens. Annu. Rev. Plant Biol. 55, 225–261.
doi:10.1146/annurev.arplant.55.031903.141729

Dong, J.-Y., Qin, L.-Q., 2011. Soy isoflavones consumption and risk of breast

    cancer incidence or recurrence: a meta-analysis of prospective studies.

    Breast Cancer Res. Treat. 125, 315–323. doi:10.1007/s10549-010-1270-8

Dowsett, M., Howell, A., 2002. Breast cancer: aromatase inhibitors take on

    tamoxifen. Nat Med 8.

Elokely, K.M., Doerksen, R.J., 2013. Docking challenge: Protein sampling and

    molecular docking performance. J. Chem. Inf. Model. 53, 1934–1945.

    doi:10.1021/ci400040d

Eyster, K.M., 2016. The Estrogen Receptors: An Overview from Different

    Perspectives, in: Eyster, K.M. (Ed.), Estrogen Receptors: Methods and

    Protocols. Springer New York, New York, NY, pp. 1–10. doi:10.1007/978-

    1-4939-3127-9_1

Fan, H., Irwin, J.J., Webb, B.M., Klebe, G., Shoichet, B.K., Sali, A., 2009.

    Molecular Docking Screens Using Comparative Models of Proteins. J.

    Chem. Inf. Model. 49, 2512–2527. doi:10.1021/ci9003706

Farina, H.G., Pomies, M., Alonso, D.F., Gomez, D.E., 2006. Antitumor and

    antiangiogenic activity of soy isoflavone genistein in mouse models of

    melanoma and breast cancer. Oncol. Rep. 16, 885–891.

Feixas, F., Lindert, S., Sinko, W., McCammon, J.A., 2014. Exploring the role of

    receptor flexibility in structure-based drug discovery. Biophys. Chem. 186,

    31–45. doi:10.1016/j.bpc.2013.10.007

Ferlay, J., Parkin, D.M., Steliarova-Foucher, E., 2010. Estimates of cancer

    incidence and mortality in Europe in 2008. Eur. J. Cancer 46, 765–81.

    doi:10.1016/j.ejca.2009.12.014

Fernandez, S. V, Russo, J., 2010. Estrogen and xenoestrogens in breast cancer. Toxicol. Pathol. 38, 110–22. doi:10.1177/0192623309354108

Fixemer, T., Remberger, K., Bonkhoff, H., 2003. Differential expression of the estrogen receptor beta (ERβ) in human prostate tissue, premalignant changes, and in primary, metastatic, and recurrent prostatic adenocarcinoma. Prostate 54, 79–87. doi:10.1002/pros.10171

Flötotto, T., Niederacher, D., Hohmann, D., Heimerzheim, T., Dall, P., Djahansouzi, S., Bender, H.G., Hanstein, B., 2004. Molecular mechanism of estrogen receptor (ER)alpha-specific, estradiol-dependent expression of the progesterone receptor (PR) B-isoform. J. Steroid Biochem. Mol. Biol. 88, 131–42. doi:10.1016/j.jsbmb.2003.11.004

Flouriot, G., Brand, H., Denger, S., Metivier, R., Kos, M., Reid, G., Sonntag-Buck, V., Gannon, F., 2000. Identification of a new isoform of the human estrogen receptor-alpha (hER-alpha) that is encoded by distinct transcripts and that is able to repress hER-alpha activation function 1., in: The EMBO Journal. England, pp. 4688–4700. doi:10.1093/emboj/19.17.4688

Forli, S., Huey, R., Pique, M.E., Sanner, M.F., Goodsell, D.S., Olson, A.J., 2016. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. Nat. Protoc. 11, 905–919.

Furr, B.J., Jordan, V.C., 1984. The pharmacology and clinical uses of tamoxifen. Pharmacol. Ther. 25, 127–205.

Ganry, O., 2005. Phytoestrogens and prostate cancer risk. Prev. Med. (Baltim). 41, 1–6. doi:10.1016/j.ypmed.2004.10.022

Geromichalos, G.D., 2007. Importance of molecular computer modeling in anticancer drug development. J. BUON. 12 Suppl 1, S101-18.

Giulivo, M., Lopez de Alda, M., Capri, E., Barcelo, D., 2016. Human exposure to endocrine disrupting compounds: Their role in reproductive systems, metabolic syndrome and breast cancer. A review. Environ. Res. 151, 251–264. doi:10.1016/j.envres.2016.07.011

Grace, P.B., Mistry, N.S., Carter, M.H., Leathem, A.J.C., Teale, P., 2007. High throughput quantification of phytoestrogens in human urine and serum using liquid chromatography/tandem mass spectrometry (LC-MS/MS). J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci. 853, 138–46. doi:10.1016/j.jchromb.2007.03.011

Green, S., Walter, P., Kumar, V., Krust, A., Bornert, J.-M., Argos, P., Chambon, P., Vyas, V.K., Ukawala, R.D., Ghate, M., Chintha, C., 1986. Human oestrogen receptor cDNA: sequence, expression and homology to v-erb-A. Nature 320, 134–139. doi:10.4103/0250-474X.102537

Greene, G.L., Gilna, P., Waterfield, M., Baker, A., Hort, Y., Shine, J., 1986. Sequence and expression of human estrogen receptor complementary DNA. Science 231, 1150–1154.

Gruber, C.J., Gruber, D.M., Gruber, I.M.L., Wieser, F., Huber, J.C., 2004. Anatomy of the estrogen response element. Trends Endocrinol. Metab. 15, 73–78. doi:http://dx.doi.org/10.1016/j.tem.2004.01.008

Hall, J.M., Couse, J.F., Korach, K.S., 2001. The multifaceted mechanisms of estradiol and estrogen receptor signaling. J. Biol. Chem. 276, 36869–72. doi:10.1074/jbc.R100029200

Hanahan, D., Weinberg, R.A., 2011. Hallmarks of cancer: the next generation. Cell 144, 646–674.

Hanahan, D., Weinberg, R.A., 2000. The hallmarks of cancer. Cell 100, 57–70.

Harris, D., Besselink, E., 2005. Phytoestrogens induce differential estrogen receptor alpha-or beta-mediated responses in transfected breast cancer cells. Exp. Biol. Med. 230, 558–568.

Holford, T.R., Cronin, K. a, Mariotto, A.B., Feuer, E.J., 2006. Changing patterns in breast cancer incidence trends. J. Natl. Cancer Inst. Monogr. 8034, 19–25. doi:10.1093/jncimonographs/lgj016

Howell, A., 2006. Fulvestrant ("Faslodex"): Current and future role in breast cancer management. Crit. Rev. Oncol. Hematol. 57, 265–273. doi:10.1016/j.critrevonc.2005.08.001

Huang, P., Chandra, V., Rastinejad, F., 2010. Structural overview of the nuclear receptor superfamily: insights into physiology and therapeutics. Annu. Rev. Physiol. 72, 247–72. doi:10.1146/annurev-physiol-021909-135917

Hüsing, A., Canzian, F., Beckmann, L., Garcia-Closas, M., Diver, W.R., Thun, M.J., Berg, C.D., Hoover, R.N., Ziegler, R.G., Figueroa, J.D., Isaacs, C., Olsen, A., Viallon, V., Boeing, H., Masala, G., Trichopoulos, D., Peeters, P.H.M., Lund, E., Ardanaz, E., Khaw, K.-T., Lenner, P., Kolonel, L.N., Stram, D.O., Le Marchand, L., McCarty, C. a, Buring, J.E., Lee, I.-M., Zhang, S., Lindström, S., Hankinson, S.E., Riboli, E., Hunter, D.J., Henderson, B.E., Chanock, S.J., Haiman, C. a, Kraft, P., Kaaks, R., 2012. Prediction of breast cancer risk by genetic risk factors, overall and by

hormone receptor status. J. Med. Genet. 49, 601–8.
doi:10.1136/jmedgenet-2011-100716

Hwang, K.-A., Choi, K.-C., 2015. Anticarcinogenic Effects of Dietary
Phytoestrogens and Their Chemopreventive Mechanisms. Nutr. Cancer
67, 796–803. doi:10.1080/01635581.2015.1040516

Jensen, E., Jacobson, H., 1962. Basic guides to the mechanism of estrogen
action. Recent Prog Horm Res 18, 387–414.

Jiang, Q., Zheng, S., Wang, G., 2013. Development of new estrogen receptor-
targeting therapeutic agents for tamoxifen-resistant breast cancer. Future
Med. Chem. 5, 1023–1035. doi:10.4155/fmc.13.63

Jones, D.T., 1999. Protein secondary structure prediction based on position-
specific scoring matrices. J. Mol. Biol. 292, 195–202.
doi:10.1006/jmbi.1999.3091

Jordan, V.C., 2004. Selective estrogen receptor modulation: concept and
consequences in cancer. Cancer Cell 5, 207–213.

Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure:
pattern recognition of hydrogen-bonded and geometrical features.
Biopolymers 22, 2577–2637. doi:10.1002/bip.360221211

Khoury, G.A., Smadbeck, J., Kieslich, C.A., Floudas, C.A., 2014. Protein
folding and de novo protein design for biotechnological applications.
Trends Biotechnol. 32, 99–109. doi:10.1016/j.tibtech.2013.10.008

Kim, H., Peterson, T.G., Barnes, S., 1998. Mechanisms of action of the soy
isoflavone genistein: emerging role for its effects via transforming growth

factor beta signaling pathways. Am. J. Clin. Nutr. 68, 1418S–1425S.

King, J., Wynne, C.H., Assersohn, L., Jones, A., 2011. Hormone replacement therapy and women with premature menopause--a cancer survivorship issue. Eur. J. Cancer 47, 1623–32. doi:10.1016/j.ejca.2011.04.007

Kitchen, D.B., Decornez, H., Furr, J.R., Bajorath, J., 2004. Docking and scoring in virtual screening for drug discovery: methods and applications. Nat. Rev. Drug Discov. 3, 935–949.

Koo, J., Cabarcas-Petroski, S., Petrie, J.L., Diette, N., White, R.J., Schramm, L., 2015. Induction of proto-oncogene BRF2 in breast cancer cells by the dietary soybean isoflavone daidzein. BMC Cancer 15, 905. doi:10.1186/s12885-015-1914-5

Kuiper, G.G., Enmark, E., Pelto-Huikko, M., Nilsson, S., Gustafsson, J.A., 1996. Cloning of a novel receptor expressed in rat prostate and ovary. Proc. Natl. Acad. Sci. U. S. A.

Kuiper, G.G., Lemmen, J.G., Carlsson, B., Corton, J.C., Safe, S.H., van der Saag, P.T., van der Burg, B., Gustafsson, J.A., 1998. Interaction of estrogenic chemicals and phytoestrogens with estrogen receptor beta. Endocrinology 139, 4252–4263. doi:10.1210/endo.139.10.6216

Kumar, R., Zakharov, M.N., Khan, S.H., Miki, R., Jang, H., Toraldo, G., Singh, R., Bhasin, S., Jasuja, R., 2011. The dynamic structure of the estrogen receptor. J. Amino Acids. doi:10.4061/2011/812540

Kurokawa, R., Mats, S., Hörlein, A., Halachmi, S., Brown, M., Rosenfeld, M.G., Glass, C.K., 1995. Polarity-specific activities of retinoic acid receptors determined by a co-repressor. Nature 377, 451–454.

doi:10.1038/377451a0

Kushner, P.J., Agard, D. a, Greene, G.L., Scanlan, T.S., Shiau, a K., Uht, R.M., Webb, P., 2000. Estrogen receptor pathways to AP-1. J. Steroid Biochem. Mol. Biol. 74, 311–7.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291. doi:10.1107/S0021889892009944

Leong, S.P.L., Shen, Z.-Z., Liu, T.-J., Agarwal, G., Tajima, T., Paik, N.-S., Sandelin, K., Derossis, A., Cody, H., Foulkes, W.D., 2010. Is breast cancer the same disease in Asian and Western countries? World J. Surg. 34, 2308–24. doi:10.1007/s00268-010-0683-1

Liu, J., Xu, Z., Ma, X., Huang, B., Pan, X., 2015. Role of ER-α36 in breast cancer by typical xenoestrogens. Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med. 36, 7355–7364. doi:10.1007/s13277-015-4006-x

López, T., López, S., Arias, C.F., 2015. The tyrosine kinase inhibitor genistein induces the detachment of rotavirus particles from the cell surface. Virus Res. 210, 141–148. doi:10.1016/j.virusres.2015.07.020

Lu, L.J., Anderson, K.E., 1998. Sex and long-term soy diets affect the metabolism and excretion of soy isoflavones in humans. Am. J. Clin. Nutr. 68, 1500S–1504S.

Lund, T.D., Munson, D.J., Haldy, M.E., Setchell, K.D.R., Lephart, E.D., Handa, R.J., 2004. Equol is a novel anti-androgen that inhibits prostate growth

and hormone feedback. Biol. Reprod. 70, 1188–95.
doi:10.1095/biolreprod.103.023713

Maddams, J., Utley, M., Moller, H., 2012. Projections of cancer prevalence in
the United Kingdom, 2010-2040. Br J Cancer 107, 1195–1202.

Manas, E.S., Xu, Z.B., Unwalla, R.J., Somers, W.S., 2004. Understanding the
selectivity of genistein for human estrogen receptor-beta using X-ray
crystallography and computational methods. Structure 12, 2197–2207.
doi:10.1016/j.str.2004.09.015

Mandušić, V., Nikolić-Vukosavljević, D., Tanić, N., Kanjer, K., Neškovic-
Konstantinović, Z., Čeleketić, D., Dimitrijević, B., 2007a. Expression of
estrogen receptor β wt isoform (ERβ1) and ERβΔ5 splice variant mRNAs
in sporadic breast cancer. J. Cancer Res. Clin. Oncol. 133, 571–579.
doi:10.1007/s00432-007-0209-x

Mandušić, V., Nikolić-Vukosavljević, D., Tanić, N., Ksenija, K., Neškovic-
Konstantinović, Z., Čeleketić, D., Dimitrijević, B., 2007b. Expression of
estrogen receptor β wt isoform (ERβ1) and ERβΔ5 splice variant mRNAs
in sporadic breast cancer. J. Cancer Res. Clin. Oncol. 133, 571–579.

Marino, M., Galluzzo, P., Ascenzi, P., 2006. Estrogen signaling multiple
pathways to impact gene transcription. Curr. Genomics 7, 497–508.

Mazur, W., 1998. Phytoestrogen content in foods. Baillieres. Clin. Endocrinol.
Metab. 12, 729–742.

McCann, S.E., Hootman, K.C., Weaver, A.M., Thompson, L.U., Morrison, C.,
Hwang, H., Edge, S.B., Ambrosone, C.B., Horvath, P.J., Kulkarni, S.A.,
2012. Dietary intakes of total and specific lignans are associated with

clinical breast tumor characteristics. J. Nutr. 142, 91–98.

doi:10.3945/jn.111.147264

McDonnell, D., Wardell, S., 2010. The molecular mechanisms underlying the

pharmacological actions of ER modulators: implications for new drug

discovery in breast cancer. Curr. Opin. Pharmacol. 10, 620–8.

doi:10.1016/j.coph.2010.09.007

Melo, F., Feytmans, E., 1998. Assessing protein structures with a non-local

atomic interaction energy. J. Mol. Biol. 277, 1141–1152.

doi:10.1006/jmbi.1998.1665

Meng, X.-Y., Zhang, H.-X., Mezei, M., Cui, M., 2011. Molecular Docking: A

powerful approach for structure-based drug discovery. Curr. Comput.

Aided. Drug Des. 7, 146–157.

Middleton, E.J., Kandaswami, C., Theoharides, T.C., 2000. The effects of

plant flavonoids on mammalian cells: implications for inflammation, heart

disease, and cancer. Pharmacol. Rev. 52, 673–751.

Moore, J.T., McKee, D.D., Slentz-Kesler, K., Moore, L.B., Jones, S. a, Horne,

E.L., Su, J.L., Kliewer, S. a, Lehmann, J.M., Willson, T.M., 1998. Cloning

and characterization of human estrogen receptor beta isoforms. Biochem.

Biophys. Res. Commun. 247, 75–8. doi:10.1006/bbrc.1998.8738

Moras, D., Gronemeyer, H., 1998. The nuclear receptor ligand-binding

domain: structure and function. Curr. Opin. Cell Biol. 10, 384–391.

doi:10.1016/S0955-0674(98)80015-X

Morris, A.L., MacArthur, M.W., Hutchinson, E.G., Thornton, J.M., 1992.

Stereochemical quality of protein structure coordinates. Proteins 12, 345–

364. doi:10.1002/prot.340120407

Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., Olson, A.J., 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comput. Chem. 19, 1639–1662. doi:10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B

Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J., 2009. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. J. Comput. Chem. 30, 2785–2791. doi:10.1002/jcc.21256

Mortensen, A., Kulling, S.E., Schwartz, H., Rowland, I., Ruefer, C.E., Rimbach, G., Cassidy, A., Magee, P., Millar, J., Hall, W.L., Kramer Birkved, F., Sorensen, I.K., Sontag, G., 2009. Analytical and compositional aspects of isoflavones in food and their biological effects. Mol. Nutr. Food Res. 53 Suppl 2, S266-309. doi:10.1002/mnfr.200800478

Mosselman, S., Polman, J., Dijkema, R., 1996. ER beta: identification and characterization of a novel human estrogen receptor. FEBS Lett. 392, 49–53.

Mueller, S.O., Simon, S., Chae, K., Metzler, M., Korach, K.S., 2004. Phytoestrogens and their human metabolites show distinct agonistic and antagonistic properties on estrogen receptor ?? (ER??) and ER?? in human cells. Toxicol. Sci. 80, 14–25. doi:10.1093/toxsci/kfh147

Nechuta, S., Caan, B.J., Chen, W.Y., Kwan, M.L., Lu, W., Cai, H., Poole, E.M., Flatt, S.W., Zheng, W., Pierce, J.P., Shu, X.O., 2013. Postdiagnosis

cruciferous vegetable consumption and breast cancer outcomes: a report from the After Breast Cancer Pooling Project. Cancer Epidemiol. Biomarkers Prev. 22, 1451–1456. doi:10.1158/1055-9965.EPI-13-0446

Nettles, K.W., Bruning, J.B., Gil, G., Nowak, J., Sharma, S.K., Hahm, J.B., Kulp, K., Hochberg, R.B., Zhou, H., Katzenellenbogen, J.A., Katzenellenbogen, B.S., Kim, Y., Joachmiak, A., Greene, G.L., 2008. NFkappaB selectivity of estrogen receptor ligands revealed by comparative crystallographic analyses. Nat. Chem. Biol. 4, 241–247. doi:10.1038/nchembio.76

Nilsson, S., Koehler, K.F., Gustafsson, J.-åke, 2011. Development of subtype-selective oestrogen receptor-based therapeutics 10. doi:10.1038/nrd3551

O'Boyle, N.M., Banck, M., James, C.A., Morley, C., Vandermeersch, T., Hutchison, G.R., 2011. Open Babel: An open chemical toolbox. J. Cheminform. 3, 33. doi:10.1186/1758-2946-3-33

Ogawa, S., Inoue, S., Watanabe, T., Orimo, A., Hosoi, T., Ouchi, Y., Muramatsu, M., 1998. Molecular cloning and characterization of human estrogen receptor βcx: A potential inhibitor of estrogen action in human. Nucleic Acids Res. 26, 3505–3512. doi:10.1093/nar/26.15.3505

Ohshiro, K., Mudvari, P., Meng, Q., Rayala, S.K., Sahin, A. a, Fuqua, S. a W., Kumar, R., 2010. Identification of a novel estrogen receptor-alpha variant and its upstream splicing regulator. Mol. Endocrinol. 24, 914–22. doi:10.1210/me.2009-0413

Onitilo, A.A., Engel, J.M., Greenlee, R.T., Mukesh, B.N., 2009. Breast Cancer Subtypes Based on ER/PR and Her2 Expression: Comparison of

Clinicopathologic Features and Survival. Clin. Med. Res. 7, 4–13. doi:10.3121/cmr.2009.825

Palmieri, C., Cheng, G.J., Saji, S., Zelada-Hedman, M., Warri, A., Weihua, Z., Van Noorden, S., Wahlstrom, T., Coombes, R.C., Warner, M., Gustafsson, J.-A., 2002. Estrogen receptor beta in breast cancer. Endocr. Relat. Cancer 9, 1–13.

Pasinetti, G.M., Ho, L., Cheng, H., Wang, J., Simon, J.E., Wu, Q.-L., Zhao, D., Carry, E., Ferruzzi, M.G., Faith, J., Valcarcel, B., Hao, K., 2017. A comprehensive database and analysis framework to incorporate multiscale data types and enable integrated analysis of bioactive polyphenols. Mol. Pharm. doi:10.1021/acs.molpharmaceut.7b00412

Pearce, S.T., Jordan, V.C., 2004. The biological role of estrogen receptors alpha and beta in cancer. Crit. Rev. Oncol. Hematol. 50, 3–22. doi:10.1016/j.critrevonc.2003.09.003

Peeters, P.H.M., Keinan-Boker, L., van der Schouw, Y.T., Grobbee, D.E., 2003. Phytoestrogens and breast cancer risk. Review of the epidemiological evidence. Breast Cancer Res. Treat. 77, 171–183.

Phillips, T., Murray, G., Wakamiya, K., Askaa, J., Huang, D., Welcher, R., Pii, K., Allred, D.C., 2007. Development of standard estrogen and progesterone receptor immunohistochemical assays for selection of patients for antihormonal therapy. Appl. Immunohistochem. Mol. Morphol. AIMM 15, 325–331. doi:10.1097/01.pai.0000213135.16783.bc

Pike, A.C., Brzozowski, A.M., Hubbard, R.E., Bonn, T., Thorsell, A.G., Engstrom, O., Ljunggren, J., Gustafsson, J.A., Carlquist, M., 1999.

Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist. EMBO J. 18, 4608–4618. doi:10.1093/emboj/18.17.4608

Pratt, W.B., Galigniana, M.D., Morishima, Y., Murphy, P.J.M., 2004. Role of molecular chaperones in steroid receptor action. Essays Biochem. 40, 41 LP-58.

Rafii, F., 2015. The role of colonic bacteria in the metabolism of the natural isoflavone daidzin to equol. Metabolites 5, 56–73. doi:10.3390/metabo5010056

Rao, J., Jiang, X., Wang, Y., Chen, B., 2011. Advances in the understanding of the structure and function of ER-α36,a novel variant of human estrogen receptor-alpha. J. Steroid Biochem. Mol. Biol. 127, 231–7. doi:10.1016/j.jsbmb.2011.08.004

Rice, S., Whitehead, S. a, 2008. Phytoestrogens oestrogen synthesis and breast cancer. J. Steroid Biochem. Mol. Biol. 108, 186–95. doi:10.1016/j.jsbmb.2007.09.003

Romano, B., Pagano, E., Montanaro, V., Fortunato, A.L., Milic, N., Borrelli, F., 2013. Novel Insights into the Pharmacology of Flavonoids. Phyther. Res. doi:10.1002/ptr.5023

Rost, B., 1999. Twilight zone of protein sequence alignments. Protein Eng. 12, 85–94.

Rothwell, J. a, Urpi-Sarda, M., Boto-Ordoñez, M., Knox, C., Llorach, R., Eisner, R., Cruz, J., Neveu, V., Wishart, D., Manach, C., Andres-Lacueva, C., Scalbert, A., 2012. Phenol-Explorer 2.0: a major update of the Phenol-

Explorer database integrating data on polyphenol metabolism and pharmacokinetics in humans and experimental animals. Database (Oxford). 2012, bas031. doi:10.1093/database/bas031

Schiavon, G., Smith, I.E., 2014. Status of adjuvant endocrine therapy for breast cancer. Breast Cancer Res. 16, 206. doi:10.1186/bcr3636

Schmidt, T., Bergner, A., Schwede, T., 2014. Modelling three-dimensional protein structures for applications in drug design. Drug Discov. Today 19, 890–897. doi:10.1016/j.drudis.2013.10.027

Schrodinger LLC, 2015. The PyMOL Molecular Graphics System, Version 1.8.

Seibold, P., Vrieling, A., Johnson, T.S., Buck, K., Behrens, S., Kaaks, R., Linseisen, J., Obi, N., Heinz, J., Flesch-Janys, D., Chang-Claude, J., 2014. Enterolactone concentrations and prognosis after postmenopausal breast cancer: assessment of effect modification and meta-analysis. Int. J. cancer 135, 923–933. doi:10.1002/ijc.28729

Setchell, K., Clerici, C., 2010. Equol: history, chemistry, and formation. J. Nutr. 3, 1355–1362. doi:10.3945/jn.109.119776.1355S

Shanle, E.K., Xu, W., 2010. Selectively targeting estrogen receptors for cancer treatment. Adv. Drug Deliv. Rev. 62, 1265–1276. doi:10.1016/j.addr.2010.08.001

Shao, Z.M., Wu, J., Shen, Z.Z., Barsky, S.H., 1998. Genistein inhibits both constitutive and EGF-stimulated invasion in ER-negative human breast carcinoma cell lines. Anticancer Res. 18, 1435–1439.

Shiau, A.K., Barstad, D., Loria, P.M., Cheng, L., Kushner, P.J., Agard, D.A., Greene, G.L., 1998. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. Cell 95, 927–937.

Shike, M., Doane, A.S., Russo, L., Cabal, R., Reis-Filho, J.S., Gerald, W., Cody, H., Khanin, R., Bromberg, J., Norton, L., 2014. The effects of soy supplementation on gene expression in breast cancer: a randomized placebo-controlled study. J. Natl. Cancer Inst. 106. doi:10.1093/jnci/dju189

Shimizu, H., Ross, R.K., Bernstein, L., Yatani, R., Henderson, B.E., Mack, T.M., 1991. Cancers of the prostate and breast among Japanese and white immigrants in Los Angeles County. Br. J. Cancer 63, 963–966.

Shou, J., Massarweh, S., Osborne, C.K., Wakeling, a. E., Ali, S., Weiss, H., Schiff, R., 2004. Mechanisms of Tamoxifen Resistance: Increased Estrogen Receptor-HER2/neu Cross-Talk in ER/HER2-Positive Breast Cancer. JNCI J. Natl. Cancer Inst. 96, 926–935. doi:10.1093/jnci/djh166

Shu, X.O., Zheng, Y., Cai, H., Gu, K., Chen, Z., Zheng, W., Lu, W., 2009. Soy food intake and breast cancer survival. JAMA 302, 2437–43. doi:10.1001/jama.2009.1783

Simoncini, T., Genazzani, A.R., 2003. Non-genomic actions of sex steroid hormones. Eur. J. Endocrinol. 148, 281–92.

Smiley, D., Khalil, R., 2009. Estrogenic compounds, estrogen receptors and vascular cell signaling in the aging blood vessels. Curr. Med. Chem. 16, 1863–1887.

Smith, R., Hubbard, R.E., Gschwend, D.A., Leach, A.R., Good, A.C., 2003. Analysis and optimization of structure-based virtual screening protocols 3. New Methods and old problems in scoring function design. J. Mol. Graph. Model. 22, 41–53.

Söding, J., 2004. Protein homology detection by HMM–HMM comparison. Bioinformatics 21, 951–960.

Speirs, V., Adams, I.P., Walton, D.S., Atkin, S.L., 2000. Identification of wild-type and exon 5 deletion variants of estrogen receptor beta in normal human mammary gland. J. Clin. Endocrinol. Metab. 85, 1601–1605. doi:10.1210/jcem.85.4.6493

Swann, R., Perkins, K.A., Velentzis, L.S., Ciria, C., Dutton, S.J., Mulligan, A.A., Woodside, J. V, Cantwell, M.M., Leathem, A.J., Robertson, C.E., Dwek, M. V, 2013. The DietCompLyf study: a prospective cohort study of breast cancer survival and phytoestrogen consumption. Maturitas 75, 232–240. doi:10.1016/j.maturitas.2013.03.018

Tanenbaum, D.M., Wang, Y., Williams, S.P., Sigler, P.B., 1998. Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains. Proc. Natl. Acad. Sci. U. S. A. 95, 5998–6003.

Thomas, C., Gustafsson, J.-A., 2011. The different roles of ER subtypes in cancer biology and therapy. Nat. Rev. Cancer 11, 597–608. doi:10.1038/nrc3093

Toft, D., Gorski, J., 1966. A receptor molecule for estrogens: isolation from the rat uterus and preliminary characterization. Proc. Natl. Acad. Sci. U. S. A. 55, 1574–81. doi:10.1073/pnas.55.6.1574

Trott, O., Olson, A.J., 2010a. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31, 455–61. doi:10.1002/jcc.21334

Trott, O., Olson, A.J., 2010b. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31, 455–461. doi:10.1002/jcc.21334

Uchoa, H.B., Jorge, G.E., Freitas Da Silveira, N.J., Camera, J.C.J., Canduri, F., De Azevedo, W.F.J., 2004. Parmodel: a web server for automated comparative modeling of proteins. Biochem. Biophys. Res. Commun. 325, 1481–1486. doi:10.1016/j.bbrc.2004.10.192

Valsta, L.M., Kilkkinen, A., Mazur, W., Nurmi, T., Lampi, A.-M., Ovaskainen, M.-L., Korhonen, T., Adlercreutz, H., Pietinen, P., 2003. Phyto-oestrogen database of foods and average intake in Finland. Br. J. Nutr. 89 Suppl 1, S31-8. doi:10.1079/BJN2002794

Velentzis, L.S., Keshtgar, M.R., Woodside, J. V, Leathem, A.J., Titcomb, A., Perkins, K.A., Mazurowska, M., Anderson, V., Wardell, K., Cantwell, M.M., 2011. Significant changes in dietary intake and supplement use after breast cancer diagnosis in a UK multicentre study. Breast Cancer Res. Treat. 128, 473–482.

Vyas, V.K., Ukawala, R.D., Ghate, M., Chintha, C., 2012. Homology modeling a fast tool for drug discovery: current perspectives. Indian J. Pharm. Sci. 74, 1–17. doi:10.4103/0250-474X.102537

Wang, T.T., Sathyamoorthy, N., Phang, J.M., 1996. Molecular effects of

genistein on estrogen receptor mediated pathways. Carcinogenesis 17, 271–275.

Wang, Y., Chirgadze, N.Y., Briggs, S.L., Khan, S., Jensen, E. V, Burris, T.P., 2006. A second binding site for hydroxytamoxifen within the coactivator-binding groove of estrogen receptor beta. Proc. Natl. Acad. Sci. U. S. A. 103, 9908–9911. doi:10.1073/pnas.0510596103

Wang, Z., Zhang, X., Shen, P., Loggie, B.W., Chang, Y., Deuel, T.F., 2005. Identification, cloning, and expression of human estrogen receptor-alpha36, a novel variant of human estrogen receptor-alpha66. Biochem. Biophys. Res. Commun. 336, 1023–1027. doi:10.1016/j.bbrc.2005.08.226

Warnmark, A., Treuter, E., Gustafsson, J.-A., Hubbard, R.E., Brzozowski, A.M., Pike, A.C.W., 2002. Interaction of transcriptional intermediary factor 2 nuclear receptor box peptides with the coactivator binding site of estrogen receptor alpha. J. Biol. Chem. 277, 21862–21868. doi:10.1074/jbc.M200764200

Webb, P., Lopez, G.N., Uht, R.M., Kushner, P.J., 1995. Tamoxifen activation of the estrogen receptor/AP-1 pathway: potential origin for the cell-specific estrogen-like effects of antiestrogens. Mol. Endocrinol. Balt. Md 9, 443–456.

Weigelt, B., Mackay, A., A'hern, R., Natrajan, R., Tan, D.S.P., Dowsett, M., Ashworth, A., Reis-Filho, J.S., 2010. Breast cancer molecular profiling with single sample predictors: a retrospective analysis. Lancet Oncol. 11, 339–49. doi:10.1016/S1470-2045(10)70008-5

Wiederstein, M., Sippl, M.J., 2007. ProSA-web: interactive web service for the

recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 35, W407-10. doi:10.1093/nar/gkm290

Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R.F., Sykes, B.D., Wishart, D.S., 2003. VADAR: a web server for quantitative evaluation of protein structure quality. Nucleic Acids Res. 31, 3316–3319.

World Health Organization WHO, 2012. Data and Statistic [WWW Document]. URL http://www.who.int/research/en/ (accessed 8.12.13).

Wu, A.H., Ursin, G., Koh, W.-P., Wang, R., Yuan, J.-M., Khoo, K.-S., Yu, M.C., 2008. Green tea, soy, and mammographic density in Singapore Chinese women. Cancer Epidemiol. Biomarkers Prev. 17, 3358–65. doi:10.1158/1055-9965.EPI-08-0132

Xu, M., Lill, M.A., 2013. Induced fit docking , and the use of QM / MM methods in docking. Drug Discov. Today Technol. 10, e411–e418. doi:10.1016/j.ddtec.2013.02.003

Yang, J., Roy, A., Zhang, Y., 2013. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. Bioinformatics 29, 2588–2595. doi:10.1093/bioinformatics/btt447

Yang, X., Belosay, A., Hartman, J.A., Song, H., Zhang, Y., Wang, W., Doerge, D.R., Helferich, W.G., 2015. Dietary soy isoflavones increase metastasis to lungs in an experimental model of breast cancer with bone micro-tumors. Clin. Exp. Metastasis 32, 323–333. doi:10.1007/s10585-015-9709-2

Yarger, J.G., Babine, R.E., Bittner, M., Shanle, E., Xu, W., Hershberger, P.,

Nye, S.H., 2013. Structurally similar estradiol analogs uniquely alter the regulation of intracellular signaling pathways. J. Mol. Endocrinol. 50, 43–57. doi:10.1530/JME-12-0083

Yuriev, E., Ramsland, P.A., 2013. Latest developments in molecular docking: 2010-2011 in review. J. Mol. Recognit. 26, 215–239. doi:10.1002/jmr.2266

Zhang, J., Li, G., Li, Z., Yu, X., Zheng, Y., Jin, K., Wang, H., Gong, Y., Sun, X., Teng, X., Cao, J., Teng, L., 2012. Estrogen-independent effects of ER-α36 in ER-negative breast cancer. Steroids 77, 666–673. doi:10.1016/j.steroids.2012.02.013

Ziegler, R.G., Hoover, R.N., Nomura, A.M., West, D.W., Wu, A.H., Pike, M.C., Lake, A.J., Horn-Ross, P.L., Kolonel, L.N., Siiteri, P.K., Fraumeni, J.F.J., 1996. Relative weight, weight change, height, and breast cancer risk in Asian-American women. J. Natl. Cancer Inst. 88, 650–660.

Appendices

Appendix 1

Appendix 1

Estrogen receptor alpha amino acid sequence

Protein uniprot ID and name	Amino acid sequence
P03372 ESR1_HUMAN Estrogen receptor Full length	MTMTLHTKASGMALLHQIQGNELEPLNRPQLKIPLERPLGEVYLDSSKPAVYNYPEGAAYEFNAAAAANAQVYGQTGLPYGPGSEAAAFGSNGLGGFPPLNSVSPSPLM LLHPPPQLSPFLQPHGQQVPYYLENEPSGYTVREAGPPAFYRPNSDNRRQGGRERLASTNDKGSMAMESAKETRYCAVCNDYASGYHYGVWSCEGCKAFFKRSIQ GHNDYMCPATNQCTIDKNRRKSCQACRLRKCYEVGMMKGGIRKDRRGGRMLKHKRQRDDGEGRGEVGSAGDMRAANLWPSPLMIKRSKKNSLALSLTADQMVS ALLDAEPPILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDML LATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEML DAHRLHAPTSRGGASVEETDQSHLATAGSTSSHSLQKYYITGEAEGFPATV
P03372 **ESR1_LBD(302 – 553)252aa** Ligand binding Domain	KKNSLALSLTADQMVSALLDAEPPILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQ GKCVEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNV VPLYDLLLEMLDAHRLHAPT

Table 1. Human estrogen receptor wild type sequence full length and ligand binding domain, as retrived from Unitprot database (06-06-13).

ESR1 cDNA sequence

ATGACCATGACCCTGCATACCAAAGCAAGCGGTATGGCACTGCTGCATCAGATTCAGGGTAATGAACTGGAACCGCTGAATCGTCCGCAGCTGAAAATTC
CGCTGGAACGTCCGCTGGGTGAAGTTTATCTGGATAGCAGCAAACCGGCAGTTTATAACTATCCGGAAGGTGCAGCATATGAATTTAATGCAGCAGCAGC
CGCAAATGCACAGGTTTATGGTCAGACAGGTCTGCCGTATGGTCCGGGTAGCGAAGCAGCAGCATTTGGTAGCAATGGTCTGGGTGGTTTTCCGCCTCT
GAATAGCGTTAGCCCGAGTCCGCTGATGCTGCTGCACCCTCCGCCTCAGCTGTCACCGTTTCTGCAGCCGCATGGCCAGCAGGTTCCGTATTATCTGGA
AAATGAACCGAGCGGTTATACCGTTCGTGAAGCAGGTCCGCCTGCATTTTATCGTCCGAATAGCGATAATCGTCGTCAGGGTGGTCGTGAACGTCTGGC
AAGCACCAATGATAAAGGTAGCATGGCAATGGAAAGCGCAAAAGAAACCCGTTATTGTGCCGTTTGTAATGATTATGCAAGCGGCTATCATTATGGTGTTT
GGAGCTGTGAAGGTTGCAAAGCATTTTTCAAACGTAGCATTCAGGGCCCAACGATTATATGTGTCCGGCAACCAATCAGTGTACCATTGATAAAAATCGT
CGTAAAAGCTGTCAGGCATGTCGTCTGCGTAAATGTTATGAAGTTGGTATGATGAAAGGTGGCATTCGTAAAGATCGTCGTGGTGGTCGTATGCTGAAAC
ACAAACGTCAGCGTGATGATGGTGAAGGTCGTGGTGAAGTGGGTAGTGCCGGTGATATGCGTGCAGCAAATCTGTGGCCGTCACCGCTGATGATTAAAC
GCAGCAAAAAAAACAGCCTGGCACTGAGCCTGACCGCAGATCAGATGGTTAGCGCACTGCTGGATGCAGAACCGCCTATTCTGTATAGCGAATATGATC
CGACCCGTCCGTTTAGTGAAGCAAGCATGATGGGTCTGCTGACCAATCTGGCAGATCGTGAACTGGTTCATATGATTAATTGGGCAAAACGTGTTCCGGG
TTTTGTTGATCTGACACTGCATGATCAGGTTCATCTGCTGGAATGTGCATGGCTGGAAATTCTGATGATTGGTCTGGTTTGGCGTAGCATGGAACATCCG
GGTAAACTGCTGTTTGCACCGAATCTGCTGCTGGATCGTAATCAGGGTAAATGTGTTGAAGGTATGGTGGAAATCTTTGATATGCTGCTGGCAACCAGCA
GCCGTTTTCGCATGATGAATCTGCAGGGTGAAGAATTTGTTTGCCTGAAAAGCATTATTCTGCTGAACAGCGGTGTTTATACCTTTCTGAGCAGCACCCTG
AAAAGTCTGGAAGAAAAAGATCATATCCATCGGGTGCTGGATAAAATCACCGATACGCTGATTCATCTGATGGCAAAAGCAGGTCTGACGCTGCAGCAAC
AGCACCAGCGTCTGGCACAACTGCTGCTGATTCTGAGCCATATTCGTCATATGAGCAATAAAGGTATGGAACACCTGTACAGCATGAAATGCAAAAATGTT
GTTCCGCTGTATGACCTGCTGCTGGAAATGCTGGATGCACATCGTCTGCATGCACCGACCAGCCGTGGTGGTGCAAGCGTTGAAGAAACCGATCAGAGC
CATCTGGCCACCGCAGGTAGCACCAGCAGTCATAGCCTGCAGAAATATTACATTACCGGTGAAGCAGAAGGTTTTCCGGCAACAGTT

Table 2. cDNA Estrogen receptor wild-type full length. Optimized sequence for *E. Coli* rare codon expression.

Primers table

Product name and size	Primer forward 5' -3'	Primer reverse 5' – 3'
pET – SUMO Plasmid	AGATTCTTGTACGACGGTATTAG	TAGTTATTGCTCAGCGGTGG
Ace-pETSumo	ATGAAAAAAAACAGCCTGGCACTGAGC	TAGTTATTGCTCAGCGGTGG
ESR1 WT full length	ATGCCATGACCCTGCATACCAAA	TTAAACTGTTGCCGGAAAACCTTCT
ESR1 LBD	ATGAAAAAAAACAGCCTGGCACTGAGC	TTAGGTCGGTGCATGCAGACGATG

Table 3. Primer designed for pET-SUMO plasmid, ESR1 full length and ESR1 LBD.

Cluster aligment ER alpha ligand binding domain wiltype and isoforms

ESR1	KKNSLALSLTADQMVSALLDAEPPILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWA
ESR1_2	--
ESR1_3	KKNSLALSLTADQMVSALLDAEPPILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWA
ESR1_4	KKNSLALSLTADQMVSALLDAEPPILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWA
ESR1	RVPGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGM
ESR1_2	----------VDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGM
ESR1_3	RVPGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGM
ESR1_4	RVPGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGM
ESR1	VEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDK
ESR1_2	VEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDK
ESR1_3	VEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDK
ESR1_4	VEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGFTISHVE------------AKKRILNLHPK
ESR1	ITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLL
ESR1_2	ITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLL
ESR1_3	ITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLL
ESR1_4	IFGNKWF--P----------------------RV--
ESR1	EMLDAHRLHAPT
ESR1_2	EMLDAHRLHAPT
ESR1_3	EMLDAHRLHAPT
ESR1_4	------------------------

Table 4. Cluster aligment ER alpha ligand binding domain wiltype and isoforms. Isoform 2 is missing KKNSLALSLTADQMVSALLDAEPPILYSEYDPTRPFSEASMMGLLTNLADREL VHMINWAKRVPG) at the LBD, Isoform 3 possess the same LBD than the wild type and isoform 44 possess aa FTISHVEAKKRILNLHPKIFGNKWFPRV in lieu of aa VYTFL…FPATV.

Cluster aligment ER Beta ligand binding domain wild-type and isoforms

```
ESR2     VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLE
ESR2_2 VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLE
ESR2_3 VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLE
ESR2_4 VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLE
ESR2_5 VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLE
ESR2_6 VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLE
ESR2_7 VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLE
ESR2_8 VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLE
ESR2_9 VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLE


ESR2     AEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
ESR2_2 AEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
ESR2_3 AEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGMRGNA--------------------
ESR2_4 AEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
ESR2_5 AEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
ESR2_6 AEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
ESR2_7 AEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPG-------------------------------
ESR2_8 AEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC
ESR2_9 AEPPHVLISRPSAPFTEASMMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESC


ESR2     WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGK--CVEGILEIFDMLLATTSRFRELKL
ESR2_2 WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGK---CVEGILEIFDMLLATTSRFRELKL
ESR2_3 ------------------------------------------------------------------------------------------------------
ESR2_4 WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGK---CVEGILEIFDMLLATTSRFRELKL
ESR2_5 WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGK---CVEGILEIFDMLLATTSRFRELKL
ESR2_6 WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGK---CVEGILEIFDMLLATTSRFRELKL
ESR2_7 ------------------------------------------------------------------------------------------------------
ESR2_8 WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRYVPSGHSDPGC-------------------------------
ESR2_9 WMEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGK---CVEGILEIFDMLLATTSRFRELKL


ESR2     QHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQS
ESR2_2 QHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQS
ESR2_3 --------------------------------------------------------------------------------------------------------
ESR2_4 QHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQS
ESR2_5 QHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQS
ESR2_6 QHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQS
```

```
ESR2_7  -----------------------------MYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQS

ESR2_8  -------------------------------------------------------------------------------------------------------------

ESR2_9 QHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQS


ESR2    MRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSEC

ESR2_2 MRLANLLMLLSHVRHARAEKASQTLTSFGM--------------------------KMET-------LLPE----ATM

ESR2_3 -------------------------------------------------------------------------------------------------------------

ESR2_4 MRLANLLMLLSHVRHASSLSLSWRLFMLRE--------------------------ASCHGVRQTPGGAHM

ESR2_5 MRLANLLMLLSHVRHARWGEKQFIHLKLS--------------------------------------------------------------

ESR2_6 MRLANLLMLLSHVRHARYAP-----------------------------------------------------------------------------

ESR2_7 MRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSEC

ESR2_8 -------------------------------------------------------------------------------------------------------------

ESR2_9 MRLANLLMLLSHVRHARSCVYK--------------------------------------------------------------------------

ESR2    SPAEDSKSKEGSQNPQSQ

ESR2_2 -EQ----------------------------

ESR2_3 --------------------------------

ESR2_5 --------------------------------

ESR2_4 SVSRSRSFEACQQPRE-

ESR2_6 --------------------------------

ESR2_7 SPAEDSKSKEGSQNPQSQ

ESR2_8 --------------------------------

ESR2_9 --------------------------------
```

Table 5. Cluster aligment ER Beta ligand binding domain wild-type and isoforms. Isoform 2 displays SNKGMEHLLN...KEGSQNPQSQ → RAEKASQTLTSFGMKMETLLPEATMEQ, Isoform 3 319-323: FVELS → MRGNA and is missing aa 324-530, Isoform 4 SNKGMEHLLN...KEGSQNPQSQ → SSLSLSWRLF...SFEACQQPRE, Isoform 5 differs from wild type sequence in aa 469-530: SNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSEC

SPAEDSKSKEGSQNPQSQ, Isoform 6 The sequence of this isoform differs from the wild type sequence 469-530: SNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAED

SKSKEGSQNPQSQ, Isoform 7 is missing aa 319-409, Isoform 8 is missing aa 376-530 and displays 365-375: DEGKCVEGILE → YVPSGHSDPGC, Isoform 9 is missing 475-530 and displays 469-474: SNKGME → RSCVYK.

LBD sequences of Estrogen Receptors isoforms

ESR1 Wild Type

KKNSLALSLTADQMVSALLDAEPPILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVE
IFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEMLDAHRL
HAPT

ESR1 Isoform 2

FVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVEIFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIHLMA
KAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEMLDAHRLHAPT

ESR1 Isoform 3

KKNSLALSLTADQMVSALLDAEPPILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVE
IFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGVYTFLSSTLKSLEEKDHIHRVLDKITDTLIHLMAKAGLTLQQQHQRLAQLLLILSHIRHMSNKGMEHLYSMKCKNVVPLYDLLLEMLDAHRL
HAPT

ESR1 Isoform 4

KKNSLALSLTADQMVSALLDAEPPILYSEYDPTRPFSEASMMGLLTNLADRELVHMINWAKRVPGFVDLTLHDQVHLLECAWLEILMIGLVWRSMEHPGKLLFAPNLLLDRNQGKCVEGMVE
IFDMLLATSSRFRMMNLQGEEFVCLKSIILLNSGFTISHVEAKKRILNLHPKIFGNKWFPRV

ESR2 Wild type

VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCW
MEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANL
LMLLSHVRHASNKGMEHLLNMKCKNVV
PVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ

ESR2 Isoform 2

VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCW
MEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANL
LMLLSHVRHARAEKASQTLTSFGMKMETLLPEATMEQ

ESR2 Isoform 3

VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMSLTKLADKELVHMISWAKKIPGMRGNA

ESR2 Isoform 4

VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCW
MEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANL
LMLLSHVRHASSLSLSWRLFMLREASCHGVRQTPGGAHMSVSRSRSFEACQQPRE

ESR2 Isoform 5

VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCW
MEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANL
LMLLSHVRHARWGEKQFIHLKLS

ESR2 Isoform 6

VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCW
MEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANL
LMLLSHVRHARYAP

ESR2 Isoform 7

VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMSLTKLADKELVHMISWAKKIPGMYPLVTATQDADSSRKL
AHLLNAVTDALVWVIAKSGISSQQQSMRLANLLMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ

ESR2 Isoform 8

VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCW
MEVLMMGLMWRSIDHPGKLIFAPDLVLDRYVPSGHSDPGC

ESR2 Isoform 9

VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCW
MEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANL
LMLLSHVRHARSCVYK

ESR2_LBD aa 215-530

VKCGSRRERCGYRLVRRQRSADEQLHCAGKAKRSGGHAPRVRELLLDALSPEQLVLTLLEAEPPHVLISRPSAPFTEASMMSLTKLADKELVHMISWAKKIPGFVELSLFDQVRLLESCW
MEVLMMGLMWRSIDHPGKLIFAPDLVLDRDEGKCVEGILEIFDMLLATTSRFRELKLQHKEYLCVKAMILLNSSMYPLVTATQDADSSRKLAHLLNAVTDALVWVIAKSGISSQQQSMRLANL
LMLLSHVRHASNKGMEHLLNMKCKNVVPVYDLLLEMLNAHVLRGCKSSITGSECSPAEDSKSKEGSQNPQSQ

Table 6. Protein sequences of LBD estrogen receptor isoforms. Isoform 1
was taken as a canonical sequence in both subtypes.

Appendix 2

Appendix 2

2.1 ER-α Wild type

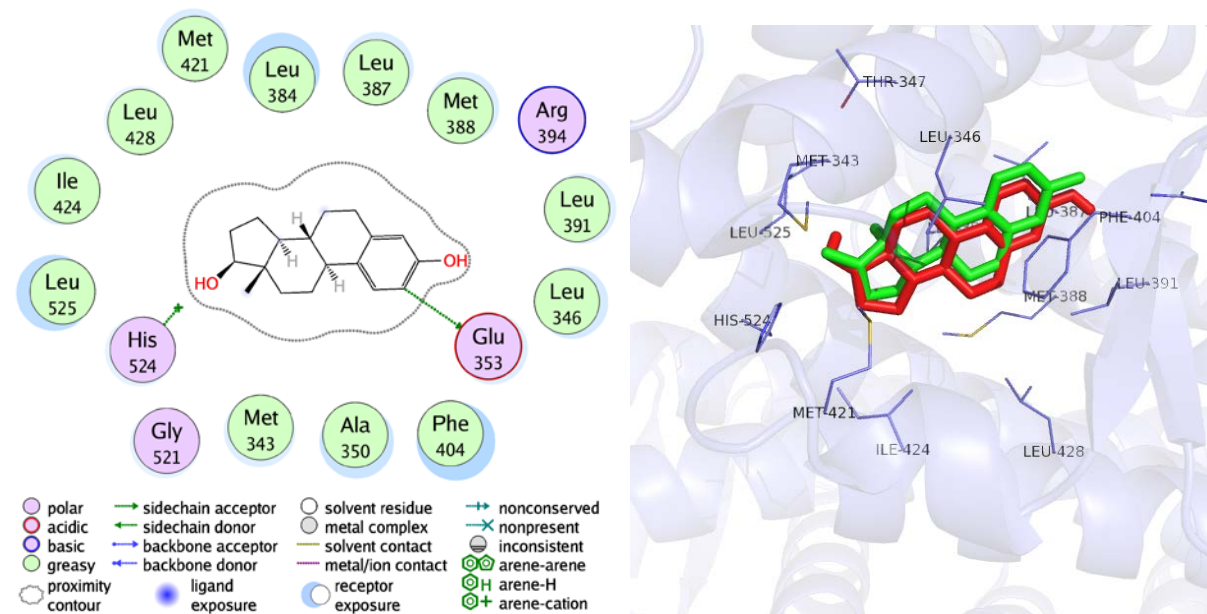**Estradiol vs ER-α WILD TYPE**

Figure 1. Estradiol binding estrogen receptor A, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand. Green ligand on the left and model was taken from PDB database (1GWR) and the red ligand is pose produced by Autodock Vina lowest binding energy (-10.7 Kcal/mol).
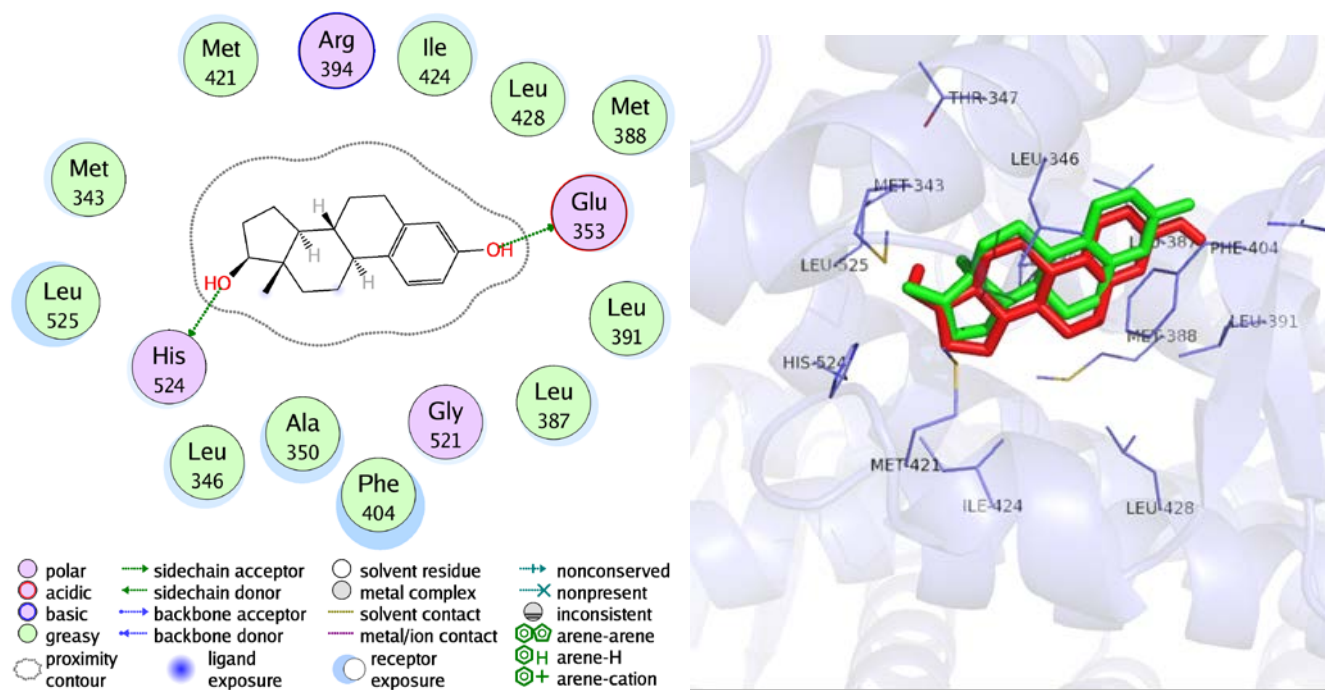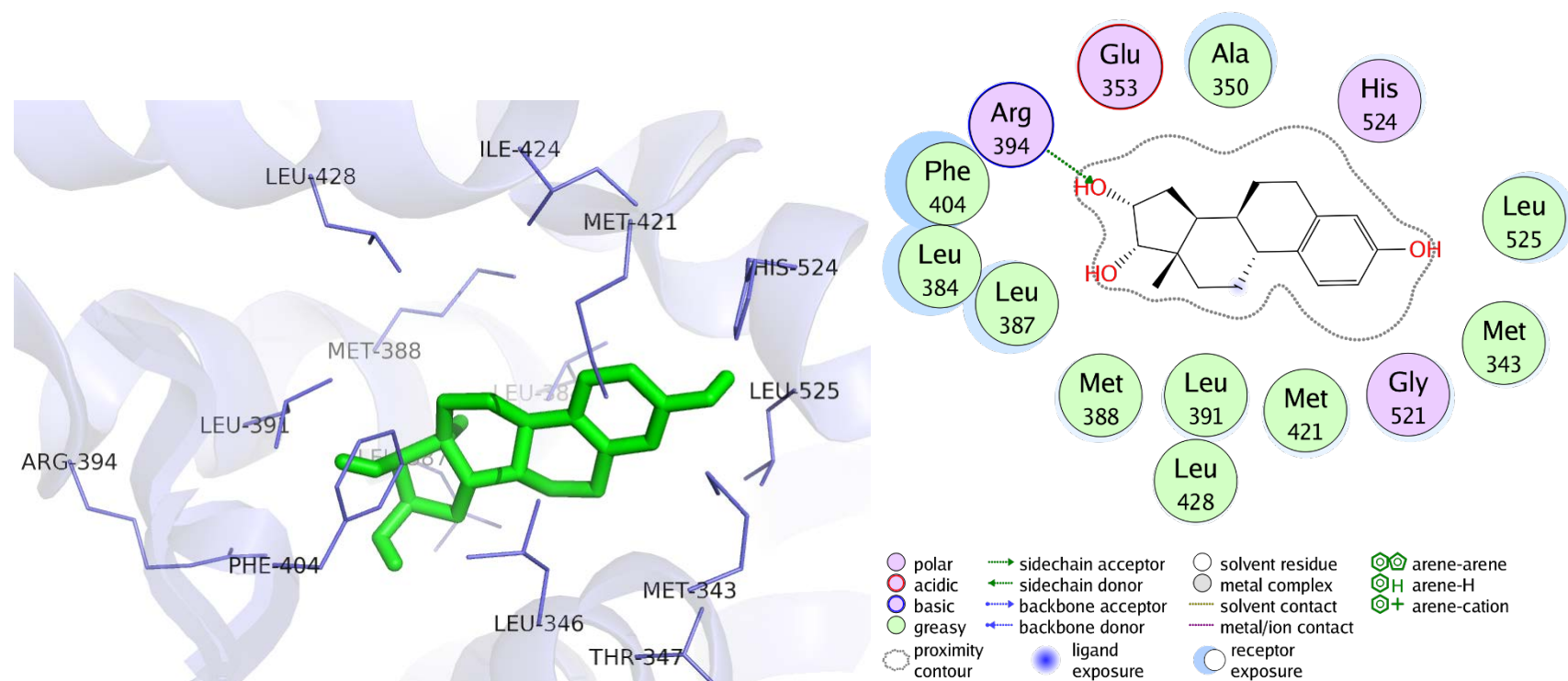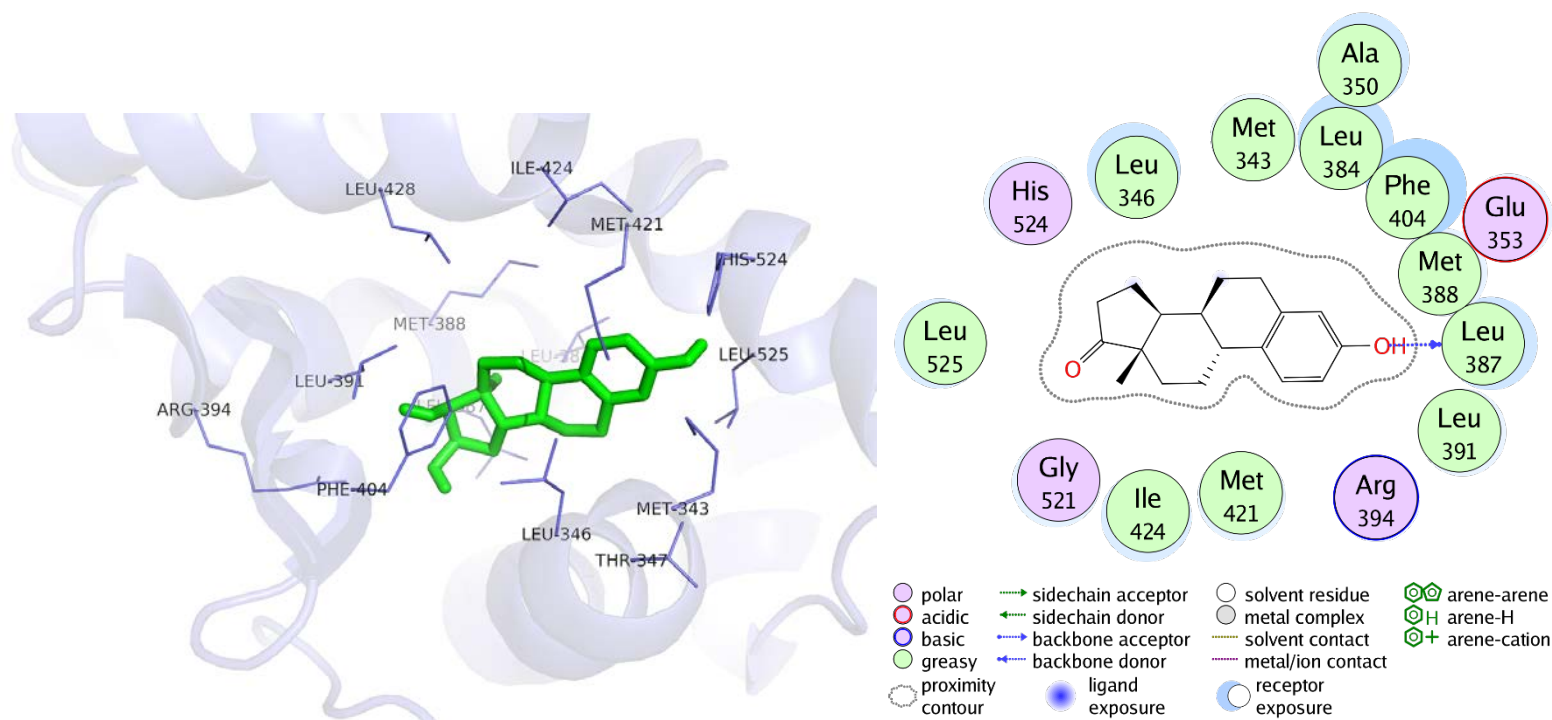
Estradiol vs ER-α WILD TYPE

Figure 2. Estradiol binding estrogen receptor α, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand based on the Autodock Vina pose. Green ligand on the left and model was taken from PDB database (1GWR) and the red ligand is pose produced by Autodock Vina lowest binding energy (-10.7 Kcal/mol).

Estriol vs ER-α WILD TYPE

Figure 3. Estriol binding estrogen receptor α, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand based on the Autodock Vina pose. Green ligand on the left and model was taken from PDB database (1GWR) and the green ligand is pose produced by Autodock Vina lowest binding energy (-10 Kcal/mol).
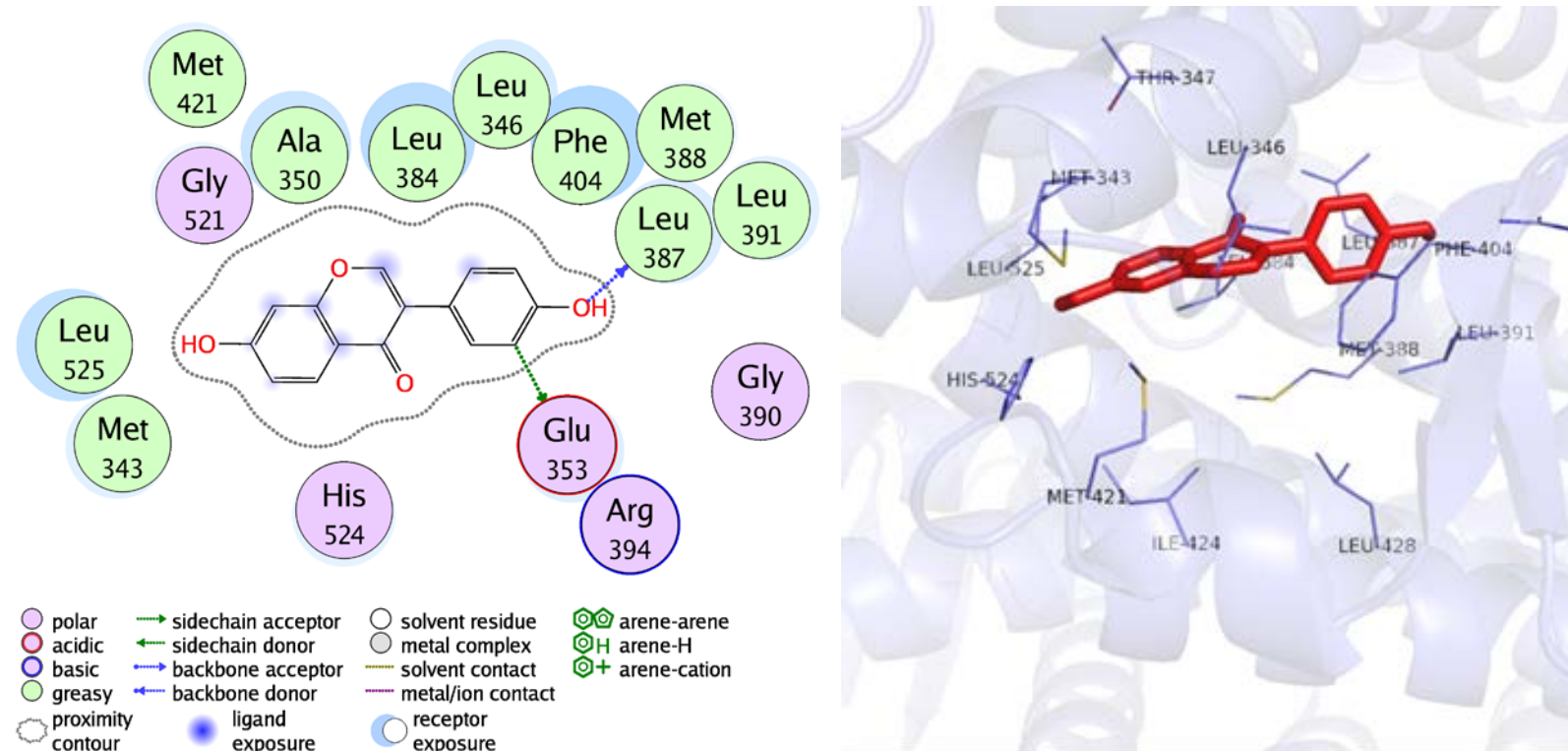
Estrone vs ER-α WILD TYPE

Figure 4. Estrone binding estrogen receptor α, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand based on the Autodock Vina pose. Green ligand on the left and model was taken from PDB database (1GWR) and the green ligand is pose produced by Autodock Vina lowest binding energy (-10.4 Kcal/mol).

141

Daidzein vs ER-α WILD TYPE

Figure 5. Daidzein binding estrogen receptor α, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand. Model template based on agonist conformation and the red ligand is pose produced by Autodock Vina lowest binding energy (-8.4 Kcal/mol).
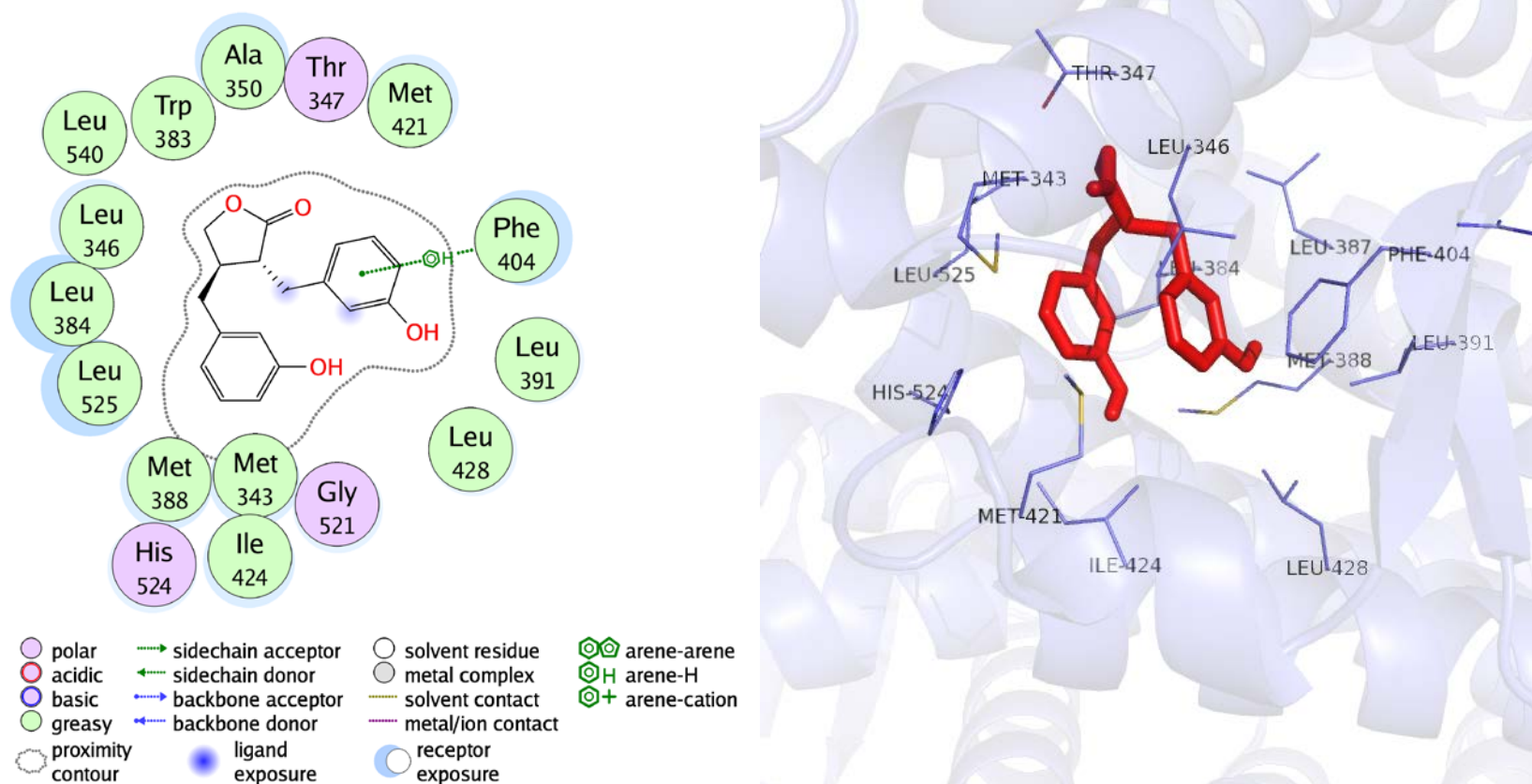
Enterolactone vs ER-α WILD TYPE

Figure 6. Enterolactone binding estrogen receptor α, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand. Model template based on agonist conformation and the red ligand is pose produced by Autodock Vina lowest binding energy (-8.6 Kcal/mol).
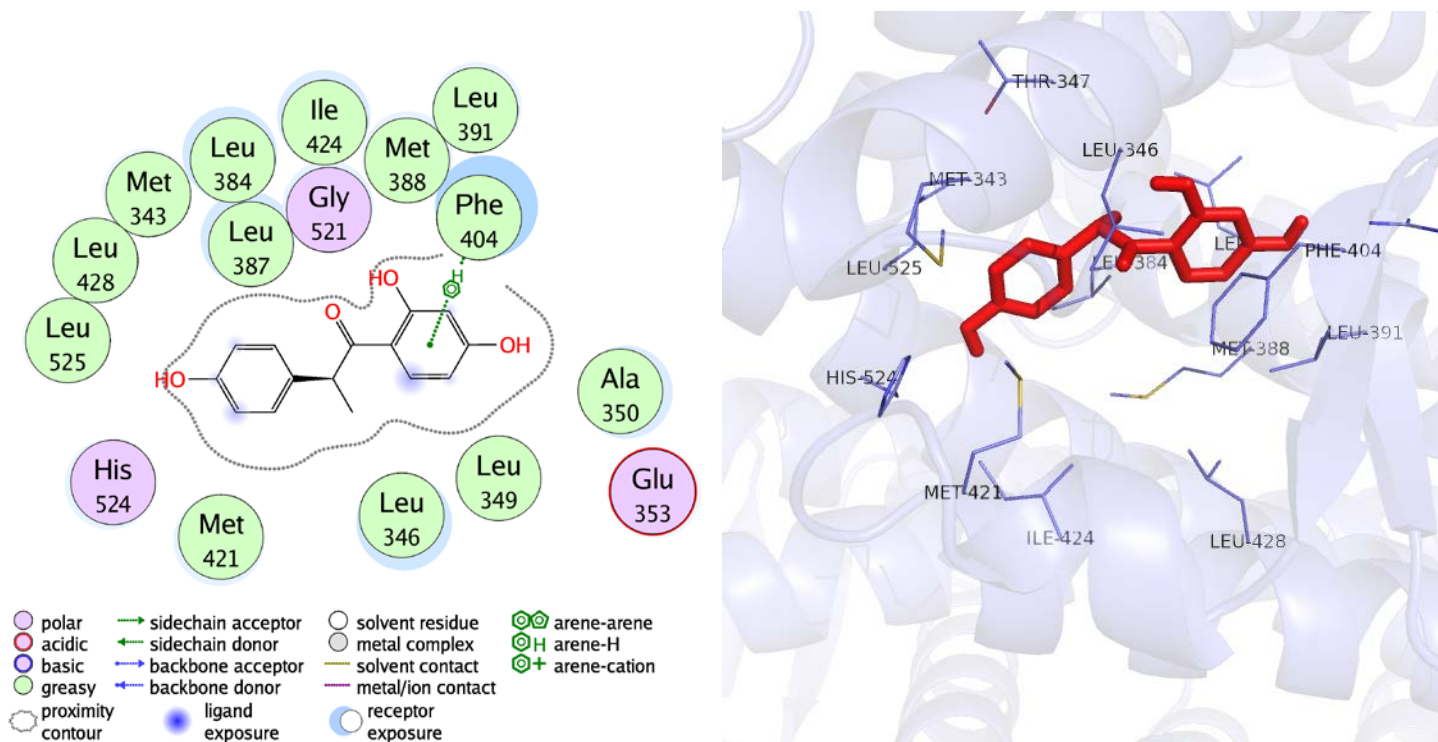
ODMA vs ER-α WILD TYPE

Figure 7. ODMA binding estrogen receptor α, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand. Model template based on agonist conformation and the red ligand is pose produced by Autodock Vina lowest binding energy (-8.5 Kcal/mol) .
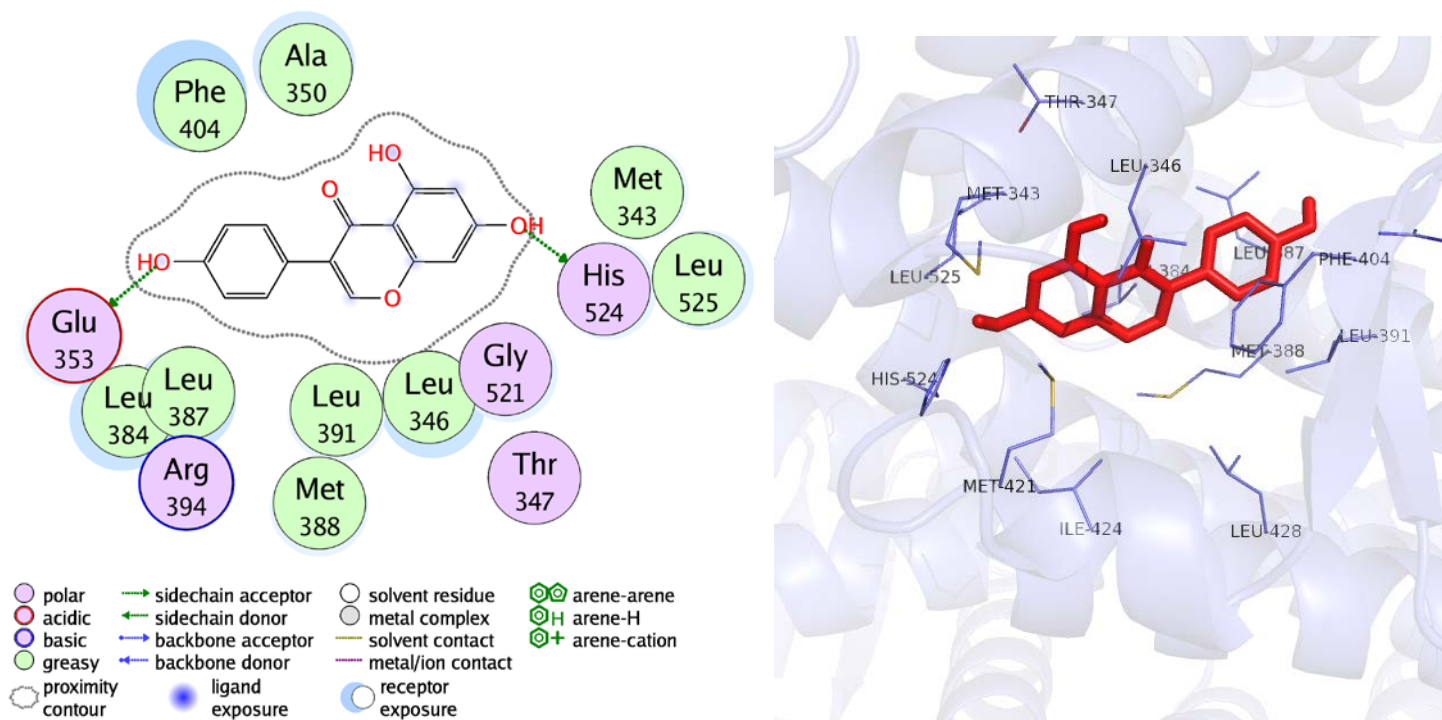
Genistein vs ER-α WILD TYPE

Figure 8. Genistein binding estrogen receptor α, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand. Model template based on agonist conformation and the red ligand is pose produced by Autodock Vina lowest binding energy (-8.3 Kcal/mol).
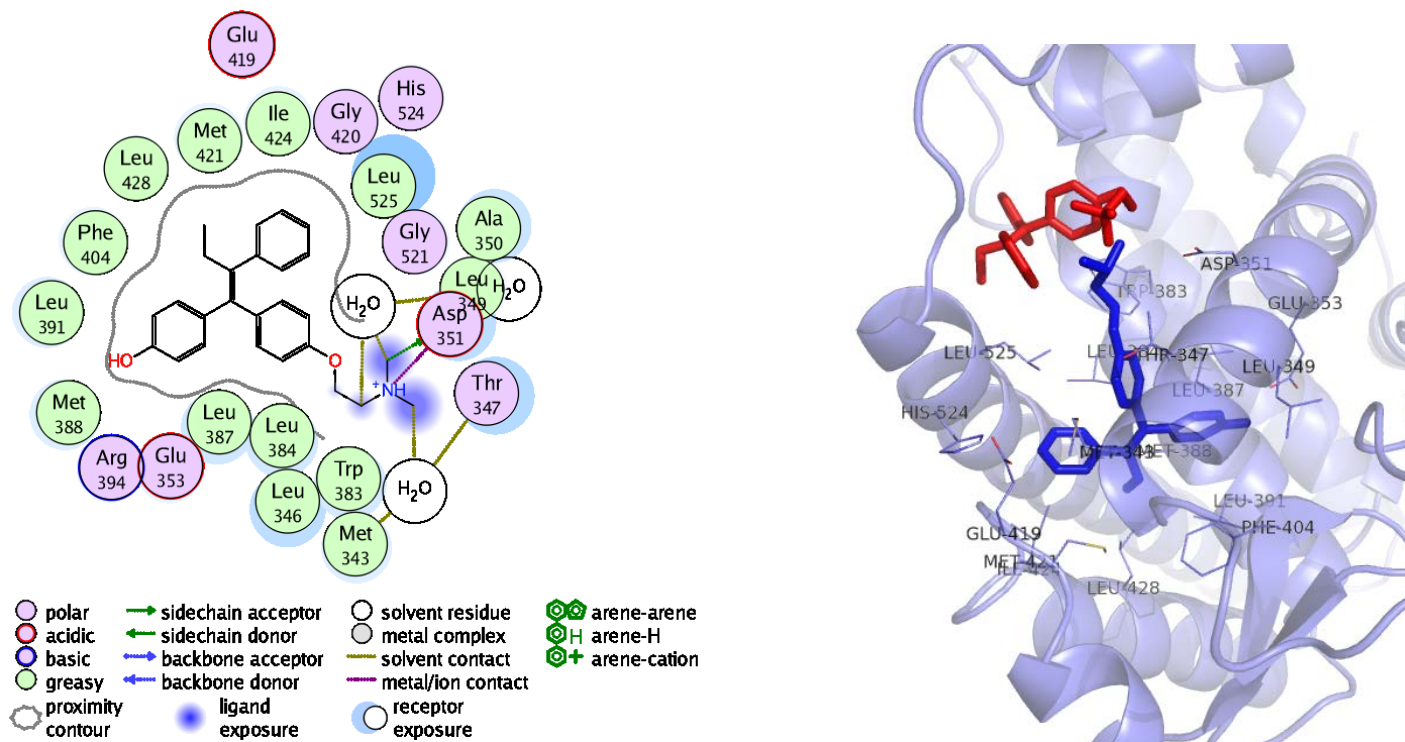
ERα Antagonist template model vs Tamoxifen

Figure 9. Tamoxifen binding estrogen receptor α, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand. Model template based on agonist conformation and the blue ligand is pose produced by Autodock Vina lowest binding energy (-6.3 Kcal/mol).

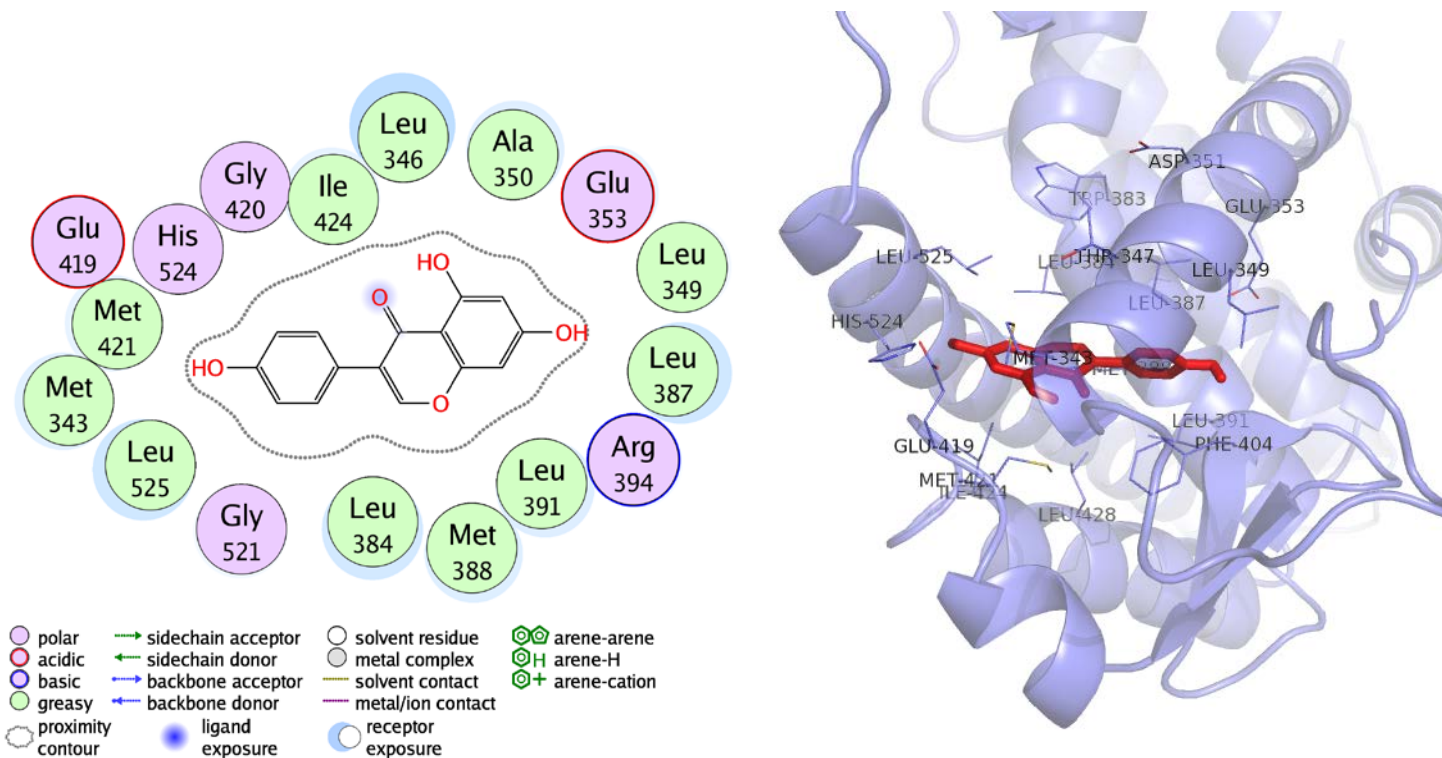**ERα antagonist template model vs genistein**

Figure 10. Genistein binding estrogen receptor α, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand. Model template based on agonist conformation and the red ligand is pose produced by Autodock Vina lowest binding energy (-8.9 Kcal/mol)

2.2 ER-β Wild type

Estradiol PDB file vs ERβ wild type

Figure 11. Estradiol binding estrogen receptor β, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand. Model template based on agonist conformation and the red ligand is pose produced by Autodock Vina lowest binding energy (-10.7kcal/mol).
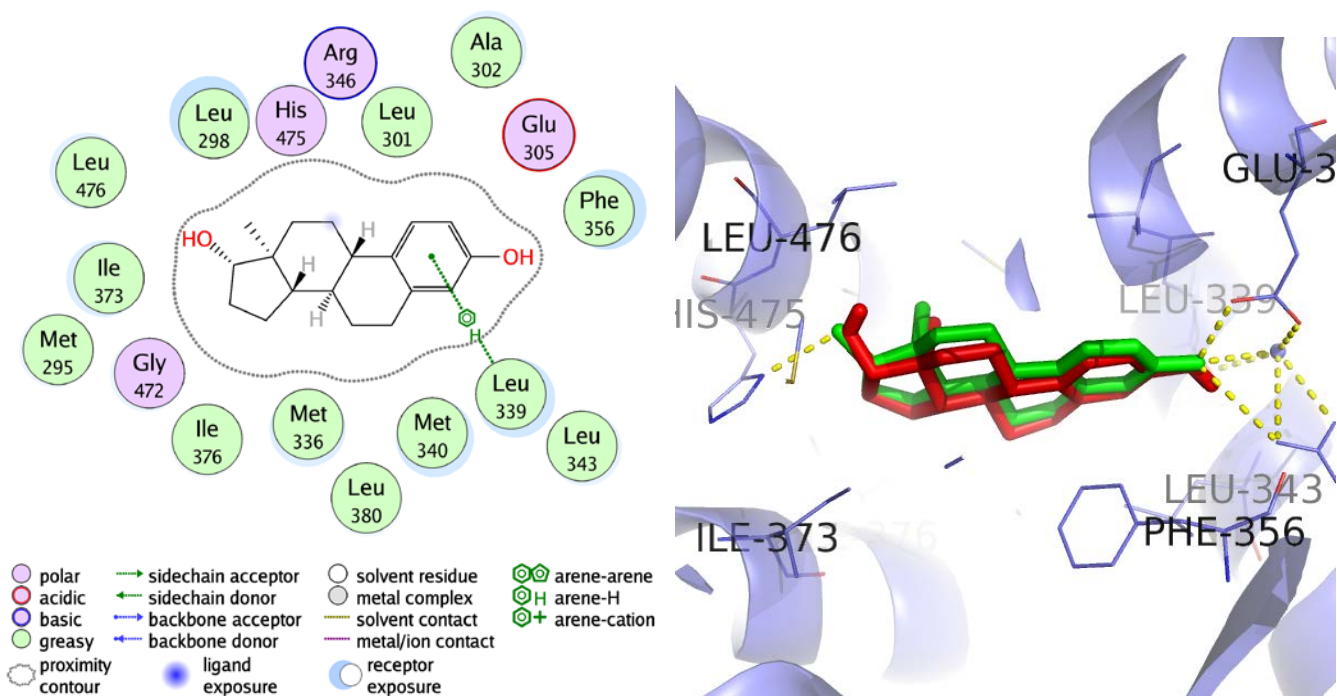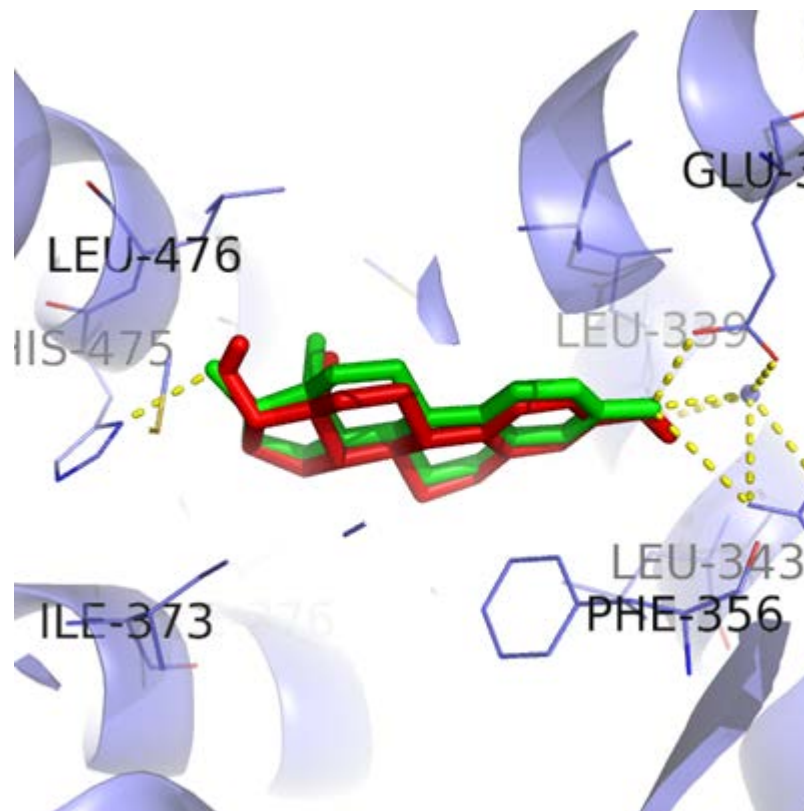
Estradiol vs ERβ wild type

Figure 12. Estradiol binding estrogen receptor β, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand. Model template based on agonist conformation and the red ligand is pose produced by Autodock Vina lowest binding energy (-10.7kcal/mol).

Figure 13. Genistein binding estrogen receptor β, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand. Model template based on agonist conformation and the red ligand is pose produced by Autodock Vina lowest binding energy (-8.3 kcal/mol).

Daidzein vs ERβ wild type

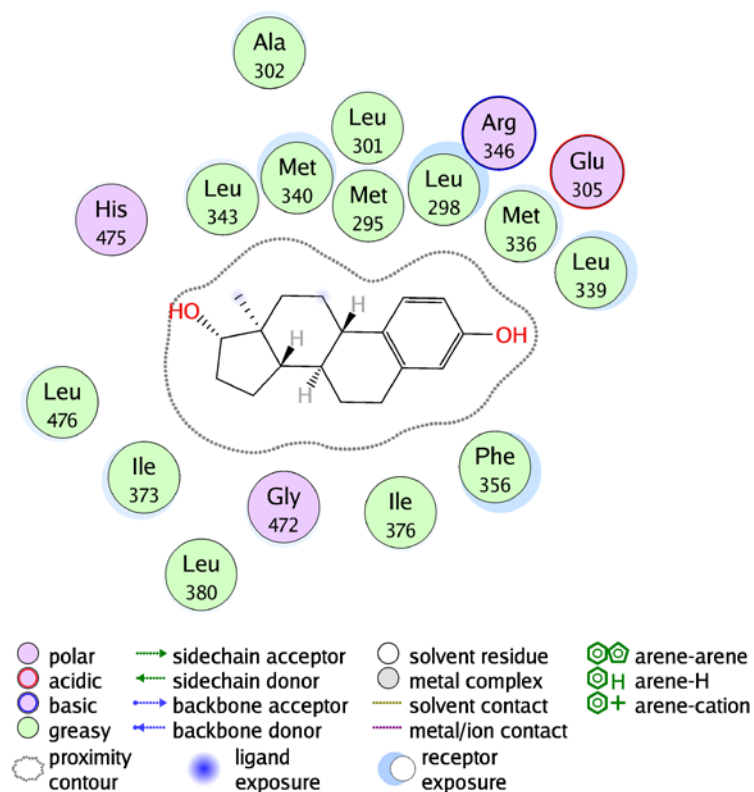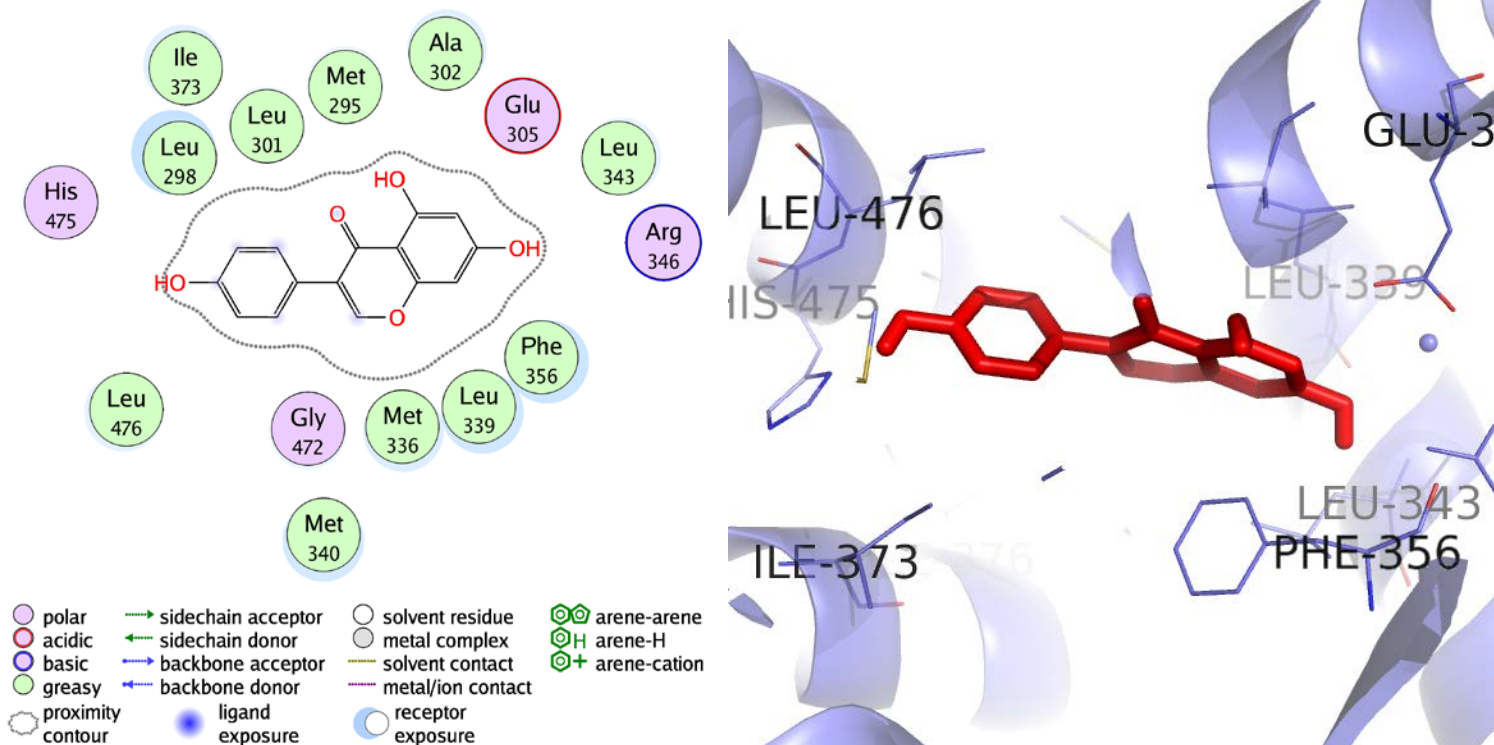

Figure 14. Daidzein binding estrogen receptor β, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand. Model template based on agonist conformation and the red ligand is pose produced by Autodock Vina lowest binding energy (-8.4 kcal/mol) .

Equol vs ERβ wild type

Figure 15. Equol binding estrogen receptor β, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand. Model template based on agonist conformation and the red ligand is pose produced by Autock Vina lowest binding energy (-8.5 kcal/mol).
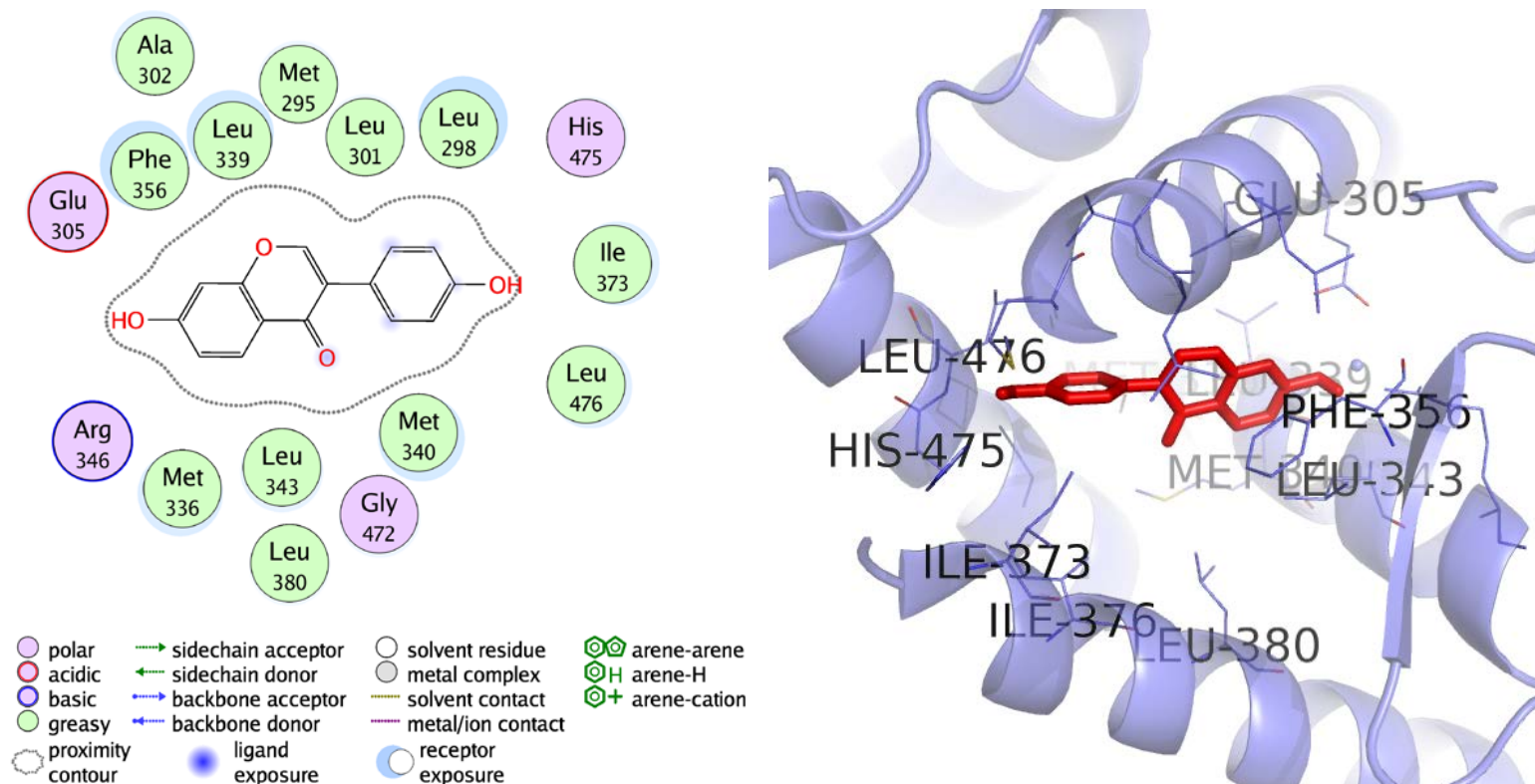
7.

ODMA vs ERβ wild type

Figure 16. ODMA binding estrogen receptor β, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand. Model template based on agonist conformation and the red ligand is pose produced by Autodock Vina lowest binding energy (-8.5 kcal/mol).

Enterolactone vs ERβ wild type

Figure 17. Enterolactone binding estrogen receptor β, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand. Model template based on agonist conformation and the red ligand is pose produced by Autodock Vina lowest binding energy (-8.6 kcal/mol).
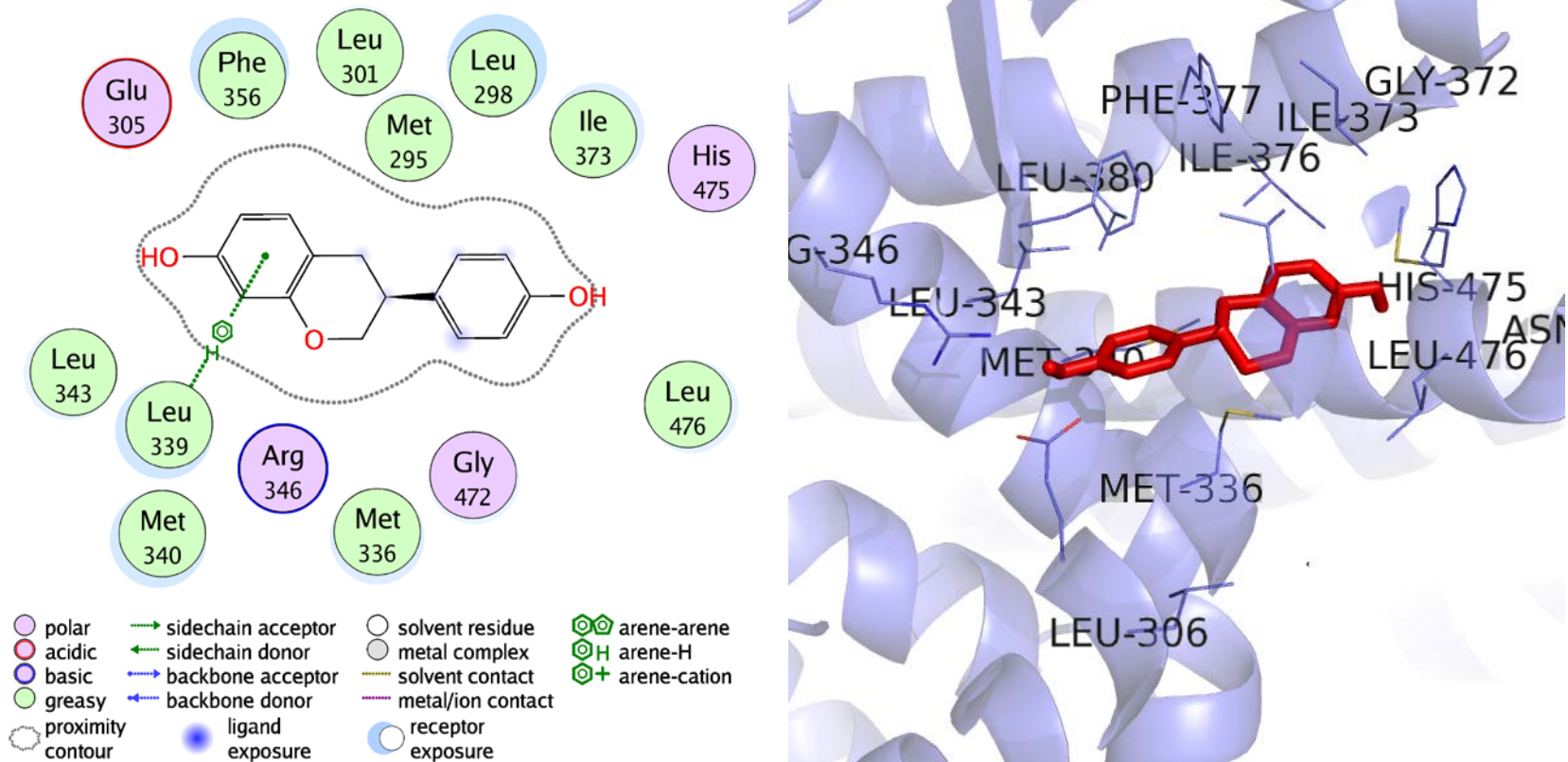
Enterolactone vs ERβ isoform 5

Figure 18. Enterolactone binding estrogen receptor β, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand. Model template based on agonist conformation and the red ligand is pose produced by Autodock Vina lowest binding energy (-6 kcal/mol).
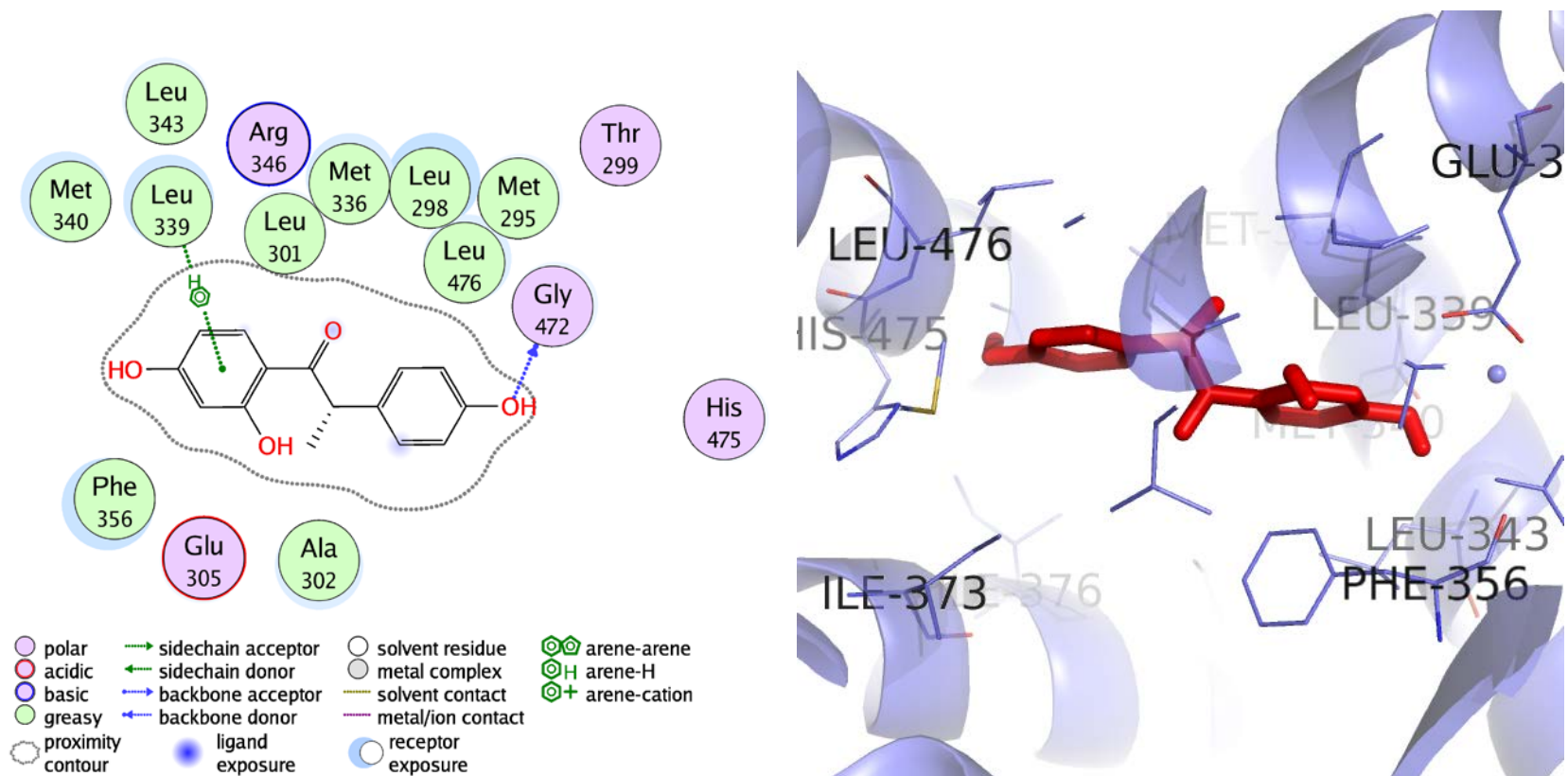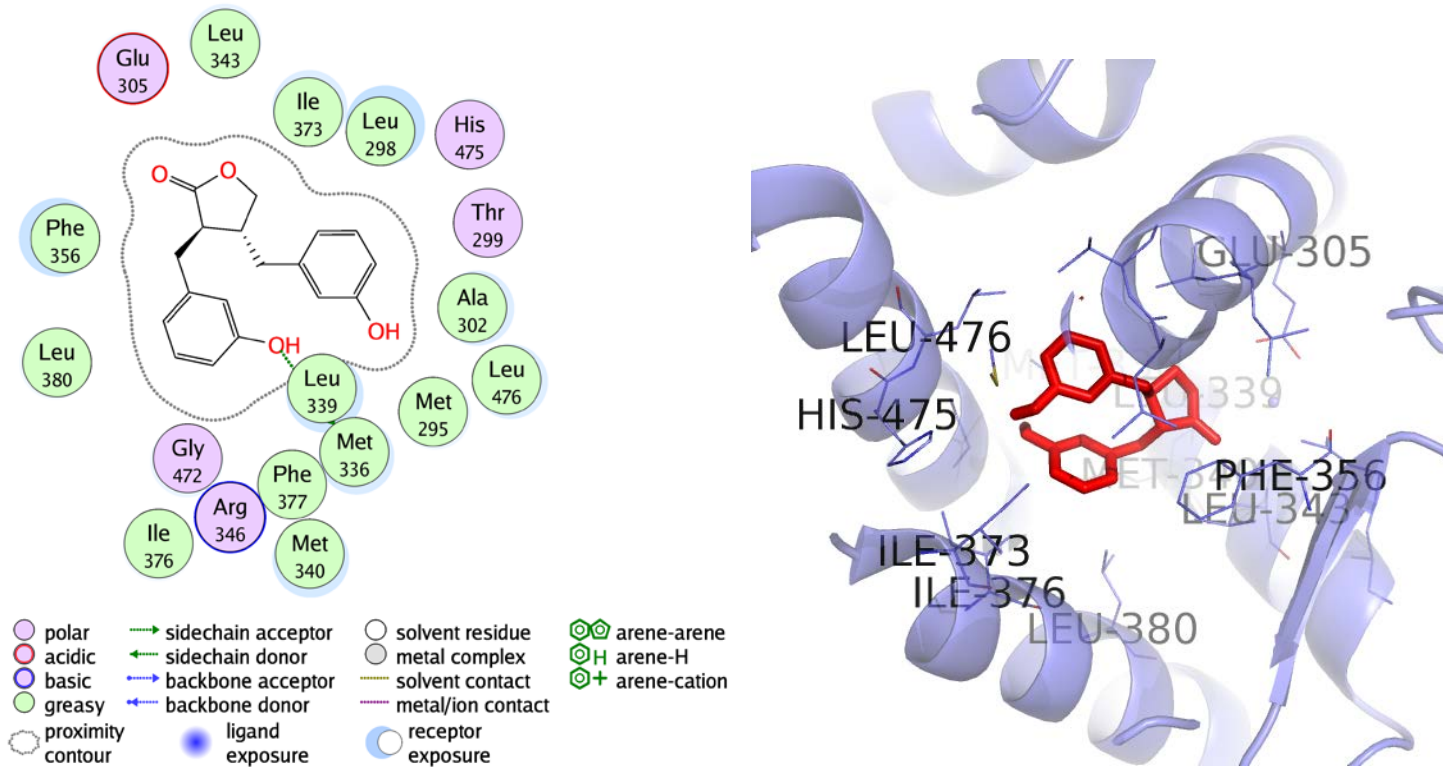
Enterolactone vs ERβ isoform 8

Figure 19. Enterolactone binding estrogen receptor β, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand. Model template based on agonist conformation and the red ligand is pose produced by Autodock Vina lowest binding energy (-7.6 kcal/mol).

ERβ Antagonist template model vs Tamoxifen

Figure 20. Tamoxifen binding estrogen receptor β, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand. Model template based on agonist conformation and the red ligand is pose produced by Autodock Vina lowest binding energy (-7kcal/mol).
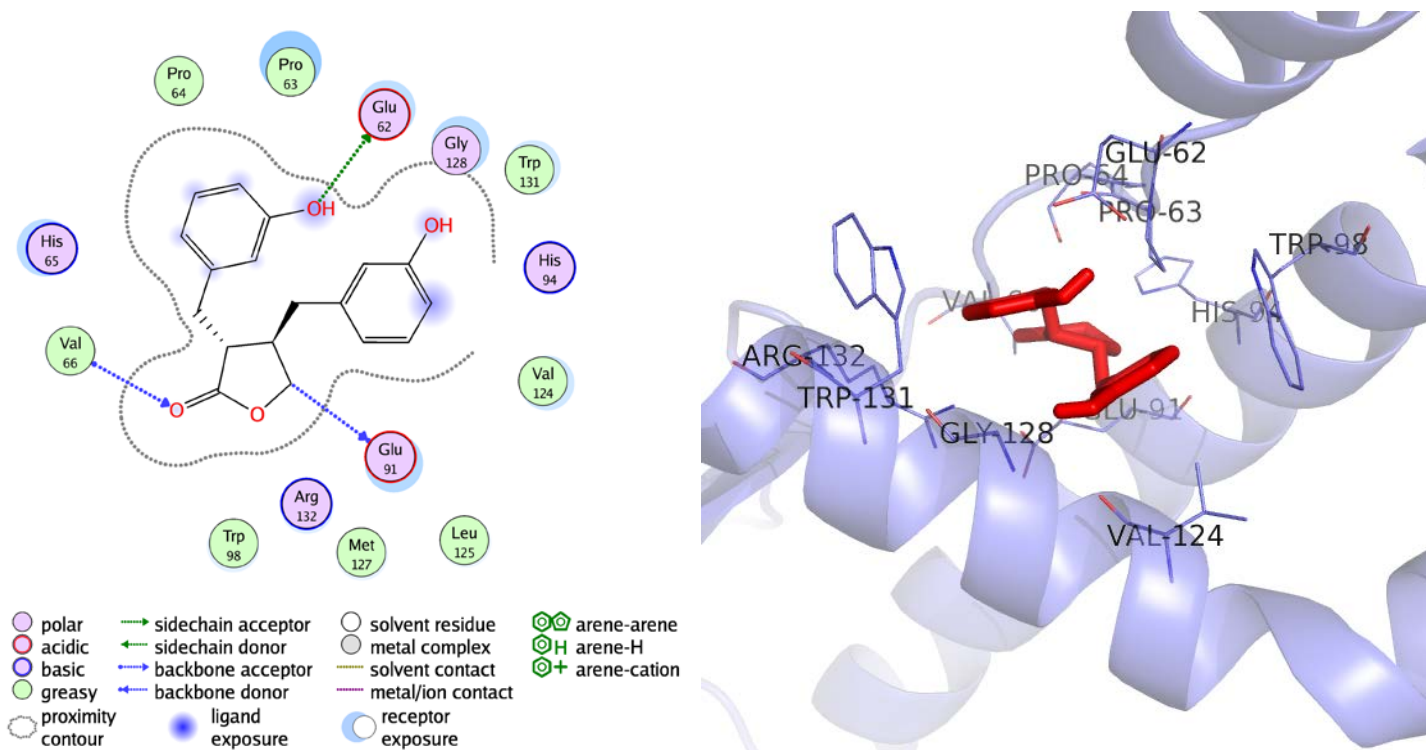
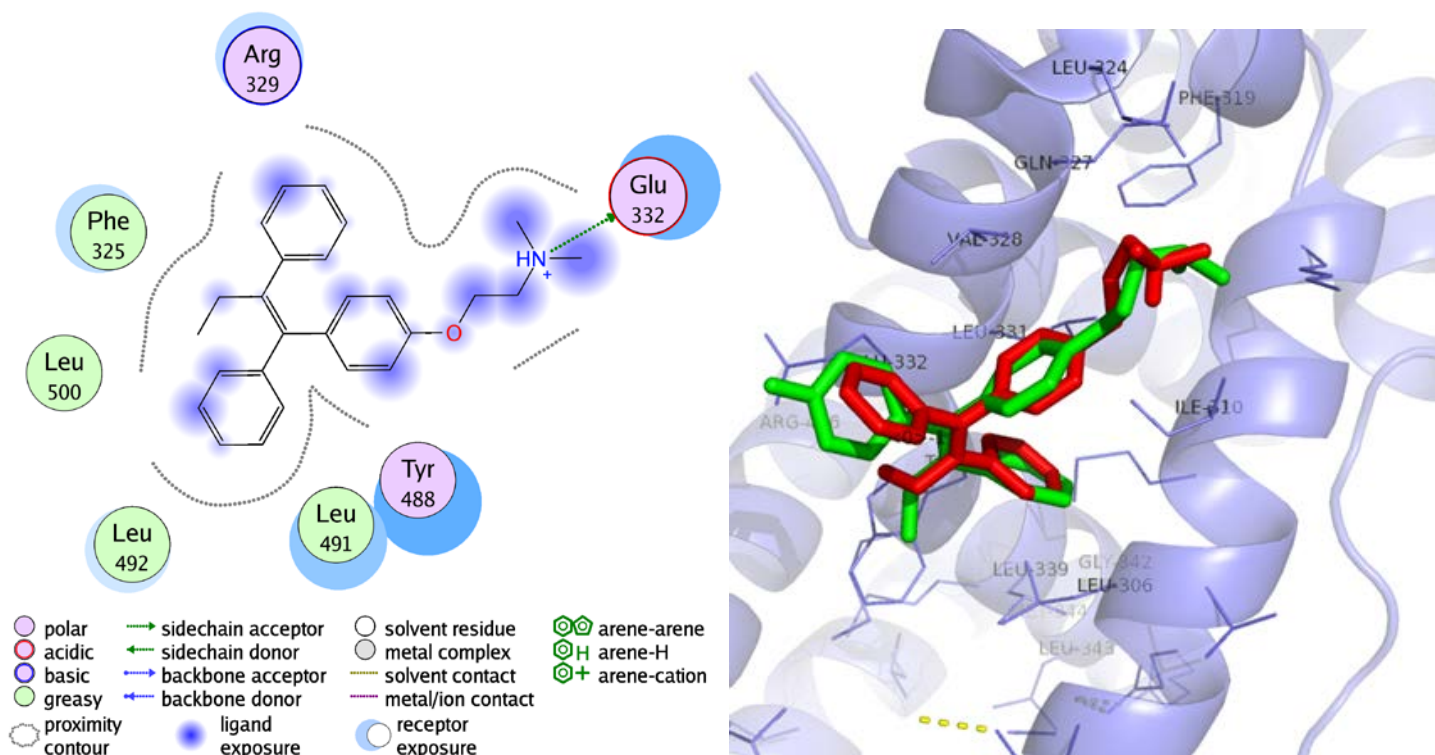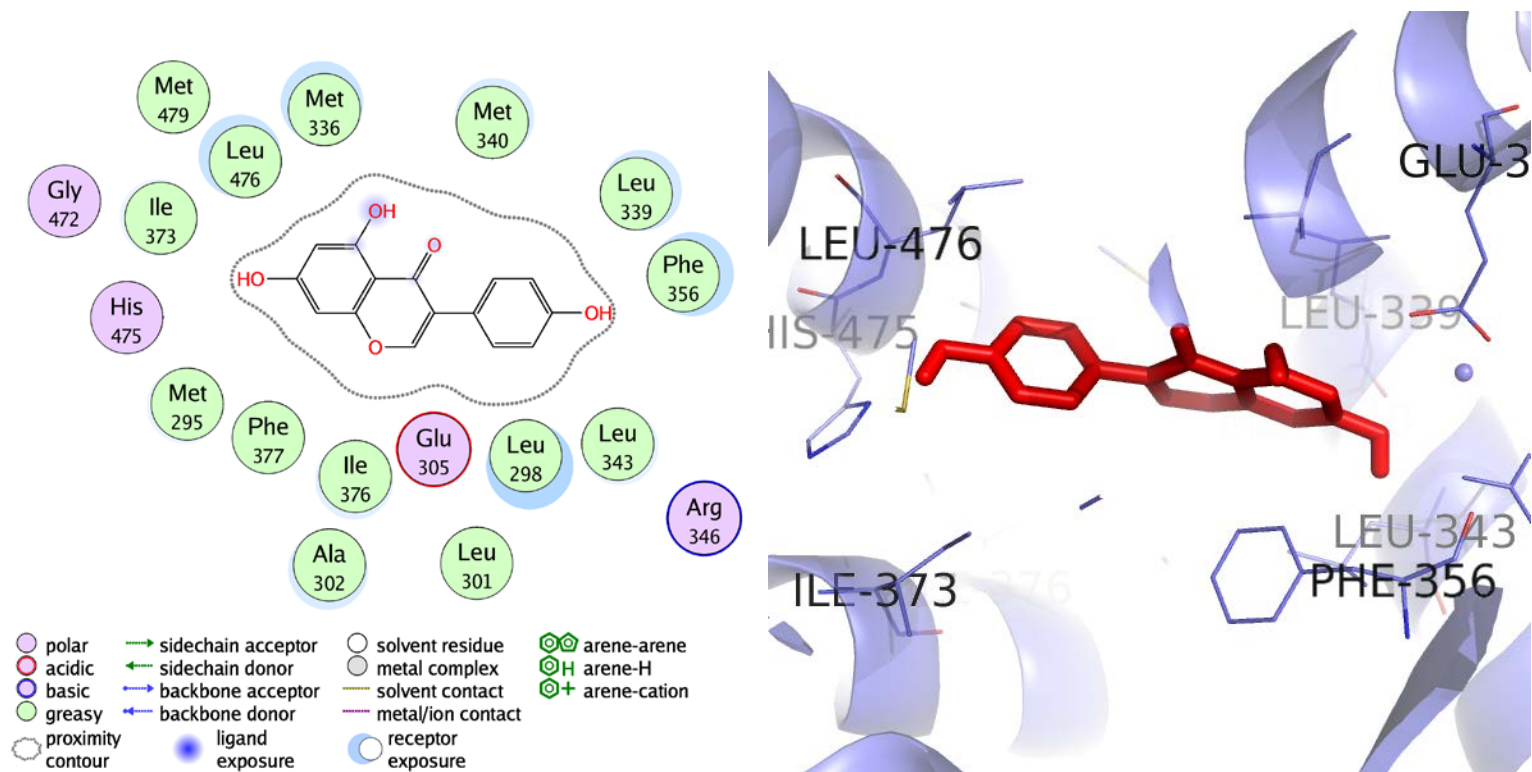**ERβ Antagonism model vs Genistein**

Figure 21. Genistein binding estrogen receptor β, on the right cartoon representation of the amino acid interactions. On the left 2D representation of binding pocket and ligand. Model template based on agonist conformation and the red ligand is pose produced by Autodock Vina lowest binding energy (-8.9 kcal/mol).

Appendix 3

Appendix 3
Validation

3.1 Methodology

3.1.1 Estrogen receptor cloning and expression in *E. coli*

3.1.1.1   Estrogen receptor full length sequence and expression

Wild type estrogen receptor protein sequence was retrieved from Uniprot database (http://www.uniprot.org). Due to the chosen expression system, ERα cDNA sequence was optimized for *E. coli* expression in order to avoid rare codons. WT ERα optimized sequence was verified and synthesized by GeneArt synthesis service at life technologies.

3.1.1.2   Insertion of WT ERα in pET-SUMO vector

pET-SUMO vector was used as expression system, allowing the direct insertion of a PCR product; primers for the insert were according to the instructions in the manual. Amplification of the WT ERα DNA was produced using PCR Supermaster mix (Invitrogen); 45 µl PCR SuperMix, 4 µl primer mix and 1µl of template DNA solution (pMX plasmid kanamycin resistant+ WT ERα supplied by Invitrogen). Reaction was mixed and loaded in the BioRad DNA Engine Thermocycler; PCR conditions were optimized for 35 cycles (94°C x 30", 60°C x 30" and 72°C x 1'). PCR products were resolved in 2% (w/v) agarose gel in Tris-borate EDTA (TBE) buffer.

3.1.1.3   Ligation

After verification of a single right size product (approximately 1865 bp), PCR product purification was performed using QIAquick PCR purification kit (Qiagen, UK) according to the manufacturer's instructions. For optimal ligation efficiency a fresh PCR product was used (less than 16 hr). The ligation reaction was prepared with 1 µl PCR product, 1 µl 10X ligation buffer, 2 µl pET-SUMO plasmid, 5 µl sterile water, 1µl T4 DNA ligase and incubated at 15 °C overnight and stored at -20 °C.

3.1.1.4   Transformation, colony selection and insert analysis

After ligation of the insert (WT ERα) into pET-SUMO, the vector was used for transformation into competent cells (One shot Mach1-T1 supplied by Invitrogen). 2 µl of the ligation reaction was pipetted into a vial of One Shot competent cells, without mixing or pipetting up or down. The respective vial was incubated on ice for 30'. After incubation cells were heat shocked for 30" at 42°C without shaking. The vial of competent cells was incubated in ice for other 15' and at room temperature pre-warm 37°C SOC media was added, mixture was shaked at 37°C at 250 rpm for 1 hr. Afterwards 100 µl of the mixture were spread on selective agar plates (LB agar, 50 µg/ml kanamycin) and incubated overnight at 37°C. After overnight incubation, one colony was used to inoculate 10 ml of LB (Kanamycin 50 µg/ml) for plasmid extraction and insert analysis. QIAgen Spin Miniprep kit was used for plasmid isolation, according to the manufacturer's instructions. Plasmid size, orientation and sequence were analysed by PCR and

sequencing (GATC Biotech Sanger sequencing), Insert orientation was evaluated using primers previously designed in the study for the insert (ESR1 forward and ESR1 reverse) and primer designed for pET-SUMO vector (pET-SUMO forward and T7 reverse) orientation test was performed in a PCR reaction using a mix primer solution (pET-SUMO forward and ESR1).

3.1.1.5   BL-21 transformation and pilot expression

pET-SUMO+ESR1 plasmid was used for pilot expression in *E. coli* BL21(DE3) supplied by Novagen, competent cells were thawed in ice and 1 µl of plasmid (pET-SUMO+ESR) was added and incubated for 30' in ice. Afterwards competent cells were heat shocked at 42 °C for 30'', 250 µl of SOC medium was added to the vial, the mixture vial was incubated at 37 °C for 1 hr and shaken at 250 rpm. The entire transformation reaction was added to 10 ml of LB with kanamycin (50 µg/ml) and incubated overnight at 37 °C with shaking at 250 rpm to proceed to the pilot expression. 10 ml of LB kanamycin 50 µg/ml was inoculated with 500 µl of the overnight culture; broth was incubated at 37 °C in at shaking incubator for 2 hr, and after reaching the mid-log phase of the culture (0.6 $OD_{600 \, nm}$) split into two 5 ml cultures. One of the vials was induced with 1 IPTG mM, 500 µl was removed from each vial (Induced and uninduced) and centrifuged for 30' at maximum speed in a micro-centrifuge, the supernatant was discarded and the cell pellet was stored a -20 °C then labelled as a zero point. Both cultures were incubated at 37 °C with shaking and an aliquot of 500 µl of time points was taken every hour for 6 hours, each time point was processed as described above.

163

3.1.1.6   Analysing samples and SDS-PAGE

Samples were analysed by SDS-PAGE on10–4 % acrylamide gels, samples were thawed and mixed with 80 µl of Laemmli sample buffer. Mixtures were boiled for 5' and briefly spun in a table micro centrifuge at maximum speed. An electrophoresis chamber (mini-PROTEAN Tetra Cell Bio Rad) was previously prepared with running buffer (Appendix 4, page 1) and two gels, 5 µl of the induced and uninduced samples were loaded in the pre-cast gels along with a protein marker (SeeBlue Pre-Stained Standard) at both gels as a duplicate. Gels were run for 1 hr at 110 V, then, gels were washed twice for 5' with ddH$_2$O. Gel was stained Coomassie Brilliant blue stain overnight at room temperature. Staining was removed by destain solution for about 4 hr.

3.1.2.1   Optimization of the expression levels

In order to increase the yield of the soluble protein and overcome proteolytic cleavage, instability and precipitation, the protein expression protocol was modified in different steps ranging from the composition of the medium to the cell lysis conditions. IPTG induction was tested under different concentrations (0.2, 0.5,1 and 1.5 mM IPTG), LB broth was supplemented with estradiol and sucrose, incubation temperatures were tested at 15, 18, 25 and 37 °C and the time for induction was varied from 18, 12, 4 and 3 hr respectively and, finally, the use of a cocktail of protease inhibitors (Sigma Aldrich; 1:100) in the cell lysis buffer (Bacterial PE LB G-Bioscience) was evaluated. All the conditions were

tested separately and finally combined. Samples were assessed by SDS-PAGE gels and Western blotting as before.

3.1.2.2 *Western blot*

To detect the expression levels of the ER by Western blotting, an anti-HisG antibody (Invitrogen) was used. Blotting was performed as shown in (Appendix 3, figure 4) with 100 Volts for 2 hours. Afterwards, membrane was blocked using 5 % w/v BSA in TBS overnight at 4ºC, blocking buffer was removed and the membrane was submerged with 5 ml of the diluted 1:2000 AP-conjugated antibody (Anti-HisG Invitrogen) into TBS for 2 hours at room temperature. NBT and BCIP substrates were used for detection of the alkaline phosphate probe; 33 µl NBT and 16.5 µl BCIP were added to 5ml of alkaline phosphatase buffer (Appendix 4). After washing the membrane in TBST twice for 5 minutes with gentle agitation, the membrane was rinsed and placed in a dish with 5 ml of the BCIP/NBT substrate solution for about 5 minutes until the colour developed. The reaction was stopped by washing the membrane in ddH$_2$O, membrane were air-dried and photographed.

3.1.3   Estrogen receptor LBD expression

According to the Uniprot database the estrogen receptor ligand binding domain (LBD) falls between aa 302 to the aa 549 of the estrogen receptor sequence (P03372 - ESR1_HUMAN). The cloning segment was based on literature reports according to its solubility and expression yield [332 aa to 552 aa] (Eiler *et al.* 2001). The optimized WT ESRα described above was use to extract the right insert optimised for *E. coli* expression. Ligation, transformation, colony

selection and expression were performed according to the instructions in the pET-SUMO manufacturer's manual and as described above (Sections 3.1.1.2/3/4/5/6). 500 µl of overnight culture of the transformation vial BL21 (pET-SUMO + ESR1/LBD) was inoculated in 10 ml of LB 50 µg/ml kanamycin at 37 °C for approximately 2 hour or until it reached the mid-log phase (0.4-$0.6_{OD600 nm}$), 500 µl of broth was added to 50 % v/v glycerol in 2 ml cryovial tube and mixed gently. Bacterial glycerol stocks were frozen at -80 °C.

3.1.3.1   *Scaling-up expression and purification*

Expression was scaled up to 500 ml bacterial culture for protein purification using HisPur Ni-NTA resin. The glycerol stock of the transformed BL21 (pET-SUMO + ESR1/LBD) cells were used. The stock was inoculated into LB containing kanamycin at 50 µg/ml and grown overnight at 37 °C. 5 ml of the overnight broth culture were used to inoculate two 2 l flasks each containing 250 ml of LB with kanamycin 50 µg/ml. The cultures were incubated for 2.5 hours at 37 °C with shaking at 259 RPM until mid-log was reached (0.6_{OD600}), flasks were cooled down in ice-cold water for 30 min. IPTG was added to final concentration of 1mM in both cultures and incubated at 25 °C for 4hr. Cells were collected by centrifugation (3000 x g for 10' at 4°C), supernatant was discarded and pellet suspended in 5 volumes of the bacterial PELB buffer, lysozyme were added to a final concentration of 100 µg/ml. The suspension was incubated for 1 hr on ice and sonicated in a beaker of ice at 40%, 5x10" bursts with intermittent cooling. Debris were removed by centrifugation at 11,000 x g, 30' 4°C. The supernatant was removed and prepared by overnight dialysis at 4°C in 1l of binding buffer. A gravity-flow column was prepared with 3 ml of Ni-NTA resin;

166

the column was equilibrated with 9 ml binding buffer; the prepared protein extract was added. 12 ml of wash buffer were added to the column and an eluted with 10 ml elution buffer, Amicon ultra-15 centrifugal filters (MWCO 10 kDa) were used for protein concentration for a final volume of 1 ml. The final protein concentration were monitored by measuring the absorbance at 280 nm. Finally, the protein was desalted using a PD-10 column (GE Healthcare) previously equilibrated with storage buffer.

3.1.4 Competitive ligand binding assay

3.1.4.1 *Fluorescent polarization Assay: Polar-Screen Estrogen receptor (ER) alpha competitor assay*

The PolarScreen™ Estrogen Receptor Alpha (ER Alpha) Competitor Assay, Green (Life technologies) was used to determine IC_{50} values of different endogenous and exogenous Compounds in the full-length ER alpha. The test compounds were 3 endogenous and natural ligands (estradiol, estriol, estrone), a commercial drug Tamoxifen and 5 different PEs (genistein, daidzein, coumestrol, secoisolaricinol and enterodiol). Firstly, it was use to determinate IC_{50} values of the tested compounds and secondly to test the LBD-ER alpha cloned in the previous experiment using the Fluormone™ ES2 Green as selective fluorescent ligand. The binding affinity of the tested compounds were measure in a range of 12 concentrations (serial dilutions), with a final concentration of 41nM of full-length ER-Alpha and 4.5 nm Fluormone ES2.

Component	Assay	Controls		
	Test compound	No receptor Control (Free fluormone)	Maximum mP Control	Minimum mP Control (Displaced Tracer)
2XSaturating Estradiol (20 µM)	-	-	-	10 µL
2XTest Compound (titrations)	10 µL	-	-	-
2XERAlpha/Fluormone ES2 Green Complex	10 µL	-	10 µL	10 µL
2X Fluormone™ ES2 Green	-	10 µL	-	-
ES2 Screening Buffer	-	10 µL	10 µL	-

Table 1. Reagents volumes, reagents amounts required to perform the competitive assay and controls in a final volume of 20 µL

Three different plates were use in the experiment the first two plates were for serial and intermediate dilutions (384-deep well, uncoated, polypropylene plates that can accommodate 240 µL per well; plate must be tolerant of 100% DMSO, Corning, Cat. no. 3363) and an assay plates (low-volume 384-well plates with non-binding surface Corning, Cat. no. 3820). Final volume of the assay per well was 20 µl and DMSO were used as a solvent. 10 µl of the NR buffer of the NR buffer was aliquoted to each well in column 1 and 2 of the final assay plate. Ligand stock were prepared at 1mM (100x), and the respective serial dilution were done in the dilution plate adding 20 µl per well from B to P rows, 30 µl of the 1mM ligand stock in A row were added per compound. 10 µl were transferred from well A1 to B1 then from B1 to C1 and this process was repeated until reached the last row (P). Afterwards, it was added 98 µl of NR buffer in the required number of columns in the intermediate dilution plates to dilute each compound from 100x to 2x after adding 2 µl from each well in the dilution plate to the respective ID well in the intermediate dilution plate. Finally, 10 µl of the 2x intermediate dilution were added of each compound by duplicate in the final assay plate. All the reagents were added according to the table 1 in the respective columns and wells, the assay plate was mixed and cover from light. After 2 hours incubation at 25 ^{0}C the Fluorescent polarization was measure in a Perkin-Elmer plate reader (EnVision 2104 Multilabel reader, software; Wallace manage version 1.12) using filter 535 nm for excitation and 590 nm for emission. Binding parameters were calculated by non-linear regression using GraphPad Prism 7 software, plotted ant fitted with a one site competition curve.

169

3.1.4.2 Ligand binding domain screening

Assessment of the LBD domain ER alpha were performed using the PolarScreen™ Estrogen Receptor Alpha (ER Alpha) Competitor Assay, Green kit. By replacing the full-length provided ER-alpha by the in house produced LBD-ER alpha and tested for estradiol binding affinity in order to find the optimal concentration to provide at least a 150 mP Delta mP (Difference between the highest concentration and the lowest).

3.2 Result

3.2.1 Estrogen receptor cloning and expression in E. coli

The full length estrogen receptor sequence (Appendix 1 table 1), reverse transcribed and optimised to take into account the rare codon usage were prepared and delivered by Invitrogen in a kanamycin resistant plasmid to facilitate selection in *E. coli*. Primers for amplification of the insert were designed according to pET-SUMO manual instructions. A pcr product was generated, purificated and checked for size (1785 bp) by electrophoresis. The 1785 bp purified product was used as an insert in the pET-SUMO linear plasmid and after ligation transformation was performed in presence of 50 µmg/ml of kanamycin agar plate. 10 colonies were chosen after overnight incubation and after plasmid extraction and purification a PCR were run using pET-SUMO primers, ESR1 primers and a mix between both in order to check the right orientation of the insert. The plasmid of the correct orientation was sent for sequencing and no mutations were observed. This was then used for transformation and expression experiments in BL21 (DE3) cells, after pilot expression and analysis using SDS-PAGE poor expression levels were observed and proteolysis apparent. In order to improve the level of expression, the solubility and to overcome issues with proteolysis, different conditions were tested as described in the materials and methods, these ranged from varying the incubation temperature, IPTG concentration, supplements and lysis buffers and inhibitors. However, protein precipitation and low expression were still observed.
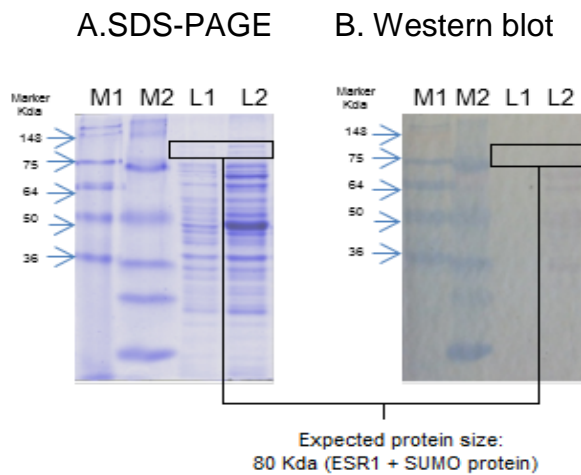
A.SDS-PAGE   B. Western blot

Expected protein size:
80 Kda (ESR1 + SUMO protein)

Figure 1. Expression of full length ESR1 WT with SUMO tag in BL21(DE3). Protein was expressed in native state in BL21(DE3), Experimental conditions: Cells were induced with 1mM IPTG and incubated at 25 ºC for 4 hr with shaking at 250 RPM. Lane M1 and M2: pre-stained protein markers (M1 SeeBlue Plus2 Invitrogen, M2 Precision plus protein Bio-Rad). L1: un-induced fraction BL21 cells lysis, L2: Induced whole cell lysed fraction. On the right A; Samples were separated using SDS-PAGE (10% gel) with Coomassie Brilliant blue staining and analysed (on the left) B; Western blot with 6HIS antibody, binding of antibody was displayed by colorimetric detention (BCIP/NBT).

3.2.2 Estrogen receptor LBD expression

After unsuccessful attempts and evaluation of various approaches to express the full length ER WT protein, it was necessary to focus in the LBD which has already been expressed under different conditions and with bacterial expression systems, and in order to improve the recombinant protein solubility the pET-SUMO plasmid system was employed as well. After plasmid construction (pET-SUMO LBD) plasmid size, orientation and sequence were assessed, these were found to display the correct sequence. Conditions used for the expression as per the flow chart after SDS-PAGE analysis, the best conditions observed were incubation for 4 hr at 25 ºC with shaking at 250 RPM, in LB broth (figure 6).
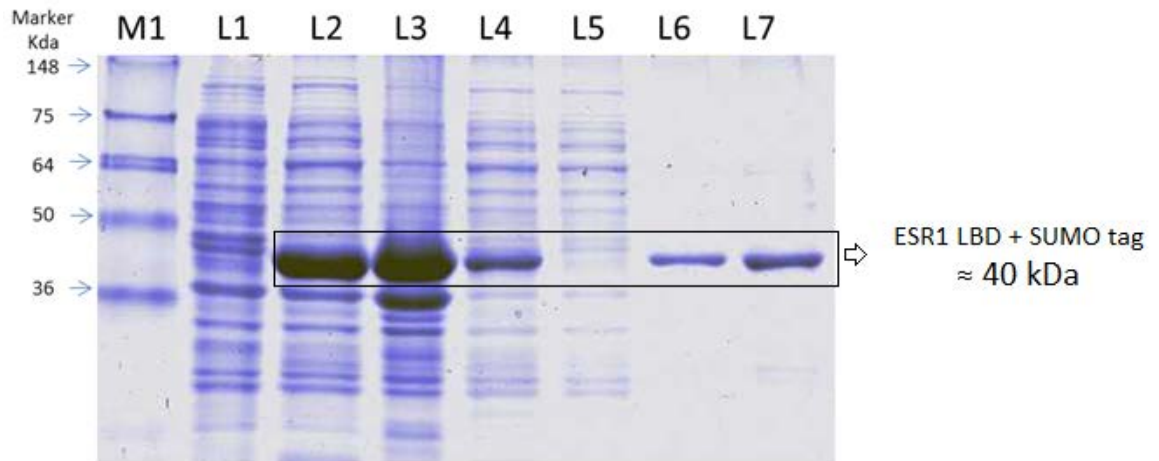
Figure 2. **Expression Estrogen receptor ligand binding domain with SUMO tag in BL21 (DE3) and purification of His-tagged protein**. Protein was expressed in native state in BL21(DE3), Experimental conditions: Cells were induced with 1mM IPTG and incubated at 25 ⁰C for 4 hr with shaking at 250 RPM. SDS-PAGE analysis of Coomassie Brilliant blue stained gel for visualisation. Lane M1: pre-stainedprotein markers (SeeBlue Plus2 Invitrogen). L1: un-induced fraction BL21 cells lysis, L2: Induced whole cell lysed fraction, L3: in-soluble fraction, L4: soluble fraction. Soluble fraction was used for affinity purification of the His-tagged protein; L5: flow-through fraction, L6: eluted fraction. All the eluted fractions were pooled and concentrated using an amicon Ultra-15 centrifugal filter, L7: concentrated recombinant purified protein (ESR1 LBD +SUMO).

During the purification steps the protein displayed less precipitation and when expression was performed at 4⁰C from 500 ml of LB broth 8 mg/ml of soluble protein was purified (Figure 6), The shelf life before precipitation was around 2 weeks at 4⁰C. This protein has now been stored at -20 ⁰C in storage buffer (25 mM NaCl, 20 mM Tris-HCl pH 8.0, 10% glycerol and 5 mM DTT) awaiting further experiments.

3.2.3　Competitive ligand binding assay

3.2.3.1　Fluorescent polarization Assay: Polar-Screen Estrogen receptor (ER) alpha competitor assay

Competition binding experiments are used to determine affinities of non-labelled ligands, in which it can be compared the relative binding affinity (IC_{50}) and K_i when the labelled ligands K_d is known. Calculating binding affinities without labelling the studied ligands is one of the benefit of fluorescent polarization (FP). FP and IC_{50} were measured using the Polar-Screen Estrogen receptor (ER) alpha competitor assay, green kit (Invitrogen, UK). Serial dilutions of the interested compounds were done ranging from 10 µM to 56 pM; FP was measured using an Perkin Elmer plate reader (EnVision serial number 1040476) using for excitation filter (FITC FP 480), Emission filter (FITC FP P-pol 535) and FITC FP S-pol 535 as a 2nd emission filter. Coordinates of the corners of the 384 plate were set at 12.13 mm x 8.99, 115.63 mm x 8.99 mm, 12.13 mm, 76.49 mm, and 115.63 mm x 76.49 mm. measure height 6.5 mm, excitation light 75%, G factor 1, Detector gain 500 and 700 and a 30 number of flashes. FP value blank corrected were calculated (1000*[(S-Bs)-G*(P-Bp)]/[(S-Bp)+G*(P-Bp)] where S; FP fluorescein channel 2 [535s], P; FP fluorescein channel 1 [535p], G; G-factor, Bs;Average of the blanks of S-channel and Bp; Average of the blank of P-channel.
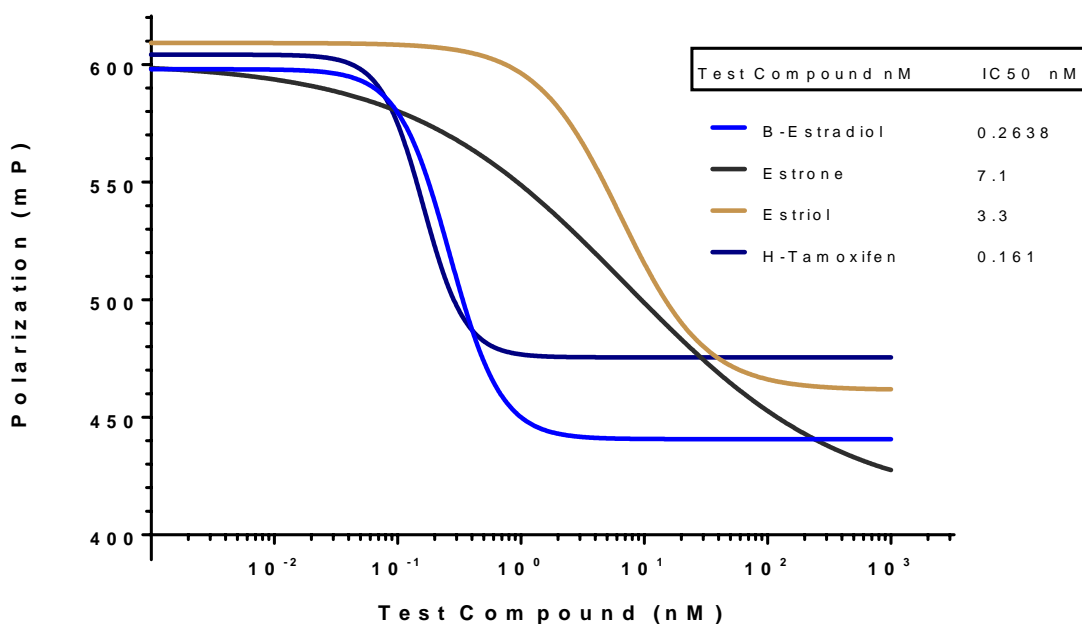
PolarScreen ER Alpha competitor assay

Test Compound nM	IC50 nM
B-Estradiol	0.2638
Estrone	7.1
Estriol	3.3
H-Tamoxifen	0.161

Figure 3. Fluorescent polarization Assay, serial dilutions of 4 well-known ER alpha ligand were done to measure mP using the polar screen kit. Binding parameters were calculated by non-linear regression using GraphPad Prism 7 software, plotted ant fitted with a one-site competition curve. The PolarScreen assay was used to test the ligand binding.

The tested compounds were 8 different PEs (Genistein, coumestrol, daidzein, equol, secoisolarisinol, enterodiol and enterolactone) in which only 4 IC_{50} were calculated by the data provided by the polarization values from the tested concentration (from 10nM to 57pM). Coumestrol displayed the lowest IC_{50} (13.75nM), secoisolarisinol, enterodiol and enterolactone needed highest dilution concentration to calculate IC_{50}.

PolarScreen ER Alpha competitor assay

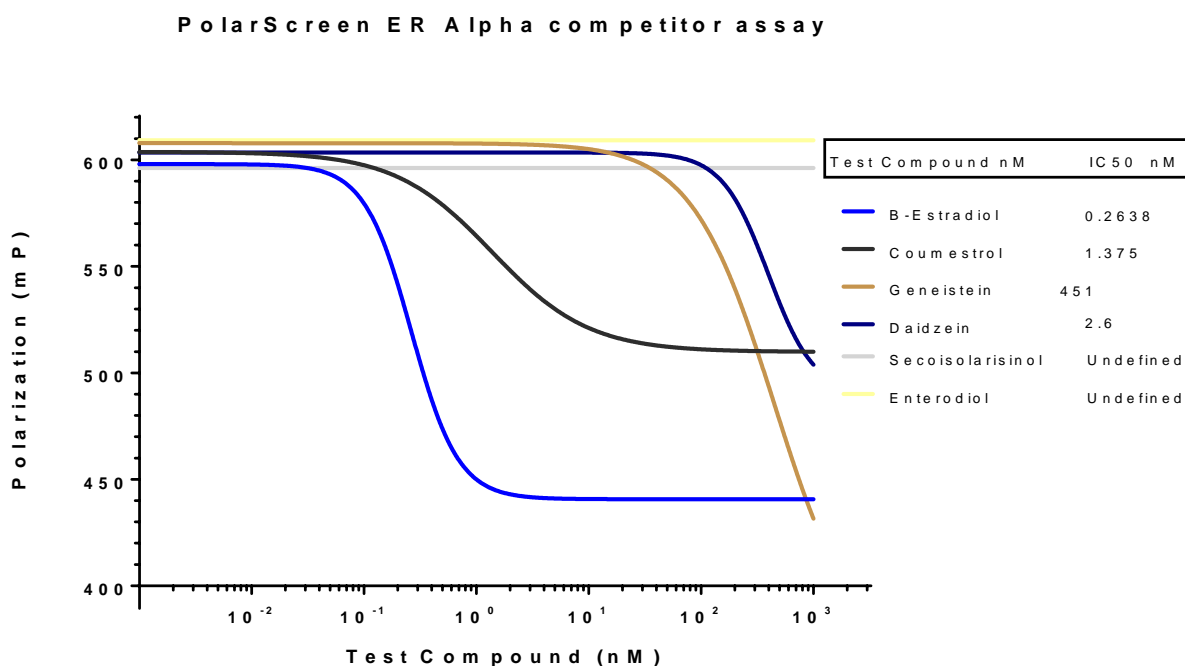| Test Compound nM	IC50 nM
B-Estradiol	0.2638
Coumestrol	1.375
Geneistein	451
Daidzein	2.6
Secoisolarisinol	Undefined
Enterodiol	Undefined

Figure 4. Fluorescent polarization Assay, serial dilutions of 8 well-known ER alpha ligand were done to measure mP using the polar screen kit data from equol (IC_{50} 395nM and enterolactone (No calculated IC_{50}) data were not displayed in this graph and . Binding parameters were calculate by non-linear regression using GraphPad Prism 7 software, plotted ant fitted with a one-site competition curve. The PolarScreen assay was used to test the ligand binding.

ER alpha is formed by several independent domains where the LBD is solely responsible for the ligand binding interactions. Full-length ER alpha was provided with the kit, and in order to prove if the in-house recombinant LBD ER alpha protein was active to interact with natural ER ligands. Estradiol was used as a control in a range of 12 different concentration. However, it was observed a clear interaction and competitive binding response, the delta mP were too low to be used to the respective calculate IC_{50}.
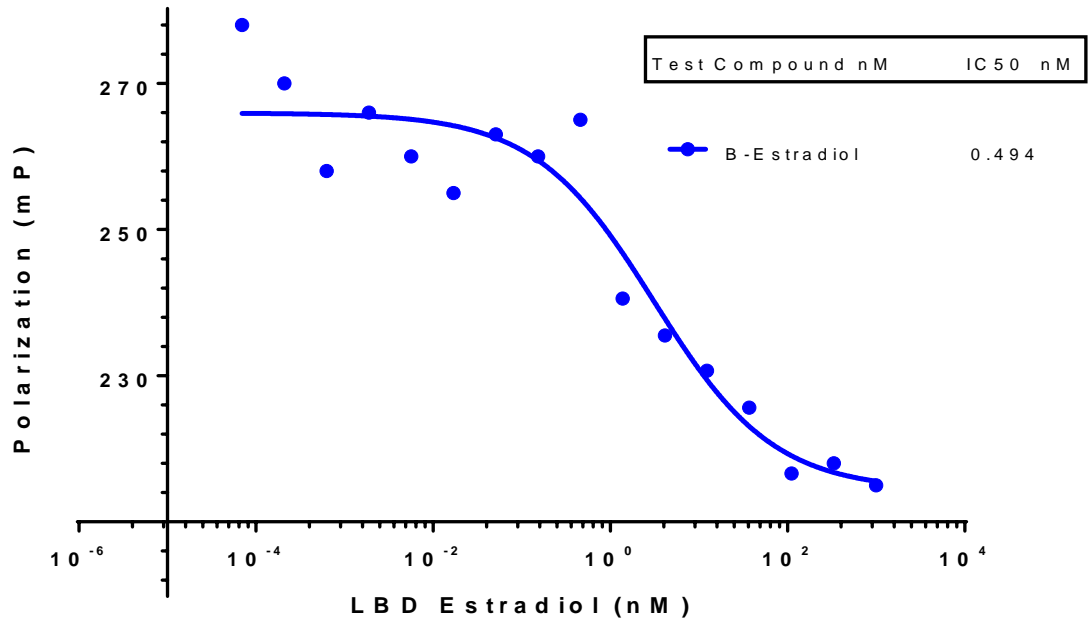
Figure 5. Fluorescent polarization Assay of the human LBD-ER recombinant protein, serial dilutions of the natural ligand 17β-estradiol that have a nanomolar affinity for the wild-type full-length ER alpha. However, the final protein concentration in the assay was no optimal to provide a desirable Delta mP, the purified LBD-ER alpha recombinant protein displayed is active and can be potentially use in future experiment and the LBD-ER beta (data no shown in the report) can be tested and used for downstream experiment.