ENDOCRINE DISRUPTORS AS LIGANDS OF BREAST CANCER PROTEINS

Ph.D. thesis by
Diana Montes Grajales

In partial fulfillment of the requirements for the degree of Doctor of Philosophy in the subject of Environmental Toxicology

Advisor
Prof. Jesús Olivero Verbel. Ph.D.

University of Cartagena
Faculty of Pharmaceutical Sciences
Ph.D. Program in Environmental Toxicology
Cartagena, Colombia
2015

DEDICATION

To the women that suffer from breast cancer caused by environmental factors, who encouraged me to work with tenacity on this project.

CONTENTS

RESUMEN

Disruptores Endocrinos como Ligandos de Proteínas Asociadas a Cáncer de Seno

Los disruptores endocrinos (EDCs) son un grupo estructuralmente diverso de compuestos químicos que tienen la capacidad de afectar el sistema endocrino. Sin embargo, sus efectos no se circunscriben exclusivamente a la desregulación de procesos controlados por hormonas, sino que han sido epidemiológicamente relacionados con una amplia gama de patologías. Entre las que se encuentra, el cáncer de seno, una enfermedad compleja por su elevada heterogeneidad en términos moleculares, y considerada el tipo de cáncer más prevalente en mujeres de acuerdo con la Organización Mundial de la Salud (OMS).

Algunos de estos compuestos han sido considerados tóxicos ubicuos. Por lo cual, existe una necesidad creciente de estudiar sus efectos adversos y mecanismos de acción. Los xenoestrógenos están presentes en numerosos productos de uso diario, tales como agentes limpiadores, maquillajes, perfumes, plásticos, electrodomésticos, computadores, e incluso alimentos. En la naturaleza, se encuentran haciendo parte de ciertas plantas y metabolitos producidos por hongos, así como de ambientes contaminados por la actividad antropogénica, afectando los recursos hídricos, el aire, el suelo y la biota.

Por lo anterior, el propósito de esta tesis ha sido evaluar la capacidad de los EDCs para interaccionar con proteínas asociadas al cáncer de mama. Este proyecto fue desarrollado en tres etapas: la creación de una base de datos de EDCs con estructuras tridimensionales disponibles, un cribado virtual inverso del bisfenol A (BPA) -uno de los xenoestrógenos a los que estamos más frecuentemente expuestos- con proteínas implicadas en diferentes vías de señalización, y un cribado de alto rendimiento (vHTS) entre los EDCs y

proteínas asociadas al cáncer de seno, incluyendo la evaluación *in vitro* de las interacciones de un complejo proteína-ligando a través de métodos espectroscópicos.

La creación de EDCs DataBank, la única base de datos de disruptores endocrinos con estructuras tridimensionales disponibles en internet (http://edcs.unicartagena.edu.co), constituyó la primera fase de este trabajo de grado. Ésta fue construida en el sistema de gestión de bases de datos relacional MySQL utilizando las listas de disruptores endocrinos *TEDX list* y *UE list of potential endocrine disruptors*, así como una amplia variedad de información proveniente de diferentes repositorios y herramientas de minería de texto. La plataforma web fue desarrollada empleando los lenguajes de programación HTML, CSS, JavaScript y PHP. Actualmente, EDCs DataBank contiene 615 moléculas, presentes en pesticidas, productos naturales e industriales, cosméticos, medicamentos y aditivos alimentarios; entre otros xenobióticos de bajo peso molecular. Por lo cual, esta base de datos puede ser utilizada para estudiar los efectos tóxicos de estas moléculas a través de diversos enfoques *in silico*. Además, esta herramienta web fue desarrollada en un entorno gráfico amigable para el usuario, con contenidos dinámicos que facilitan el análisis de la información, por lo tanto también puede ser utilizada con propósitos académicos, y para la población en general.

Uno de los EDCs más comunes es el BPA (2,2-bis-(4-hidroxifenil)propano), un precursor monomérico empleado en la fabricación de plásticos, resinas y retardantes de llama. Debido a la versatilidad de estos materiales y a la mala gestión de los residuos que lo contienen, el BPA está disperso en diferentes productos de uso diario y en el medio ambiente, siendo uno de los compuestos químicos con mayores volúmenes de producción a nivel industrial. La exposición de los humanos a este xenoestrógeno es muy frecuente y se produce generalmente a través de la vía oral, al ser transferido a los alimentos y bebidas que están en contacto con recipientes plásticos hechos con BPA. Éste ha sido detectado en casi todos los norteamericanos testeados, y relacionado con numerosas condiciones de salud, tales como diabetes, trastornos reproductivos y diferentes tipos de cáncer. Por lo tanto, la OMS considera que la evaluación de sus efectos adversos es prioritaria.

En su mayoría, los mecanismos de toxicidad del BPA han estado relacionados a la activación del receptor de estrógenos (ER) y a cambios en los patrones epigenéticos, sin embargo hay diferentes vías de señalización implicadas en el desarrollo de estas patologías que podrían verse afectadas por este contaminante, a través de un mecanismo de interacción proteína-ligando. Por consiguiente, el objetivo de la segunda etapa fue utilizar un enfoque bioinformático que permitiera la identificación de nuevos blancos proteicos para el BPA. Para esto, se realizaron estudios de acoplamiento entre la estructura optimizada del BPA y 271 proteínas relacionadas con diversos procesos bioquímicos, seleccionadas a través de minería de texto. Subsiguientemente, experimentos de refinamiento del acoplamiento y análisis conformacionales, utilizando LigandScout 3.0, se llevaron a cabo para las macromoléculas escogidas, empleando como criterio de clasificación una elevada afinidad (*Affinity* ≤ -8,0 kcal/mol). Distintas proteínas, entre las que se encuentran ESRRG (-9.9 kcal/mol), y proteínas quinasas de especificidad dual CLK4 (-9.5kcal/mol), CLK1 (-9.1 kcal/mol) y CLK2 (-9.0kcal/mol), presentaron altos valores de afinidad para el BPA *in silico*. Interesantemente, estas macromoléculas están involucradas con el corte y empalme alternativo, y la alteración de este proceso está asociada a la patogénesis de cáncer. Las interacciones entre estas proteínas y el BPA fueron en su mayoría de naturaleza hidrofóbica con la presencia de algunos enlaces de hidrógeno. Por lo tanto, este estudio sugiere que este xenoestrógeno puede tener otros blancos moleculares diferentes al ER.

Las enfermedades asociadas a los EDCs van desde diabetes hasta efectos neurodegenerativos. Sin embargo, éstos afectan de manera contundente los sistemas endocrino y reproductor, siendo de especial preocupación su vinculación al cáncer de seno. Por lo tanto, la tercera etapa de esta tesis consistió en un vHTS para evaluar la afinidad de 189 proteínas relacionadas con este tipo de cáncer, tales como el receptor de estrógenos 1 (ESR1), ERBB2, el receptor de progesterona (PGR), BRCA1 y la globulina fijadora de hormonas sexuales (SHBG), con EDCs de origen urbano. Para la predicción de los complejos proteína-ligando, se utilizó una estrategia de acoplamiento ciego por triplicado en AutoDock Vina 2.0, usando las afinidades de unión como criterio de clasificación. Las estructuras tridimensionales fueron

obtenidas previamente de EDCs DataBank y Protein Data Bank (PDB), preparadas y optimizadas por Sybyl X-2.0. Algunos de los productos químicos que mostraron los mejores puntajes de afinidad para las proteínas relacionadas con cáncer de mama en cada categoría fueron 1,3,7,8-tetraclorodibenzo-p-dioxina, derivados del BPA, ácido sulfónico de perfluorooctano (PFOS) y benzo(a)pireno, para la catalasa, varias proteínas, SHBG y el citocromo P450 1A2, respectivamente. Posteriormente, se realizó la validación experimental de las interacciones de un complejo que obtuvo una afinidad de unión moderada *in silico*, SHBG/BPA, la proteína se produjo utilizando la tecnología de ADN recombinante, y la interacción con el BPA fue evaluada a través de técnicas espectroscópicas. Encontrando, que el BPA se une con la proteína recombinante SHBG, y esto se traduce en un aumento de su contenido de hélices alfas. En resumen, este trabajo muestra el potencial de varios EDCs para interaccionar con proteínas asociadas con cáncer de mama, lo cual sirve de guía para priorizar compuestos en la realización de análisis *in vitro* y para beneficiar a la regulación o la prevención de la exposición de la población.

En general, esta tesis aborda el tema de los EDCs como ligandos de proteínas de cáncer de seno, así como de otras macromoléculas implicadas en rutas de señalización cruciales, en el caso del BPA. Las principales contribuciones realizadas son una nueva base de datos de EDCs adecuada para estudios *in silico*, así como para la población académica y general. La identificación de nuevas dianas para el BPA que podrían guiar los estudios *in vitro* e *in vivo*; así como de EDCs con potencial de perturbar la proteómica relacionada con el cáncer de mama a través de interacciones proteína-ligando. Dos puntos a resaltar de este último estudio fueron la habilidad de los nuevos sustituyentes del BPA de interaccionar con diferentes proteínas asociadas al cáncer de seno con alta afinidad, y el comportamiento promiscuo de algunos xenoestrógenos que hace difícil establecer y controlar sus mecanismos de toxicidad, ya que podrían actuar a través de diferentes rutas bioquímicas.

ABSTRACT

Endocrine Disruptors as Ligands of Breast Cancer Proteins

Endocrine disrupting chemicals (EDCs) are a broad range of compounds that affect the endocrine system. However, their effects are not confined to the deregulation of processes controlled by hormones, but have been epidemiologically linked to a wide range of diseases. Including breast cancer, a complex and heterogeneous disease, which is the considered the most prevalent cancer type in women according to the World Health Organization (WHO).

Some of these compounds are ubiquitous pollutants. Therefore, there is a growing need to study their adverse effects and mechanisms of action. Xenoestrogens are present in many everyday products such as cleansers, makeup, perfumes, plastics, appliances, computers, and even food. In nature, they are being part of certain plants and metabolites produced by fungi and contaminated environments by anthropogenic activity, such as water, air, soil and biota.

Therefore the aim of this thesis was to evaluate the plausibility of EDCs to bind breast cancer proteins, through a three-steps approach: creation of a database of EDCs with three-dimensional structure available, an inverse virtual screening of bisphenol A (BPA) -one of the xenoestrogens to which we are most frequently exposed to- against proteins involved in different signaling pathways, and a virtual high-throughput screening (vHTS) among EDCs and proteins involved in this disease with an evaluation of the interactions of a protein-ligand complex through spectroscopic methods.

The creation of EDCs DataBank, the only database of endocrine disruptors with three-dimensional structures available online (http://edcs.unicartagena.edu.co), constituted the first phase of this project. It was built in relational database management system MySQL using the TEDX list and EU list of potential endocrine disruptors, as well as a wide variety of information from different repositories and text mining tools. The web platform was developed using the languages HTML, CSS, JavaScript and PHP. Currently, EDCs DataBank contains 615 molecules, including pesticides, natural and industrial products, cosmetics, pharmaceuticals and food additives; among other low molecular weight xenobiotics. Therefore, this database can be used to study the toxic effects of these molecules through various *in silico* approaches. In addition, this tool was developed in a user-friendly graphical environment with dynamic contents that facilitates the analysis of information. Consequently, this can also be utilized for academic and general population.

One of the most common EDCs is BPA (2,2-bis(4-hydroxyphenyl)propane), a monomeric precursor used in the manufacture of plastics, resins and flame retardants. This is a chemical with high volumes of production worldwide dispersed in urban and natural environments, due to the versatility of these materials and mismanagement of wastes that contain it. The exposure of humans to this xenoestrogen is very common and usually occurs through oral way, by contaminated foods that have been in contact with materials that can deliver it. This has been detected in nearly all Americans tested and linked to numerous health conditions such as diabetes, reproductive disorders and various cancers. Therefore, WHO considers that the assessment of its adverse effects is a priority.

The mechanisms of toxicity of BPA have been mostly related to the activation of the estrogen receptor (ER) and changes in epigenetic patterns, however there are different signaling pathways involved in the development of these diseases that could be affected by this pollutant, through a protein-ligand interaction mechanism. Therefore, the objective of the second step was to use a bioinformatic approach to identify possible new targets for BPA. Docking studies were performed between the optimized structure of BPA and 271

proteins related to different biochemical processes, as selected by text-mining. Then, refinement docking experiments and conformational analyses, using LigandScout 3.0, were carried out for the proteins selected through affinity ranking (*Affinity* ≤ -8.0 kcal/mol). Several proteins including the estrogen-related receptor gamma (ESRRG; -9.9kcal/mol), and dual specificity protein kinases CLK4 (-9.5kcal/mol), CLK1 (-9.1kcal/mol) and CLK2 (-9.0kcal/mol) presented great *in silico* binding affinities for BPA. Interestingly, these proteins are involved in alternative splicing, and the deregulation of this process is associated with the pathogenesis of cancer. The interactions between those proteins and BPA were mostly hydrophobic with the presence of some hydrogen bonds. Therefore, this study suggests that this endocrine disruptor may have other targets different from the ER.

EDCs have been associated with diseases ranging from diabetes to neurodegenerative disorders. However, the main impact of these pollutants to human health is related to the endocrine and reproductive systems, being of special concern the relationship between them and breast cancer. Then, the third step of this thesis consisted on a vHTS to evaluate the affinity of 189 proteins related to this cancer type, such as the estrogen receptor 1 (ESR1), ERBB2, progesterone receptor (PGR), breast cancer type 1 susceptibility protein (BRCA1) and sex hormone-binding globulin (SHBG), with EDCs from urban sources. A blind docking strategy was employed to screen each protein-ligand pair by triplicate in AutoDock Vina 2.0, using the computed binding affinities as ranking criteria. The three-dimensional structures were previously obtained from EDCs DataBank and Protein Data Bank (PDB), prepared and optimized by Sybyl X-2.0. Some of the chemicals that exhibited the best affinity scores for breast cancer proteins in each category were 1,3,7,8-tetrachlorodibenzo-p-dioxin, BPA derivatives, perfluorooctane sulfonic acid (PFOS) and benzo(a)pyrene, for catalase, several proteins, SHBG and cytochrome P450 1A2, respectively. The experimental validation of the interactions in a complex that presented moderate binding affinity *in silico* SHBG/BPA was carried out, the protein was obtained using DNA recombinant technology and the physical interaction with BPA assessed through spectroscopic techniques. As a result, BPA binds on the recombinant SHBG, and it causes an increase of its alpha helix content. In short, this work

shows the potential of several EDCs to bind breast cancer associated proteins, as a tool to prioritize compounds to perform *in vitro* analysis to benefit the regulation or exposure prevention by the population.

In general, this thesis addressed the issue of EDCs as ligands of breast cancer proteins, as well as other macromolecules involved in crucial pathways, in the case of the BPA. The main contributions were a new database of EDCs suitable for *in silico* studies, as well as for general and academic population. The identification of new targets for BPA that could guide further *in vitro* and *in vivo* studies; as well as EDCs with potential to disrupt the proteomics related to breast cancer through protein-ligand interactions. Two points to highlight of this last study were the ability of the new BPA substituents to bind different breast cancer proteins with high affinity, and the promiscuous behavior of some xenoestrogens which makes difficult to establish and control their mechanisms of toxicity as they could be acting through different ways.

CHAPTER 1

Thesis overview

Endocrine disrupting chemicals (EDCs) are a group of molecules capable to interfere with the homeostasis of the body, usually through the mimicry of natural hormones leading to activation or blockage of its receptors[1]. Recently, many of these compounds have been associated with the development or increased susceptibility to breast cancer[2]. As hundreds of chemicals belong to this group[3] and humans are highly exposed to these through various everyday products[4]; the main purpose of this project was the *in silico* identification of EDCs-breast cancer proteins complexes, in order to help to prioritize them for *in vitro* and *in vivo* studies. In this chapter, the rationale, hypothesis, pertinence, objectives and structure of this thesis will be discussed.

1.1. RATIONALE

The risk of breast cancer has two main components: inheritable and environmental related or lifestyle. Interestingly, the inheritable component has a small contribution, as between 5 to 27 % of all breast cancers are attributed to factors such as specific gene mutations, certain traits, and metabolic issues; being the environmental risks the major factor[5]. These include the exposure to chemicals, foods, waste, occupational hazards, cosmetics, drugs and stress, among others. Although these stimuli have not been clearly characterized, EDCs are known to have the ability to affect the normal breast development, and lead to adverse effects associated with a decreased longevity, increasing sensitivity to carcinogenic chemicals, hyperplasia or tumor[6].

Currently, the US Environmental Protection Agency (EPA) estimates there exist approximately 10,000 EDCs among the common daily exposures that could impose any risk of disease[7]. Some critical points in the debate around these compounds include the potential to affect the health of humans and wildlife, as well as their progeny[8]; the massive volume of production of some of them and the chronic exposure to the general population through the environment and everyday products[9]. These points have enhanced the scientific concern regarding the safety of these chemicals in recent years. Therefore, the first step of this thesis project consisted on the construction of a repository with this kind of chemicals suitable for *in silico* approaches, general population and educational purposes.

One of the industrial chemicals with the highest production volume (10^9 kg/year) worldwide is bisphenol A (BPA)[10]. This is a known endocrine disrupting chemical (EDC)[11] with estrogenic action demonstrated both *in vivo* and *in vitro* studies[12], and used in the production of plastics, resins and flame retardants[13]. This pollutant contaminates food through its release from plastic containers, which is of special concern since it is the primary exposure route in humans[14], and it has been detected in over 90% of individuals in North America[15], which means we are all exposed to this pollutant. Recent influential publications have associated BPA exposure to several diseases[12a], such as diabetes, obesity, reproductive disorders, cardiovascular diseases, birth defects, chronic respiratory and kidney diseases, as well as breast cancer[16]. Therefore and based on the existing data regarding the relationship between BPA exposure and several pathologies, we hypothesized that some of their related proteins could be targets of BPA. Accordingly, in the second step of this thesis, an inverse docking strategy was carried out to evaluate *in silico* the possible interaction between BPA with proteins having a role in critical disease pathways[17].

The exposure to EDCs induces adverse effects, specially related to reproduction, development and different cancer types[18], including breast cancer[19] , which according to the World Health Organization (WHO) is the most common cancer in women (http://www.who.int/cancer/detection/breastcancer/en/index1.html). The

classic hallmark mechanism in breast cancer is the activation of estrogen receptor (ER). For this reason, the responses due to EDCs exposures have been mainly explored and attributed to this signaling pathway[20]. In general, the understanding on the underlying mechanisms of breast cancer[8] are limited because of its complexity and heterogeneity; nevertheless, there are some proteins recognized as important in its initiation and progression, specially hormone receptors such as ER, progesterone receptor (PGR), and other proteins, in particular, the human epidermal growth factor receptor 2 (HER-2), breast cancer type 1 susceptibility protein (BRCA1) and breast cancer type 2 susceptibility protein (BRCA2)[21]. Therefore, computational toxicology approaches are needed to help to prioritize chemicals for screening[22]. In the third step of this thesis, an *in silico* evaluation of the potential of EDCs to target breast cancer proteins was carried out, through a virtual high-throughput screening (vHTS) and the *in vitro* validation of the interactions among one of the predicted protein/ligand complexes by spectroscopic methods.

1.2. HYPOTHESIS

Proteins associated with breast cancer are molecular targets for EDCs.

1.3. PERTINENCE

Breast cancer is the most common cancer in women worldwide (http://www.who.int/cancer/detection/breastcancer/en/index1.html). The WHO estimated that over 508,000 women died in 2011 due to this disease (http://www.who.int/cancer/detection/breastcancer/en/index1.html). In Latin America, the annual incidence of breast cancer is 114,900 cases, causing 37,000 deaths per year, making it the highest rate of all cancers suffered by women, with a tendency to increase. In fact, twice as many cases in 2030 are expected. In Latin America about 30-40% of diagnoses correspond to metastasis, indicating poor survival, in part due to late diagnosis and poor

access to treatment[23]. In Colombia, according to the latest data published by the WHO in 2011, deaths from breast cancer reached the 2,120 or 1.23% of total deaths (http://www.worldlifeexpectancy.com/colombia-breast-cancer).

The prevalence of reproductive disorders and cancer has been increasing due to exposure to chemicals that affect the action of natural hormones, which have been associated with these diseases. Environmental factors as EDCs have the ability to affect the normal breast development and have adverse consequences on life expectancy, especially when exposure occurs early in life[5]. Therefore, it is important to identify EDCs and their targets of action[24], in order to contribute to the generation of data leading to decision making and generation of relevant legislation in the fields of public policy on health and environment.

Chemical compounds that interfere with estrogen signaling affecting endocrine homeostasis, also constitute a potential for the occurrence of breast cancer[25], diabetes[26] and other metabolic diseases. The EPA aware of the potential for these substances to accumulate and affect subsequent generations, disrupting human development have indicated the need to assess their levels and effects, including phthalate esters, alkylphenols and BPA[13c]. This is of special concern due to the high volumes of production of some of these chemicals worldwide, and their presence in many everyday goods, therefore the use of *in silico* methods such as vHTS is helpful to facilitate the search for new molecular targets reducing costs in terms of time and money[24].

1.4. OBJECTIVES AND THESIS STRUCTURE

This project was carried out in order to identify target proteins of EDCs directly or indirectly involved in the breast cancer disease. This was structured in three main steps, as follows:

- Construction of a database of EDCs with three-dimensional structures available for virtual screening.
- Search of protein targets of BPA using computer aided tools.
- *In silico* identification of molecular targets of EDCs among the breast cancer proteins and *in vitro* evaluation of the interactions between a protein-ligand pair with good theoretical affinity.

The introduction is presented in Chapter 2. This describes the main antecedents and state-of-the-art knowledge on endocrine disruptors and their association with breast cancer, as well as a brief explanation of the techniques used. The subsequent chapters show the development, results, discussions and conclusions of each single step.

The building of the database, EDCs DataBank, is explained in Chapter 3. The main purpose of this step was the creation of a repository with structural data useful for computational toxicology approaches, an emerging area that needs to be incentivized to deal with the challenges that the field of EDCs are facing nowadays. Other objectives of this step were the implementation of strategies to make the database useful to inform general population about the potential risk of exposure in everyday products, as well as for educational purposes at different levels of expertise.

Chapter 4 presents the study of the potential interaction of BPA with several pathways involved in important physiological processes by an inverse virtual screening and a validation using available experimental data. The purpose of this step was to identify molecular targets of this pollutant, with a role in different diseases pathways, using computer aided tools.

Finally, the interaction between urban EDCs and breast cancer proteins is presented in Chapter 5. According to the sources of exposure, these were classified in dioxins and related molecules, plastics and other types of polymers, everyday products and miscellaneous compounds. In addition, the *in vitro* validation of the interaction of one protein-ligand complex by spectroscopic methods is presented in this chapter. The objectives of this

phase were to identify proteins related to breast cancer employing data mining tools and to conduct a vHTS between proteins associated with breast cancer and EDCs present in urban areas. Furthermore, the *in silico* analysis of the interactions of the best complexes, and *in vitro* evaluation of one of them were performed. This last, required the obtaining of a protein by recombinant DNA technology and the assessment of the conformational changes of the protein after its incubation with the EDC by circular dichroism.

The main contributions of this thesis project are presented in Chapter 6, entitled conclusions and final remarks, and a broad range of supporting data is presented in the annexes.

REFERENCES

1. Bonefeld-Jørgensen, E. C.; Ghisari, M.; Wielsøe, M.; Bjerregaard-Olesen, C.; Kjeldsen, L. S.; Long, M., Biomonitoring and Hormone-Disrupting Effect Biomarkers of Persistent Organic Pollutants *In vitro* and Ex Vivo. Basic Clin. Pharmacol. 2014, 115 (1), 118-128.
2. Yang, M.; Ryu, J.-H.; Jeon, R.; Kang, D.; Yoo, K.-Y., Effects of bisphenol A on breast cancer and its risk factors. Arch Toxicol 2009, 83 (3), 281-285.
3. Bergman, A.; Heindel, J. J.; Kasten, T.; Kidd, K. A.; Jobling, S.; Neira, M.; Zoeller, R. T.; Becher, G.; Bjerregaard, P.; Bornman, R., The impact of endocrine disruption a consensus statement on the state of the science. Environ. Health Perspect. 2013, 121 (4), A104.
4. (a) Lomenick, J. P.; Calafat, A. M.; Castro, M. S. M.; Mier, R.; Stenger, P.; Foster, M. B.; Wintergerst, K. A., Phthalate exposure and precocious puberty in females. J. Pediatr. 2010, 156 (2), 221-225; (b) MacLusky, N. J.; Hajszan, T.; Leranth, C., The environmental estrogen bisphenol A inhibits estradiol-induced hippocampal synaptogenesis. Environ. Health Perspect. 2005, 675-679.
5. Macon, M. B.; Fenton, S. E., Endocrine disruptors and the breast: early life effects and later life disease. J. Mammary Gland Biol. 2013, 18 (1), 43-61.
6. Soto, A. M.; Sonnenschein, C., Environmental causes of cancer: endocrine disruptors as carcinogens. Nat Rev Endocrinol 2010, 6 (7), 363-370.
7. Teitelbaum, S. L.; Belpoggi, F.; Reinlib, L., Advancing research on endocrine disrupting chemicals in breast cancer: Expert panel recommendations. Reprod. Toxicol. 2015, 54 (0), 141-147.
8. Ihara, M.; Ihara, M. O.; Kumar, V.; Narumiya, M.; Hanamoto, S.; Nakada, N.; Yamashita, N.; Miyagawa, S.; Iguchi, T.; Tanaka, H., Co-occurrence of Estrogenic and Antiestrogenic Activities in Wastewater: Quantitative Evaluation of Balance by *in vitro* ERα Reporter Gene Assay and Chemical Analysis. Environ. Sci. Technol. 2014, 48 (11), 6366-6373.
9. (a) Wu, R. S. S.; Zhou, B. S.; Randall, D. J.; Woo, N. Y. S.; Lam, P. K. S., Aquatic Hypoxia Is an Endocrine Disruptor and Impairs Fish Reproduction. Environ. Sci. Technol. 2003, 37 (6), 1137-

1141; (b) Kanno, J.; Kato, H.; Iwata, T.; Inoue, T., Phytoestrogen-Low Diet for Endocrine Disruptor Studies. J. Agric. Food Chem. 2002, 50 (13), 3883-3885; (c) Benotti, M. J.; Trenholm, R. A.; Vanderford, B. J.; Holady, J. C.; Stanford, B. D.; Snyder, S. A., Pharmaceuticals and Endocrine Disrupting Compounds in U.S. Drinking Water. Environ. Sci. Technol. 2009, 43 (3), 597-603; (d) Westerhoff, P.; Yoon, Y.; Snyder, S.; Wert, E., Fate of Endocrine-Disruptor, Pharmaceutical, and Personal Care Product Chemicals during Simulated Drinking Water Treatment Processes. Environ. Sci. Technol. 2005, 39 (17), 6649-6663; (e) Guenther, K.; Heinke, V.; Thiele, B.; Kleist, E.; Prast, H.; Raecker, T., Endocrine Disrupting Nonylphenols Are Ubiquitous in Food. Environ. Sci. Technol. 2002, 36 (8), 1676-1680; (f) Park, B. J.; Palace, V.; Wautier, K.; Gemmill, B.; Tomy, G., Thyroid Axis Disruption in Juvenile Brown Trout (Salmo trutta) Exposed to the Flame Retardant ß-Tetrabromoethylcyclohexane (ß-TBECH) via the Diet. Environ. Sci. Technol. 2011, 45 (18), 7923-7927.

10. Doerge, D. R.; Twaddle, N. C.; Vanlandingham, M.; Fisher, J. W., Pharmacokinetics of bisphenol A in neonatal and adult Sprague-Dawley rats. Toxicol. Appl. Pharmacol. 2010, 247 (2), 158-165.
11. (a) Becker, K.; Güen, T.; Seiwert, M.; Conrad, A.; Pick-Fuß, H.; Müller, J.; Wittassek, M.; Schulz, C.; Kolossa-Gehring, M., GerES IV: Phthalate metabolites and bisphenol A in urine of German children. Int. J. Hyg. Environ. Health. 2009, 212 (6), 685-692; (b) Li, D.-K.; Zhou, Z.; Miao, M.; He, Y.; Wang, J.; Ferber, J.; Herrinton, L. J.; Gao, E.; Yuan, W., Urine bisphenol-A (BPA) level in relation to semen quality. Fertil Steril. 2011, 95 (2), 625-630.e4.
12. (a) Balakrishnan, A.; Stearns, A. T.; Rounds, J.; Irani, J.; Giuffrida, M.; Rhoads, D. B.; Ashley, S. W.; Tavakkolizadeh, A., Diurnal rhythmicity in glucose uptake is mediated by temporal periodicity in the expression of the sodium-glucose cotransporter (SGLT1). Surgery 2008, 143 (6), 813-818; (b) Berger, R. G.; Foster, W. G.; deCatanzaro, D., Bisphenol-A exposure during the period of blastocyst implantation alters uterine morphology and perturbs measures of estrogen and progesterone receptor expression in mice. Reprod. Toxicol. 2010, 30 (3), 393-400.
13. (a) Alamdari, A.; Nasab, S. S.; Jahanmiri, A., Kinetic study of adductive crystallization of bisphenol-A. Chem. Eng. Res. Des. 2010, 88 (12), 1615-1623; (b) Avissar-Whiting, M.; Veiga, K. R.; Uhl, K. M.; Maccani, M. A.; Gagne, L. A.; Moen, E. L.; Marsit, C. J., Bisphenol A exposure leads to specific microRNA alterations in placental cells. Reprod. Toxicol. 2010, 29 (4), 401-406; (c) Ballesteros, O.; Zafra, A.; Navalón, A.; Vílchez, J. L., Sensitive gas chromatographic–mass spectrometric method for the determination of phthalate esters, alkylphenols, bisphenol A and their chlorinated derivatives in wastewater samples. J. Chromatogr. A. 2006, 1121 (2), 154-162.
14. (a) Cantonwine, D.; Meeker, J.; Hu, H.; Sanchez, B.; Lamadrid-Figueroa, H.; Mercado-Garcia, A.; Fortenberry, G.; Calafat, A.; Tellez-Rojo, M., Bisphenol a exposure in Mexico City and risk of prematurity: a pilot nested case control study. Environ. Health 2010, 9 (1), 62; (b) Crain, D. A.; Eriksen, M.; Iguchi, T.; Jobling, S.; Laufer, H.; LeBlanc, G. A.; Guillette Jr, L. J., An ecological assessment of bisphenol-A: Evidence from comparative biology. Reprod. Toxicol. 2007, 24 (2), 225-239.
15. Cabaton, N. J.; Wadia, P. R.; Rubin, B. S.; Zalko, D.; Schaeberle, C. M.; Askenase, M. H.; Gadbois, J. L.; Tharp, A. P.; Whitt, G. S.; Sonnenschein, C.; Soto, A. M., Perinatal Exposure to Environmentally Relevant Levels of Bisphenol A Decreases Fertility and Fecundity in CD-1 Mice. Environ. Health Perspect. 2010, 119 (4).

16. Rezg, R.; El-Fazaa, S.; Gharbi, N.; Mornagui, B., Bisphenol A and human chronic diseases: Current evidences, possible mechanisms, and future perspectives. Environ. Int. 2014, 64, 83-90.
17. Kemsley, J., Rethinking breast cancer toxicology. Chem. Eng. News 2010, 88 (10), 40-41.
18. (a) Hauser, R.; Skakkebaek, N. E.; Hass, U.; Toppari, J.; Juul, A.; Andersson, A. M.; Kortenkamp, A.; Heindel, J. J.; Trasande, L., Male Reproductive Disorders, Diseases, and Costs of Exposure to Endocrine-Disrupting Chemicals in the European Union. J. Clin. Endocrinol. Metab. 2015, 100 (4), 1267-1277; (b) Bellanger, M.; Demeneix, B.; Grandjean, P.; Zoeller, R. T.; Trasande, L., Neurobehavioral Deficits, Diseases, and Associated Costs of Exposure to Endocrine-Disrupting Chemicals in the European Union. J. Clin. Endocrinol. Metab. 2015, 100 (4), 1256-1266; (c) Bedia, C.; Dalmau, N.; Jaumot, J.; Tauler, R., Phenotypic malignant changes and untargeted lipidomic analysis of long-term exposed prostate cancer cells to endocrine disruptors. Environ. Res. 2015, 140 (0), 18-31; (d) Prins, G. S., Endocrine disruptors and prostate cancer risk. Endocr. Relat. Cancer 2008, 15 (3), 649-656.
19. Brody, J. T., Breast cancer risk in relation to occupations with exposure to carcinogens and endocrine disruptors: a Canadian case-control study. National Network on Environments and Women's Health: 2012.
20. Lee, H.-R.; Hwang, K.-A.; Nam, K.-H.; Kim, H.-C.; Choi, K.-C., Progression of breast cancer cells was enhanced by endocrine-disrupting chemicals, triclosan and octylphenol, via an estrogen receptor-dependent signaling pathway in cellular and mouse xenograft models. Chem. Res. Toxicol. 2014, 27 (5), 834-842.
21. Hulka, B. S.; Moorman, P. G., Breast cancer: hormones and other risk factors. Maturitas 2001, 38 (1), 103-113.
22. Erickson, B. E., EPA RETOOLS ENDOCRINE PROGRAM. Chem. Eng. News 2013, 91 (11), 30-32.
23. Justo, N.; Wilking, N.; Jönsson, B.; Luciani, S.; Cazap, E., A review of breast cancer care and outcomes in Latin America. Oncologist 2013, 18 (3), 248-256.
24. Nashev, L. G.; Schuster, D.; Laggner, C.; Sodha, S.; Langer, T.; Wolber, G.; Odermatt, A., The UV-filter benzophenone-1 inhibits 17β-hydroxysteroid dehydrogenase type 3: Virtual screening as a strategy to identify potential endocrine disrupting chemicals. Biochem. Pharmacol. 2010, 79 (8), 1189-1199.
25. Jenkins, S.; Betancourt, A. M.; Wang, J.; Lamartiniere, C. A., Endocrine-active chemicals in mammary cancer causation and prevention. J. Steroid. Biochem. Mol. Biol. 2012, 129 (3), 191-200.
26. Sakurai, K.; Kawazuma, M.; Adachi, T.; Harigaya, T.; Saito, Y.; Hashimoto, N.; Mori, C., Bisphenol A affects glucose transport in mouse 3T3-F442A adipocytes. Br. J. Pharmacol. 2004, 141 (2), 209-214.

CHAPTER 2

Introduction

A conceptual review of the most relevant information regarding the topic of this thesis will be made in this chapter. It starts with the state-of-art knowledge of the EDCs, followed by the epidemiological relationship between these pollutants and breast cancer, the molecular mechanisms of this pathology, and finally a brief description of the main techniques used in this project will be presented.

2.1. ENDOCRINE DISRUPTING CHEMICALS

EDCs are a broad range of molecules with potential to affect the endocrine system through different ways, metabolic, epigenetic and DNA damage[1b-e]. Most of them have the capability to alter certain signal transduction systems associated with hormones, including those related to estrogen, androgen and thyroid hormone receptors[2]. However, other proteins and toxicity mechanisms are also involved[3]. These chemicals were classified for first time in the 1990's and defined by the EPA as exogenous agents with the ability to interfere with the synthesis, secretion, transport, binding, action, or elimination of natural hormones in the body which are responsible for the maintenance of homeostasis, reproduction, development, and/or behavior[4].

Some xenoestrogens cause adverse ecological and human health effects[5]. Indeed, an increasing evidence from both experimental animal and epidemiology studies suggests that there is a link between exposure to EDCs and many conditions[6]. The main adverse consequences in the human body are associated with reproductive disorders[7], dysplasia and the development or

progression of cancers[8], as well as obesity, diabetes, heart disease, neurodevelopmental and neurodegenerative disorders[9].

These pollutants contaminate indoor and outdoor areas. Xenoestrogens can be bioaccumulated and deposited in the environment, mainly in sediments of water bodies and biosolids, due to the high octanol-water partition coefficients that exhibit a good proportion of them[10]. In urban areas, these are found in various products in daily use, such as plastics, fire retardants, food containers; as well as in wastewater treatment plants[10]. In fact, hundreds of them have been detected in humans and wildlife, even in remote parts of the Earth, such as the Arctic. Therefore, it is almost impossible to find a population that is not exposed to EDCs[10].

Consequently, there is a growing international concern[11], due to the potential risk that those represent to general population[12]. Actually, the WHO and the United Nations Environment Programme (UNEP) (http://www.who.int/ceh/publications/endocrine/en/) indicates EDCs are a worldwide problem requiring global solutions. In addition, the identification of compounds with estrogenic activity remains a challenge and what is known of them can be comparable with the tip of an iceberg, being of particular interest the large production volumes of some of them[13].

These belong to diverse chemical groups with very dissimilar structures, biochemical properties and mechanisms of action[14], such as metals, organic and inorganic compounds. Some of the most popular EDCs are presented in the following sections.

2.1.1. Synthetic Xenoestrogens

A wide variety of man-made EDCs currently in use or that have been utilized in the past with different purposes, and continues to contaminate the environment due to their persistence (Figure 2.1).

2.1.1.1. Flame retardants. These are additives used in different materials, such as electronics, textiles and plastics, to prevent fire[15]. Some of the brominated organic compounds in this group are: polybrominated diphenyl ethers (PBDEs), hexabromocyclododecane (HBCD), tetrabromobisphenol A (TBBPA) and polybrominated biphenyls (PBBs)[16]. The main concern about these chemicals is their fate and stability, as organohalogenated compounds are some of the most persistent classes of environmental pollutants related to adverse health effects in humans and wildlife, including endocrine disruption[17]. Three of these compounds are included in the Stockholm convention (HBCD, pentabromodiphenyl ether and octabromodiphenyl ether)[18]. In Colombia, there is not specific legislation for this kind of pollutants, however, the partial regulation of hazardous waste is regulated by "Decreto 4741 de 2005" (http://www.ficem.org/normas/Colombia/Decreto_4741.pdf).

2.1.1.2. Polychlorinated biphenyls (PCBs), dibenzo-p-dioxins (PCDDs), dibenzofurans (PCDFs), and related compounds. These are industrial compounds or derivatives, which have been widely identified in the environment and in waste dumpsites[19]. PCBs are persistent organic pollutants (POPs) that have been used for decades in several industrial applications. Although their production was restricted from 1970's in most countries, substantial amounts have been detected in various environmental and biological matrices[20]. In Colombia, their use is regulated by "Manual de manejo de PCBs para Colombia" (http://www.crc.gov.co/files/Respel/Manual_PCBs.pdf).

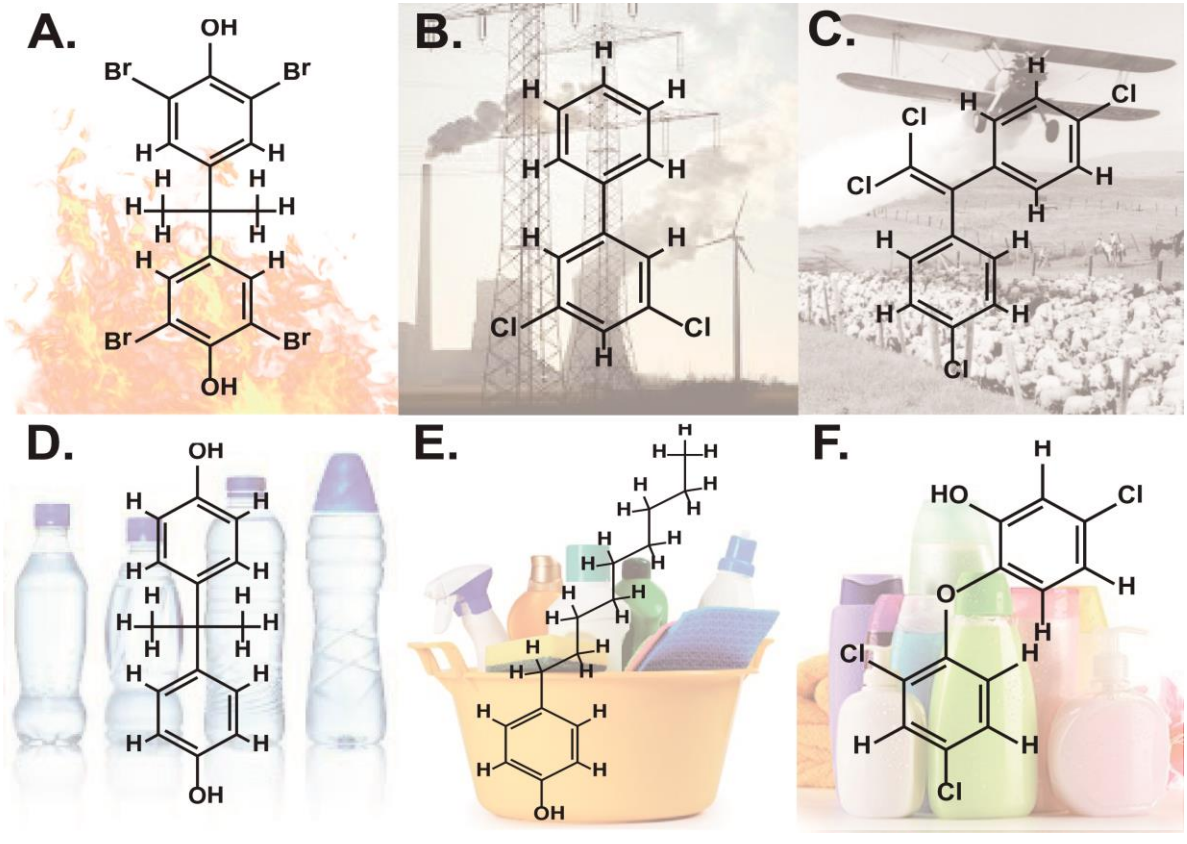


Figure 2.1. Synthetic xenoestrogens in everyday products. (A) tetrabromobisphenol A (TBBPA), (B) 3,5-dichlorobiphenyl, (C) p,p'-DDE, (D) BPA, (E) 4-nonylphenol and (F) triclosan.

PCDDs are formed during the combustion of chlorine-containing fuels in municipal solid waste incinerators and coal fired power plants[21], as well as thermal and industrial processes[22], being 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD), the most toxic dioxin[19]. In Colombia, the release of these compounds is regulated by establishing rules and maximum permissible emission limits for incinerators and crematoria of solid and liquid waste by "Resolución 058 de 2002" (http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=14363).
PCDFs, are produced unintentionally due to incomplete combustion, during manufacturing of other chlorinated compounds, and in the chlorine bleaching of wood-pulp[23]. In Colombia, these are partially regulated by "Decreto 4741 de 2005" (http://www.ficem.org/normas/Colombia/Decreto_4741.pdf) that contemplates the prevention and management of hazardous waste.

2.1.1.3. Plasticizers. The main EDCs in this group are BPA and the phthalates: bis(2-ethylhexyl)phthalate (DEHP) and dibutyl phthalate (DBP). BPA is utilized in the production of polycarbonate plastics and epoxy resins and have the capability to interact with human ER and acts as antagonist with the androgen receptor (AR)[24]. This pollutant is present in food containers, and recently different analogs have appeared in the market to substitute it[25]. DEHP is widely employed as a plasticizer in manufacturing of articles made of polyvinyl compounds and it is considered a reproductive and developmental toxicant in humans and animals. DBP is a phthalate used primarily as plasticizer to add flexibility to plastics, as a component in latex adhesives, cosmetics and as a solvent for dyes[26].

2.1.1.4. Pesticides. Many chemicals that have been identified as EDCs are pesticides. Interestingly, an increased rate of breast cancer in women from areas of high contamination by these pollutants is reported in the literature[27]. Historically, *p,p'*-DDE, a chemical compound formed by dehydrohalogenation of the insecticide DDT, have received much attention due their effect as endocrine disrupter acting primarily as anti-androgenic[28]. Nowadays, probably one of the most popular pesticides with endocrine activity is the molluscicide, tributyltin. It is an organotin compound that can be biomagnified in marine organism, and can have severe effects in biota

including imposex, which consists on an irreversibly affectation of the sex organs, due to the aromatase inhibition[3a]. In Colombia, the use of pesticides is regulated by "Decreto 775 del 16 de Abril de 1990" (http://www.icbf.gov.co/cargues/avance/docs/Decreto_0775_1990.htm), and the use of some organochloride pesticides with known toxicity effects, such as aldrin, heptachlor, dieldrin, chlordane and camphechlor by "Decreto 305 de 1988" (http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=14520) and "Decreto 704 de 1986" (http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=14521).

2.1.1.5. Household and personal care products. The presence of xenoestrogens in these products is quite alarming as some of them are directly applied in our body. Surfactants, antimicrobial agents, UV blockers and cosmetics, are some of the products that contain EDCs[29]. Probably some of the most popular chemicals in this group are parabens, 4-nonylphenol and triclosan. Parabens are widely employed as antimicrobial preservatives in cosmetics[30]. 4-nonylphenol is utilized for the production of nonionic surfactants, detergents, emulsifiers, pesticides, lubricants and oil additives, utilized in daily life and at industrial level. This is a persistent pollutant, widely detected in nature, treatment plants, sewage, sludge, air and even in drinking water and food[31]. Triclosan is added as an antibacterial agent in liquid toothpaste, soap, shampoo, and cosmetics[32]

2.1.2. Natural xenoestrogens

A well-known group of natural occurring EDCs are phytoestrogens, plant-derived compounds that are structurally or functionally similar to estradiol. These are present in different foods, such as soy products, tofu, alfalfa, apples, green tea, sesame, and wheat, among others[33]. Some of them are presented in Figure 2.2.

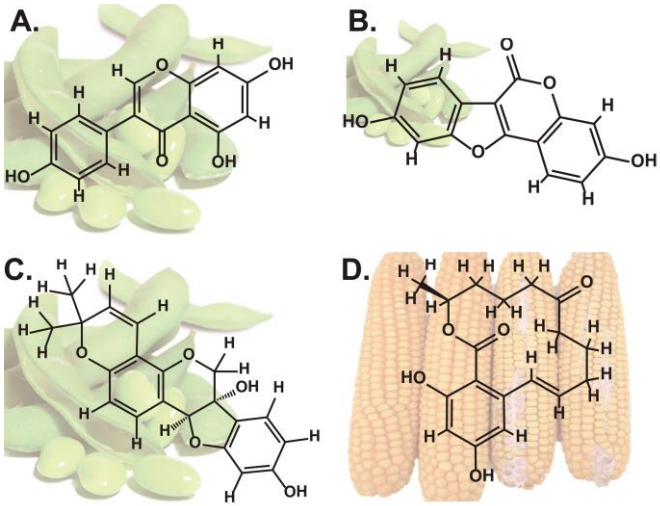


Figure 2.2. Natural xenoestrogens in foods. (A) genistein, (B) coumestrol, (C) glyceollin I and (D) zearalenone.

They can act as endocrine modulators through mimicking natural hormones[34], some of them are genistein, coumestrol[7] and glyceollin[35]. Their consumption have been associated to beneficial health effects, and even when their protective effects against breast cancer have been suggested the studies in humans are not yet conclusive[36]. These compounds are able to bind mammalian ER and may alter risk of breast cancer by their weak inhibitory effect on aromatase, thereby lowering the amount of circulating estrogen. As a result, breast tumor proliferation may be decreased. However, other mechanisms such as inhibition of cancer cell growth and angiogenesis[37], as well as stimulation of apoptosis, are also under discussion[38]. Some of the known targets are cyclin dependent kinases (p21/p27), the tumor suppressor gene p53, pro-apoptotic and anti-apoptotic genes including Bax and Blc-2[39], as well as the fibroblast growth factor receptor 2 (FGFR2), which has been associated with risk of sporadic postmenopausal breast cancer[40]. Besides, their protective effects through changes in the epigenetics patterns of genes with a pivotal role in this disease such as BRCA1 and BRCA2 have been reported for some of them[41]. Phytoestrogens are also good antioxidants with demonstrated both reactive oxygen species (ROS) scavenging ability and anti-

cancer characteristics[42]. ROS exposure is a key point during oxidative damage to cellular components including DNA, protein, and lipids, and play a role in mechanisms of tumorogenesis[43].

Other natural occurring EDCs are some mycotoxins such as zearalenone. Zearaleone is produced by a variety of *Fusarium* species, in particular *Fusarium graminearum*, which grows on crops such as maize and wheat. This xenoestrogen is more potent than some synthetic EDCs, and has been reported to produce infertility[44] and cancer[45].

2.2. STATE OF THE SCIENCE OF EDCs 2012

This document provides an assessment of the global status of scientific knowledge on exposure to EDCs and their effects. It was prepared by a group of experts for the UNEP and WHO and its last update was in 2012. This includes scientific information on the main aspects to consider in making decisions about the future of human health and wildlife.

The key points discussed in the State of the Science of Endocrine Disrupting Chemicals, 2012 (http://www.who.int/ceh/publications/endocrine/en/index.html) are:

- The human health and wildlife depends on their ability to reproduce and develop normally, which is not possible without a healthy endocrine system.
- The high incidence of many endocrine-related disorders in humans with increasing trends.
- The frequency of genital malformations, such as undescended testicles and malformations of the penis (hypospadias), in male babies has increased over time or unfavorably stabilized at high rates.
- The occurrence of adverse pregnancy outcomes such as premature delivery and low birth weight has augmented in many countries.

- Neurobehavioral disorders associated with thyroid disorders affect a large proportion of children in some countries have increased in recent decades.
- The overall rates of cancer related to the endocrine system (breast, endometrium, ovary, prostate, testes and thyroid) have been growing in the past 40-50 years.
- The trend toward earlier breast development in girls in all countries, which is a risk factor for the onset of breast cancer.
- The prevalence of obesity and type 2 diabetes has increased dramatically worldwide over the past 40 years.
- There are around 800 chemicals known or suspected to be capable of interfering with the hormone receptors, hormones synthesis or their conversion. Nevertheless, only a small fraction of these chemicals have been investigated in tests capable of identifying endocrine effects in an intact organism.
- The vast majority of chemicals in commercial use today have not been tested at all. This lack of data makes significant uncertainties about the true extent of the risks of chemicals that could disrupt the endocrine system.

2.3. ENDOCRINE DISRUPTION AND BREAST CANCER

An increase in non-communicable diseases in humans and wildlife has taken place in the past 40 years, which indicates the pivotal role of the etiology of environmental diseases[46]. The exposure to EDCs is an important component of this, beyond nutrition and other factors. Therefore, its reduction could has a significant impact on disease prevention, which is better than treatment, not only in economic terms, but also in quality of life[47].

Interestingly, the increase in breast cancer incidence matches with an enormous growth of man-made and persistent chemicals in our environment, some of which have endocrine disrupting features[48]. There exists numerous reports indicating that exposure to EDCs may increase the risk of disorders of the reproductive tract[49] and breast cancer[50], one of the most prevalent cancer

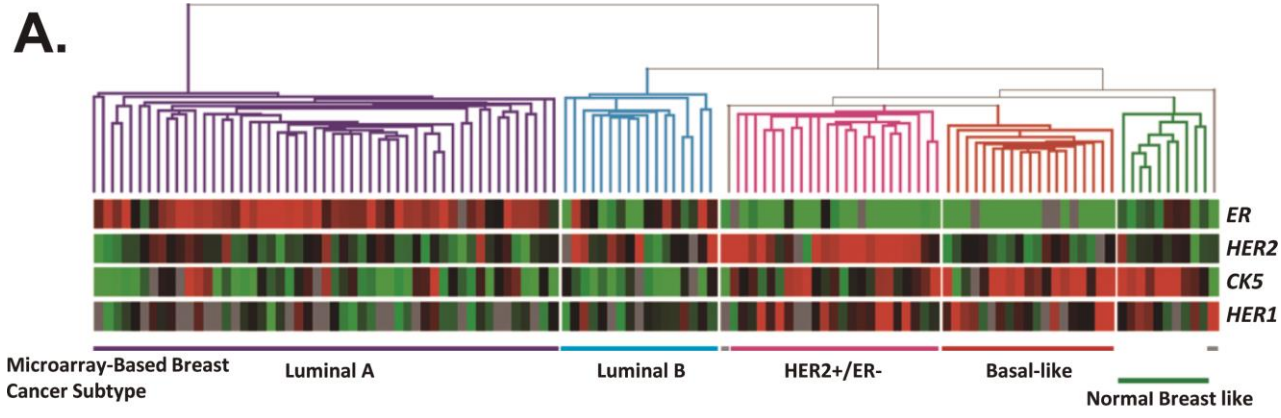
types in woman[51]. The epidemiological evidence is solid indicating that the susceptibility to get breast cancer is linked to the reproductive history and changes in the hormonal microenvironment, which could be mediated by xenoestrogens[48]. In addition, they can affect the normal breast tissue development and increase the sensitivity to carcinogenics[3b] through different ways, such as metabolic, epigenetic and DNA damage [1a-d, 52]. However, a causative link between exposure to these EDCs and human diseases is difficult to establish as many of these pollutants are ubiquitous and no unexposed controls exist[48, 53].

2.4. BREAST CANCER

Breast cancer is a hormone-dependent disease[54] and the most common cancer type in women worldwide[55]. Some risk factors identified by epidemiologic studies include age[56], race, ethnicity, family history of cancer and genetic traits, as well as exposure to chemicals, such as alcohol and xenoestrogens[57]. Due to its heterogeneity, this disease can be classified in different subtypes according to histological, biological and molecular characteristics[58]. These are presented in Figure 2.3, showing a hierarchical cluster with the expression patterns of some of the immunohistochemical markers in breast cancer, for a group of human patients[59].

Luminal A is the most common subtype of breast cancer and it is characterized by the expression of genes activated by the ER transcription factor that are typically expressed in the luminal epithelium lining the mammary ducts and by a low expression of genes related to cell proliferation[60]. Luminal B have a more aggressive phenotype with higher recurrence rates and shorter survival times after relapse and presents an increased expression of proliferation genes[61]. HER-positive is characterized by a high expression of the HER-2 and other genes associated with its pathway, as well as cell proliferation genes[60]. Basal-like is characterized by the expression of genes usually present in normal breast myoepithelial cells and luminal epithelium, but at levels significantly lower than those of luminal carcinomas.

A.



B.

Subtype	Prevalence among total breast cancer (%)	Immunohistochemical profile
Luminal A	50-60	ER+ and/or PR+, HER2-
Luminal B	10-20	ER+ and/or PR+, HER2+
HER2-positive	15-20	ER-, PR-, HER+
Basal-like	10–20	ER-, PR-, HER2-, CK5/6+ and/or HER1+
Normal breast-like	5–10	-

Figure 2.3. Breast cancer subtypes. (A) Hierarchical cluster with the expression patterns of ER, HER-2, cytokeratin (CK5) and the human epidermal growth factor receptor (HER-1), for the main subtypes of breast cancer (Adapted from Carey et al.[59] and Eroles et al.[60]) and (B) their immunohistochemical profiles.

In clinical practice, this is also known as triple negative because the lack of expression of the ER, PGR and HER-2. Some of its features are aggressive clinical course, earlier age of diagnosis, and lack of efficient treatment[60, 62]. Normal breast subtype expresses genes characteristic of adipose tissue and lacks the expression of ER, HER-2 and PGR, so these tumors can also be classified as triple negative, presenting an intermediate prognosis between luminal and basal-like subtypes and usually do not respond to neo-adjuvant chemotherapy[60].

2.4.1. Pathways associated to breast cancer

The most representative hallmarks shared by the different cancer types are the ability to sustain chronic proliferation, evading growth suppressors, resisting cell death, enabling replicative immortality, inducing angiogenesis, and activating invasion and metastasis. Some of the signaling pathways involved in these processes are PI3-kinase (PI3K), AKT, and mTOR (mammalian target of rapamycin), stimulated by survival signals to block apoptosis and autophagy; signaling by oncoproteins such as RAS, MYC, and RAF; tumor suppressors that operate in various ways to limit cell growth and proliferation like RB (retinoblastoma-associated), growth factor-beta (TGF-ß) and tumor protein (TP53); as well as apoptotic and antiapoptotic pathways like Bcl-2, Bax and Bak[63].

In breast cancer, the main circuits are HER-2, ER, IGF1R (type 1 insulin-like growth factor), PI3K/AKT, mTOR, AMPK (AMP-activated protein kinase) and those related to angiogenesis. The de-regulation of HER-2 pathway in breast cancer leads to sustained proliferative signaling, as once the tyrosine kinase function is activated, this membrane receptor can activate Ras/Raf/MAPK (Ras–mitogen-activated protein kinase), JAK/Stat and PI3K/AKT/mTOR pathways, involved in cell growth, survival, proliferation, division, metabolism, apoptosis and migration (Figure 2.4). Therefore, HER-2 positive breast cancer types are the most aggressive[60].

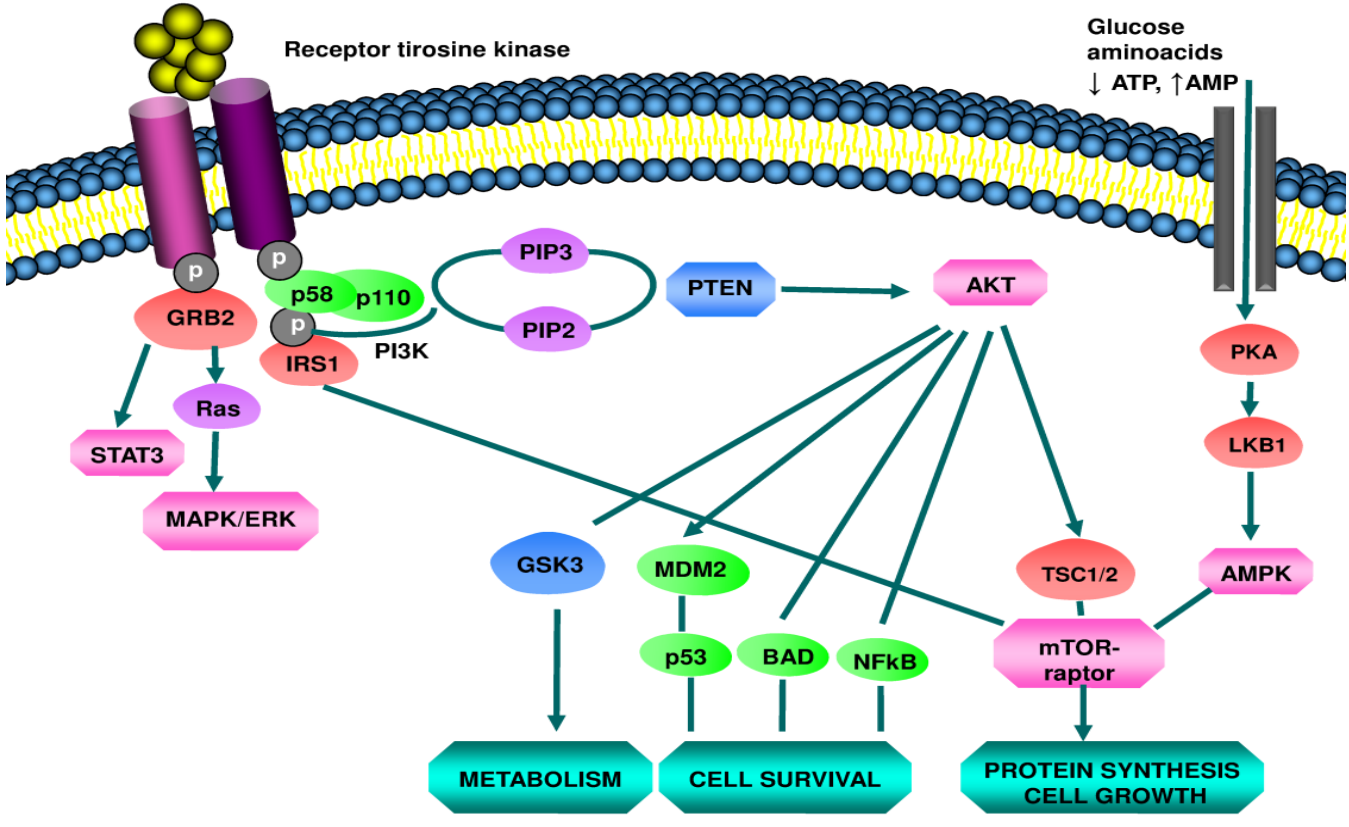


Figure 2.4. Pathway associated to the activation of the membrane tyrosine kinase receptors (RTK), such as the HER family and IGF1R (Adapted from Eroles et al.[60]).

The ER plays a fundamental role in the occurrence and development of breast cancer. Acting in two ways genomic and non-genomic, leading to activation of gene families involved in the onset of tumorigenesis, such as: Fos/Jun[64]. The ligand-mediated activation of the ER leads the activation of pathways of kinases, for example MEK/MAPK and PI3K/AKT[65], producing a double result: first, the cell cycle progression and survival are stimulated due to the activation of their regulators, and second, the phosphorylation of the ER, becoming more active at the genomic level[64]. Finally, the pathways implicated in cell growth and metabolic regulation, such as IGF1R/PI3K/AKT/mTOR, are important in breast cancer to obtain a constant source of energy for the sustained proliferative capacity of the neoplastic cell[60]. Crosstalk between ER and growth factors signaling pathways is shown in Figure 2.5.

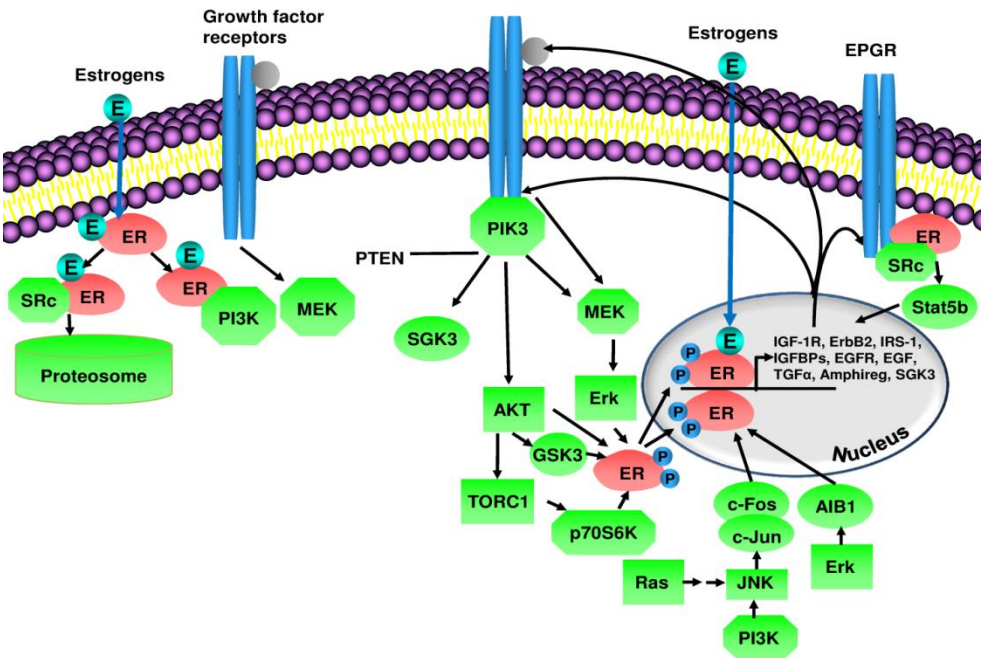


Figure 2.5. Crosstalk between ER and growth factor receptor signaling pathways (Adapted from Di Leo et al. [66]).

2.5. TECHNIQUES USED IN THIS PROJECT

The scientific techniques used in this thesis will be presented in the following sections; these include docking, protein expression and purification, microscale thermophoresis, and circular dichroism.

2.5.1. Docking

This is an *in silico* procedure to predict the non-covalent binding between two molecules, normally a protein and a low molecular weight ligand, based on their structure (Figure 2.6)[67]. The information derived from this process normally allows the prediction of the amino acids surrounding the binding site that interact with the docked molecule[68], and in some cases, the agonistic or antagonistic behavior of the ligands can be determined according to the binding site and conformational changes[69]. AutoDock/Vina, GOLD, FlexX and FRED are some of the most used programs in this field[70]. There are three main categories of scoring functions. The methods based on knowledge use the Boltzmann-weighted mean eld derived from the statistical analysis of inter-atomic contacts in the ligand-receptor complexes supported by available data in Protein Data Bank (PDB). The second is based on the master equation which estimates the energy contributions of various types of interactions in a semiquantitative manner. The third one is based on regression methods that have biological activity data available for performing training groups of receptor-ligand pairs; and the forth, the Poisson-Boltzman equation solvers that address electrostatics and incorporate solvent effects[71]. There are different approaches of docking studies that can be used. Some of them are vHTS that allows the evaluation of large libraries of small molecules against the proteins of interest[72], and the inverse virtual screening that predicts possible protein targets for a specific ligand[73].

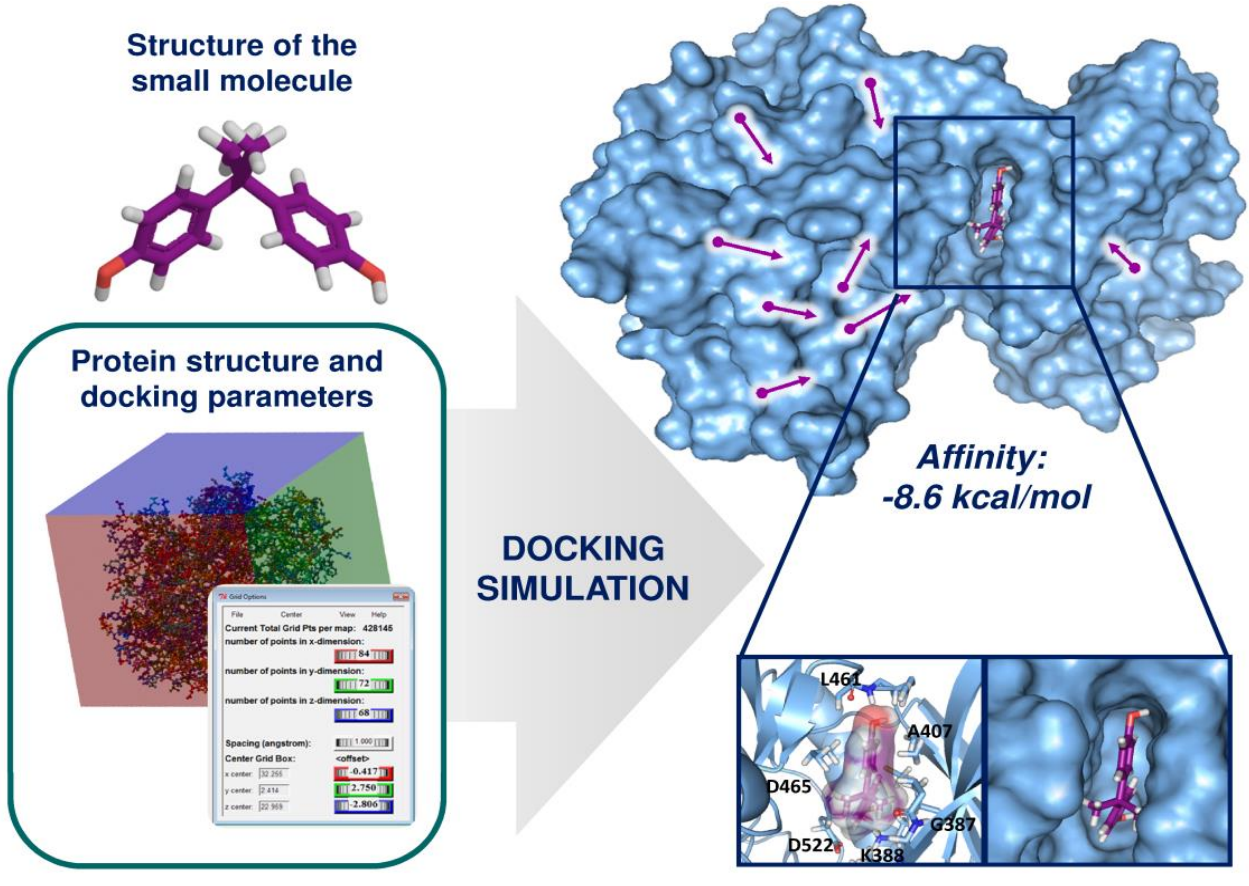


Figure 2.6. Docking simulation.

2.5.2. Protein expression and purification

The expression and purification of recombinant proteins are important tools in molecular biology[74]. DNA technology is generally used for these purposes. It involves altering the genetic machinery of an organism[75], mainly employed for the production of proteins[76]. Basically, the recombinant DNA is formed by introducing a DNA fragment of interest within a molecule called vector that is able to replicate independently, once the recombinant molecule is obtained, this is inserted into an appropriate host, one of the most used is *Escherichia coli*. The bacteria colonies having the plasmid are normally selected by an antibiotic resistance gene also incorporated in the recombinant DNA[77], and the protein of interest is expressed by growing them[78]. This can be purified by various methods, including one of the most widely used, affinity chromatography[79]. However there are other alternatives such as Dynabeads, which is a new technique that uses magnetic particles which enable the specific binding to polyhistidine-tagged proteins and their posterior release to purify them in native state[80].

2.5.3. Microscale thermophoresis (MST)

Microscale thermophoresis (MST) allows the analysis of protein interactions with low sample consumption, based on thermophoresis -the directed motion of molecules in temperature gradients[81]- which allows the determination of protein-ligand binding measured by the induced changes in the movement of molecules along a temperature gradient produced during this process[82]. This technique have gained its importance because of its capability to evaluate the interactions of proteins or small molecules in biological liquids such as blood serum or cell lysate[83].

2.5.4. Circular dichroism (CD)

Circular dichroism (CD) is an excellent tool for rapid determination of the secondary structures of proteins[84]. This measures the difference in absorption between right and left circularly polarized light when an asymmetric molecule is in the trajectory of these forms of light[85]. CD can be used to establish the conformational changes made by a protein after its binding with an small molecule, as the macromolecule undergoes a series of perturbations

and rearrangements that are observed in comparison to the protein and the protein-ligand complex CD spectra[86].

REFERENCES

1. (a) Lee, H.-R.; Hwang, K.-A.; Nam, K.-H.; Kim, H.-C.; Choi, K.-C., Progression of breast cancer cells was enhanced by endocrine-disrupting chemicals, triclosan and octylphenol, via an estrogen receptor-dependent signaling pathway in cellular and mouse xenograft models. Chem. Res. Toxicol. **2014,** 27 (5), 834-842; (b) Knower, K. C.; To, S. Q.; Leung, Y.-K.; Ho, S.-M.; Clyne, C. D., Endocrine disruption of the epigenome: a breast cancer link. Endocr. Relat. Cancer **2014,** 21 (2), T33-T55; (c) Martos, S. N.; Tang, W.-y.; Wang, Z., Elusive inheritance: Transgenerational effects and epigenetic inheritance in human environmental disease. Prog. Biophys. Mol. Biol. **2015,** 118 (1–2), 44-54; (d) Anway, M. D.; Cupp, A. S.; Uzumcu, M.; Skinner, M. K., Epigenetic Transgenerational Actions of Endocrine Disruptors and Male Fertility. Science **2005,** 308 (5727), 1466-1469; (e) Izzotti, A.; Kanitz, S.; D'Agostini, F.; Camoirano, A.; De Flora, S., Formation of adducts by bisphenol A, an endocrine disruptor, in DNA *in vitro* and in liver and mammary tissue of mice. Mutat. Res. **2009,** 679 (1–2), 28-32.
2. Jung, H.; Hong, Y.; Lee, D.; Pang, K.; Kim, Y., The association between some endocrine disruptors in human plasma and the occurrence of congenital hypothyroidism. Environ. Toxicol. Pharmacol. **2013,** 35 (2), 278-283.
3. (a) Sharan, S.; Nikhil, K.; Roy, P., Effects of low dose treatment of tributyltin on the regulation of estrogen receptor functions in MCF-7 cells. Toxicol. Appl. Pharmacol. **2013,** 269 (2), 176-186; (b) Macon, M. B.; Fenton, S. E., Endocrine disruptors and the breast: early life effects and later life disease. J. Mammary Gland Biol. **2013,** 18 (1), 43-61; (c) Scholz, S.; Renner, P.; Belanger, S. E.; Busquet, F.; Davi, R.; Demeneix, B. A.; Denny, J. S.; Leonard, M.; McMaster, M. E.; Villeneuve, D. L., Alternatives to *in vivo* tests to detect endocrine disrupting chemicals (EDCs) in fish and amphibians–screening for estrogen, androgen and thyroid hormone disruption. Critical reviews in toxicology **2013,** 43 (1), 45-72.
4. Futran Fuhrman, V.; Tal, A.; Arnon, S., Why endocrine disrupting chemicals (EDCs) challenge traditional risk assessment and how to respond. J. Hazard. Mater. **2015,** 286 (0), 589-611.
5. Kumar, A.; Xagoraraki, I., Pharmaceuticals, personal care products and endocrine-disrupting chemicals in U.S. surface and finished drinking waters: A proposed ranking system. Sci. Total Environ. **2010,** 408 (23), 5972-5989.
6. Gore, A.; Balthazart, J.; Bikle, D.; Carpenter, D.; Crews, D.; Czernichow, P.; Diamanti-Kandarakis, E.; Dores, R.; Grattan, D.; Hof, P., Reprint of: Policy decisions on endocrine disruptors should be based on science across disciplines: A response to Dietrich, et al. Front. Neuroendocrinol. **2014,** 35 (1), 2.

7. Gutendorf, B.; Westendorf, J., Comparison of an array of *in vitro* assays for the assessment of the estrogenic potential of natural and synthetic estrogens, phytoestrogens and xenoestrogens. Toxicology **2001,** 166 (1–2), 79-89.
8. (a) Xi, Y.; Li, D.; San, W., Exposure to the endocrine disruptor nonylphenol alters structure and function of thyroid gland in rats. Regul. Pept. **2013,** 185, 52-56; (b) Prins, G. S., Endocrine disruptors and prostate cancer risk. Endocr. Relat. Cancer **2008,** 15 (3), 649-656; (c) Bellanger, M.; Demeneix, B.; Grandjean, P.; Zoeller, R. T.; Trasande, L., Neurobehavioral Deficits, Diseases, and Associated Costs of Exposure to Endocrine-Disrupting Chemicals in the European Union. J. Clin. Endocrinol. Metab. **2015,** 100 (4), 1256-1266; (d) Hauser, R.; Skakkebaek, N. E.; Hass, U.; Toppari, J.; Juul, A.; Andersson, A. M.; Kortenkamp, A.; Heindel, J. J.; Trasande, L., Male Reproductive Disorders, Diseases, and Costs of Exposure to Endocrine-Disrupting Chemicals in the European Union. J. Clin. Endocrinol. Metab. **2015,** 100 (4), 1267-1277; (e) Diamanti-Kandarakis, E.; Bourguignon, J.-P.; Giudice, L. C.; Hauser, R.; Prins, G. S.; Soto, A. M.; Zoeller, R. T.; Gore, A. C., Endocrine-Disrupting Chemicals: An Endocrine Society Scientific Statement. Endocr. Rev. **2009,** 30 (4), 293-342.
9. (a) Patisaul, H. B., Effects of environmental endocrine disruptors and phytoestrogens on the kisspeptin system. In Kisspeptin Signaling in Reproductive Biology, Springer: 2013; pp 455-479; (b) Addis, A.; Moretti, M. E.; Ahmed Syed, F.; Einarson, T. R.; Koren, G., Fetal effects of cocaine: an updated meta-analysis. Reprod. Toxicol. **2001,** 15 (4), 341-369.
10. Stasinakis, A. S.; Gatidou, G.; Mamais, D.; Thomaidis, N. S.; Lekkas, T. D., Occurrence and fate of endocrine disrupters in Greek sewage treatment plants. Water Res. **2008,** 42 (6–7), 1796-1804.
11. Rousselle, C.; Ormsby, J.; Schaefer, B.; Lampen, A.; Platzek, T.; Hirsch-Ernst, K.; Warholm, M.; Oskarsson, A.; Nielsen, P.; Holmer, M., Meeting report: International workshop on endocrine disruptors: Exposure and potential impact on consumers health. Regul. Toxicol. Pharmacol. **2013,** 65 (1), 7-11.
12. Meeker, J. D., Exposure to environmental endocrine disrupting compounds and men's health. Maturitas **2010,** 66 (3), 236-241.
13. Birnbaum, L. S., State of the science of endocrine disruptors. Environ. Health Perspect. **2013,** 121 (4), a107.
14. Stojić, N.; Erić, S.; Kuzmanovski, I., Prediction of toxicity and data exploratory analysis of estrogen-active endocrine disruptors using counter-propagation artificial neural networks. J. Mol. Graph. Model. **2010,** 29 (3), 450-460.
15. de Wit, C. A., An overview of brominated flame retardants in the environment. Chemosphere **2002,** 46 (5), 583-624.
16. Watanabe, I.; Sakai, S.-i., Environmental release and behavior of brominated flame retardants. Environ. Int. **2003,** 29 (6), 665-682.
17. Legler, J.; Brouwer, A., Are brominated flame retardants endocrine disruptors? Environ. Int. **2003,** 29 (6), 879-885.
18. Newton, S.; Sellström, U.; de Wit, C. A., Emerging Flame Retardants, PBDEs, and HBCDDs in Indoor and Outdoor Media in Stockholm, Sweden. Environ. Sci. Technol. **2015,** 49 (5), 2912-2920.

19. Safe, S., Polychlorinated Biphenyls (PCBs), Dibenzo-p-Dioxins (PCDDs), Dibenzofurans (PCDFs), and Related Compounds: Environmental and Mechanistic Considerations Which Support the Development of Toxic Equivalency Factors (TEFs). Crit. Rev. Toxicol. **1990,** 21 (1), 51-88.
20. Arrebola, J.; Mutch, E.; Cuellar, M.; Quevedo, M.; Claure, E.; Mejía, L.; Fernández-Rodríguez, M.; Freire, C.; Olea, N.; Mercado, L., Factors influencing combined exposure to three indicator polychlorinated biphenyls in an adult cohort from Bolivia. Environ. Res. **2012,** 116, 17-25.
21. Korell, J.; Paur, H.-R.; Seifert, H.; Andersson, S., Simultaneous removal of mercury, PCDD/F, and fine particles from flue gas. Environ. Sci. Technol. **2009,** 43 (21), 8308-8314.
22. Lanting, C. I.; Patandin, S.; Fidler, V.; Weisglas-Kuperus, N.; Sauer, P. J. J.; Boersma, E. R.; Touwen, B. C. L., Neurological condition in 42-month-old children in relation to pre- and postnatal exposure to polychlorinated biphenyls and dioxins. Early Hum. Dev. **1998,** 50 (3), 283-292.
23. Chandra, R.; Chaudhary, S., Persistent organic pollutants in environment and their health hazards. IJBio. **2013,** 2 (09), 1232-1238.
24. Takayanagi, S.; Tokunaga, T.; Liu, X.; Okada, H.; Matsushima, A.; Shimohigashi, Y., Endocrine disruptor bisphenol A strongly binds to human estrogen-related receptor γ (ERRγ) with high constitutive activity. Toxicol. lett. **2006,** 167 (2), 95-105.
25. Rosenmai, A. K.; Dybdahl, M.; Pedersen, M.; Alice van Vugt-Lussenburg, B. M.; Wedebye, E. B.; Taxvig, C.; Vinggaard, A. M., Are Structural Analogues to Bisphenol A Safe Alternatives? Toxicol. Sci. **2014,** 139 (1), 35-47.
26. Manikkam, M.; Tracey, R.; Guerrero-Bosagna, C.; Skinner, M. K., Plastics Derived Endocrine Disruptors (BPA, DEHP and DBP) Induce Epigenetic Transgenerational Inheritance of Obesity, Reproductive Disease and Sperm Epimutations. PLoS ONE **2013,** 8 (1), e55387.
27. Schug, T. T.; Janesick, A.; Blumberg, B.; Heindel, J. J., Endocrine disrupting chemicals and disease susceptibility. The Journal of steroid biochemistry and molecular biology **2011,** 127 (3), 204-215.
28. Davis, L. K.; Visitacion, N.; Riley, L. G.; Hiramatsu, N.; Sullivan, C. V.; Hirano, T.; Grau, E. G., Effects of o, p'-DDE, heptachlor, and 17ß-estradiol on vitellogenin gene expression and the growth hormone/insulin-like growth factor-I axis in the tilapia, Oreochromis mossambicus. Comp. Biochem. Physiol. C. Toxicol. Pharmacol. **2009,** 149 (4), 507-514.
29. Birkett, J. W.; Lester, J. N., Endocrine Disrupters in Wastewater and Sludge Treatment Processes. Taylor & Francis: 2002.
30. Ahn, H.-J.; An, B.-S.; Jung, E.-M.; Yang, H.; Choi, K.-C.; Jeung, E.-B., Parabens inhibit the early phase of folliculogenesis and steroidogenesis in the ovaries of neonatal rats. Mol. Reprod. Dev. **2012,** 79 (9), 626-636.
31. Chen, B.-S.; Yen, J.-H., Effect of endocrine disruptor nonylphenol on physiologic features and proteome during growth in Arabidopsis thaliana. Chemosphere **2013,** 91 (4), 468-474.
32. Ishibashi, H.; Matsumura, N.; Hirano, M.; Matsuoka, M.; Shiratsuchi, H.; Ishibashi, Y.; Takao, Y.; Arizono, K., Effects of triclosan on the early life stages and reproduction of

medaka Oryzias latipes and induction of hepatic vitellogenin. Aquat. Toxicol. **2004,** 67 (2), 167-179.

33. Umland, E. M.; Cauffield, J. S.; Kirk, J. K.; Thomason, T. E., Phytoestrogens as therapeutic alternatives to traditional hormone replacement in postmenopausal women. Pharmacotherapy **2000,** 20 (8), 981-990.
34. Eden, J. A., Phytoestrogens for menopausal symptoms: A review. Maturitas **2012,** 72 (2), 157-159.
35. Bratton, M. R.; Martin, E. C.; Elliott, S.; Rhodes, L. V.; Collins-Burow, B. M.; McLachlan, J. A.; Wiese, T. E.; Boue, S. M.; Burow, M. E., Glyceollin, a novel regulator of mTOR/p70S6 in estrogen receptor positive breast cancer. The Journal of Steroid Biochemistry and Molecular Biology **2015,** 150, 17-23.
36. Cassidy, A.; Albertazzi, P.; Nielsen, I. L.; Hall, W.; Williamson, G.; Tetens, I.; Atkins, S.; Cross, H.; Manios, Y.; Wolk, A.; Steiner, C.; Branca, F., Critical review of health effects of soyabean phyto-oestrogens in post-menopausal women. Proc. Nutr. Soc. **2006,** 65 (01), 76-92

M3 - 10.1079/PNS2005476.

37. (a) Anderson, L. N.; Cotterchio, M.; Boucher, B. A.; Kreiger, N., Phytoestrogen intake from foods, during adolescence and adulthood, and risk of breast cancer by estrogen and progesterone receptor tumor subgroup among Ontario women. International Journal of Cancer **2013,** 132 (7), 1683-1692; (b) Salvo, V. A.; Boué, S. M.; Fonseca, J. P.; Elliott, S.; Corbitt, C.; Collins-Burow, B. M.; Curiel, T. J.; Srivastav, S. K.; Shih, B. Y.; Carter-Wientjes, C.; Wood, C. E.; Erhardt, P. W.; Beckman, B. S.; McLachlan, J. A.; Cleveland, T. E.; Burow, M. E., Antiestrogenic Glyceollins Suppress Human Breast and Ovarian Carcinoma Tumorigenesis. Clinical Cancer Research **2006,** 12 (23), 7159-7164.
38. Zaineddin, A. K.; Buck, K.; Vrieling, A.; Heinz, J.; Flesch-Janys, D.; Linseisen, J.; Chang-Claude, J., The Association Between Dietary Lignans, Phytoestrogen-Rich Foods, and Fiber Intake and Postmenopausal Breast Cancer Risk: A German Case-Control Study. Nutrition and Cancer **2012,** 64 (5), 652-665.
39. Rice, S.; Whitehead, S., Targets for the Action of Phytoestrogens in Breast Cancer—Focus on Isoflavones and Resveratrol. Curr Breast Cancer Rep **2014,** 6 (2), 88-95.
40. Hunter, D. J.; Kraft, P.; Jacobs, K. B.; Cox, D. G.; Yeager, M.; Hankinson, S. E.; Wacholder, S.; Wang, Z.; Welch, R.; Hutchinson, A.; Wang, J.; Yu, K.; Chatterjee, N.; Orr, N.; Willett, W. C.; Colditz, G. A.; Ziegler, R. G.; Berg, C. D.; Buys, S. S.; McCarty, C. A.; Feigelson, H. S.; Calle, E. E.; Thun, M. J.; Hayes, R. B.; Tucker, M.; Gerhard, D. S.; Fraumeni, J. F.; Hoover, R. N.; Thomas, G.; Chanock, S. J., A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet **2007,** 39 (7), 870-874.
41. Bosviel, R.; Dumollard, E.; Déchelotte, P.; Bignon, Y.-J.; Bernard-Gallon, D., Can Soy Phytoestrogens Decrease DNA Methylation in BRCA1 and BRCA2 Oncosuppressor Genes in Breast Cancer? OMICS: A Journal of Integrative Biology **2012,** 16 (5), 235-244.
42. Tong, L.; Chuang, C.-C.; Wu, S.; Zuo, L., Reactive oxygen species in redox cancer therapy. Cancer Letters **2015**.

43. Sung, M.-K.; Bae, Y.-J., Chapter 13 - Iron, Oxidative Stress, and Cancer. In Cancer, Preedy, V., Ed. Academic Press: San Diego, 2014; pp 139-149.
44. Hartmann, N. Environmental exposure to estrogenic mycotoxins. Diss., Umweltnaturwissenschaften, Eidgenössische Technische Hochschule ETH Zürich, Nr. 17751, 2008, 2008.
45. Grosse, Y.; Chekir-Ghedira, L.; Huc, A.; Obrecht-Pflumio, S.; Dirheimer, G.; Bacha, H.; Pfohl-Leszkowicz, A., Retinol, ascorbic acid and α-tocopherol prevent DNA adduct formation in mice treated with the mycotoxins ochratoxin A and zearalenone. Cancer Lett. **1997,** 114 (1), 225-229.
46. Barouki, R.; Gluckman, P. D.; Grandjean, P.; Hanson, M.; Heindel, J. J., Developmental origins of non-communicable disease: implications for research and public health. Environ. Health **2012,** 11 (42), 10-1186.
47. Bergman, A.; Heindel, J. J.; Kasten, T.; Kidd, K. A.; Jobling, S.; Neira, M.; Zoeller, R. T.; Becher, G.; Bjerregaard, P.; Bornman, R., The impact of endocrine disruption a consensus statement on the state of the science. Environ. Health Perspect. **2013,** 121 (4), A104.
48. Brisken, C., Endocrine disruptors and breast cancer. CHIMIA Int. J. Chem. **2008,** 62 (5), 406-409.
49. Safea, S., Endocrine disruptors and human health: is there a problem. Toxicology **2004,** 205 (1), 3-10.
50. Yang, M.; Ryu, J.-H.; Jeon, R.; Kang, D.; Yoo, K.-Y., Effects of bisphenol A on breast cancer and its risk factors. Arch Toxicol **2009,** 83 (3), 281-285.
51. Takaku, M.; Adomas, A.; Grimm, S. A.; Takashi, S.; Wade, P. A., GATA3 mutations in breast cancer. Cancer Res. **2014,** 74 (19 Supplement), 475-475.
52. Izzotti, A.; Kanitz, S.; D'Agostini, F.; Camoirano, A.; De Flora, S., Formation of adducts by bisphenol A, an endocrine disruptor, in DNA *in vitro* and in liver and mammary tissue of mice. Mutat. Res. Genet. Toxicol. Environ. Mutagen. **2009,** 679 (1–2), 28-32.
53. Darbre, P. D.; Fernandez, M. F., Environmental oestrogens and breast cancer: long-term low-dose effects of mixtures of various chemical combinations. J. Epidemiol. Community Health **2012**.
54. Hopp, T. A.; Weiss, H. L.; Parra, I. S.; Cui, Y.; Osborne, C. K.; Fuqua, S. A., Low levels of estrogen receptor ß protein predict resistance to tamoxifen therapy in breast cancer. Clin. Cancer Res. **2004,** 10 (22), 7490-7499.
55. Nwozo, S. O.; Solomon, O.; Abimbola, O. O.; Kikelomo, D. O., Comparative study of biochemical and nutritional status of breast cancer patients on chemotherapy/radiotherapy in ibadan. Am. J. Cancer. **2013,** 2 (1), 51-60.
56. Bidgoli, S. A.; Ahmadi, R.; Zavarhei, M. D., Role of hormonal and environmental factors on early incidence of breast cancer in Iran. Sci. Total Environ. **2010,** 408 (19), 4056-4061.
57. Han, D.; Nie, J.; Bonner, M.; Ambrosone, C.; Marian, C.; Shields, P.; Trevisan, M.; Edge, S.; Freudenheim, J., Clustering of place of birth for women with breast cancer: differences by tumor characteristics. Cancer Cause. Control **2013,** 24 (3), 587-594.
58. Weigelt, B.; Geyer, F. C.; Reis-Filho, J. S., Histological types of breast cancer: How special are they? Mol. Oncol. **2010,** 4 (3), 192-208.

59. Carey, L. A.; Perou, C. M.; Livasy, C. A.; et al., Race, breast cancer subtypes, and survival in the carolina breast cancer study. JAMA **2006,** 295 (21), 2492-2502.
60. Eroles, P.; Bosch, A.; Alejandro Pérez-Fidalgo, J.; Lluch, A., Molecular biology in breast cancer: Intrinsic subtypes and signaling pathways. Cancer. Treat. Rev. **2012,** 38 (6), 698-707.
61. Das, P.; Siegers, G. M.; Postovit, L.-M., Illuminating luminal B: QSOX1 as a subtype-specific biomarker. Breast Cancer Res. **2013,** 15 (3), 104.
62. Vaca-Paniagua, F.; Alvarez-Gomez, R.; Maldonado-Martínez, H.; Pérez-Plasencia, C.; Fragoso-Ontiveros, V.; Lasa-Gonsebatt, F.; Herrera, L.; Cantú, D.; Bargallo-Rocha, E.; Mohar, A., Revealing the Molecular Portrait of Triple Negative Breast Tumors in an Understudied Population through Omics Analysis of Formalin-Fixed and Paraffin-Embedded Tissues. PloS one **2015,** 10 (5), e0126762.
63. Hanahan, D.; Weinberg, Robert A., Hallmarks of Cancer: The Next Generation. Cell **2011,** 144 (5), 646-674.
64. Giancotti, V., Breast cancer markers. Cancer Lett. **2006,** 243 (2), 145-159.
65. Chen, Y.; Guggisberg, N.; Jorda, M.; Gonzalez-Angulo, A.; Hennessy, B.; Mills, G. B.; Tan, C.-K.; Slingerland, J. M., Combined Src and aromatase inhibition impairs human breast cancer growth *in vivo* and bypass pathways are activated in AZD0530-resistant tumors. Clin. Cancer Res. **2009,** 15 (10), 3396-3405.
66. Di Leo, A.; Curigliano, G.; Diéras, V.; Malorni, L.; Sotiriou, C.; Swanton, C.; Thompson, A.; Tutt, A.; Piccart, M., New approaches for improving outcomes in breast cancer in Europe. Breast **2015,** 24 (4), 321-330.
67. Trott, O.; Olson, A. J., AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. **2010,** 31 (2), 455-461.
68. Wolber, G.; Langer, T., LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. J. Chem. Inf. Model. **2005,** 45 (1), 160-169.
69. Nose, T.; Tokunaga, T.; Shimohigashi, Y., Exploration of endocrine-disrupting chemicals on estrogen receptor α by the agonist/antagonist differential-docking screening (AADS) method: 4-(1-Adamantyl)phenol as a potent endocrine disruptor candidate. Toxicol. Letters **2009,** 191 (1), 33-39.
70. Azam, S. S.; Abbasi, S. W., Molecular docking studies for the identification of novel melatoninergic inhibitors for acetylserotonin-O-methyltransferase using different docking routines. Theor. Biol. Med. Model. **2013,** 10, 63-63.
71. Gad, S. C., Drug Discovery Handbook. Wiley: 2005.
72. Seifert, M. H. J.; Wolf, K.; Vitt, D., Virtual high-throughput *in silico* screening. Biosilico **2003,** 1 (4), 143-149.
73. Scrima, M.; Lauro, G.; Grimaldi, M.; Di Marino, S.; Tosco, A.; Picardi, P.; Gazzerro, P.; Riccio, R.; Novellino, E.; Bifulco, M., Structural Evidence of N 6-Isopentenyladenosine As a New Ligand of Farnesyl Pyrophosphate Synthase. J. Med. Chem. **2014,** 57 (18), 7798-7803.
74. Castro, J. E. Z., Manual de Técnicas Básicas de Biología Molecular. UADY: 2005.

75. Sandhu, S. S., Recombinant DNA Technology. I.K. International Publishing House Pvt., Limited: 2010.
76. Mahesh, S.; Vedamurthy, A. B., Biotechnology-4: Including Recombinant DNA Technology, Environmental Biotechnology, Animal Cell Culture. New Age International (P) Limited: 2003.
77. Mazzotta, G. C., Identificación Por ADN. Ediciones Jurídicas Cuyo: 2000.
78. Vallejo, F. G., El nómada molecular: la historia molecular del virus linfotrópico humano tipo (HTLV-1). Universidad del Valle: 2004.
79. Hausman, G.; Cooper, G. M.; Hausman, R. E., The Cell: A Molecular Approach. Sinauer Associates Incorporated: 2009.
80. Corchero, J. L.; Villaverde, A., Biomedical applications of distally controlled magnetic nanoparticles. Trends Biotechnol. **2009,** 27 (8), 468-476.
81. Seidel, S. A. I.; Dijkman, P. M.; Lea, W. A.; van den Bogaart, G.; Jerabek-Willemsen, M.; Lazic, A.; Joseph, J. S.; Srinivasan, P.; Baaske, P.; Simeonov, A.; Katritch, I.; Melo, F. A.; Ladbury, J. E.; Schreiber, G.; Watts, A.; Braun, D.; Duhr, S., Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. Methods **2013,** 59 (3), 301-315.
82. Corin, K.; Baaske, P.; Ravel, D. B.; Song, J.; Brown, E.; Wang, X.; Geissler, S.; Wienken, C. J.; Jerabek-Willemsen, M.; Duhr, S.; Braun, D.; Zhang, S., A Robust and Rapid Method of Producing Soluble, Stable, and Functional G-Protein Coupled Receptors. PLoS ONE **2011,** 6 (10), e23036.
83. Wienken, C. J.; Baaske, P.; Rothbauer, U.; Braun, D.; Duhr, S., Protein-binding assays in biological liquids using microscale thermophoresis. Nat. Commun. **2010,** 1, 100.
84. Whitmore, L.; Woollett, B.; Miles, A. J.; Klose, D. P.; Janes, R. W.; Wallace, B. A., PCDDB: the protein circular dichroism data bank, a repository for circular dichroism spectral and metadata. Nucleic Acids Res. **2011,** 39 (suppl 1), D480-D486.
85. McReynolds, K. D.; Gervay-Hague, J., Examining the secondary structures of unnatural peptides and carbohydrate-based compounds utilizing circular dichroism. Tetrahedron Asymmetry **2000,** 11 (2), 337-362.
86. Rodger, A.; Marrington, R.; Roper, D.; Windsor, S., Circular Dichroism Spectroscopy for the Study of Protein-Ligand Interactions. In Protein-Ligand Interactions, Ulrich Nienhaus, G., Ed. Humana Press: 2005; Vol. 305, pp 343-363.

CHAPTER 3

EDCs DataBank

The initial phase of this thesis project was the creation of EDCs DataBank, which is the first database of xenoestrogens with three-dimensional structures available for virtual screening on the web (http://edcs.unicartagena.edu.co/). The main motivation for its building was to incentivize computational toxicological approaches in this field helpful to guide *in vitro* and *in vivo* studies, as well as to inform the community regarding the potential adverse effects of some consumer products and to systematically organize the existent key information about these pollutants for its easy use with educational and research purposes.

In the context of this database, EDCs are those molecules with potential to affect the endocrine system with reported data in the scientific literature. It includes a wide range of compounds such as insecticides, plasticizers and monomers, flame retardants, natural products, fragrances, industrial chemicals, organic pollutants and food contact materials, among others[1].

In this chapter, the whole building process of EDCs DataBank is described, as well as its features and usage. Finally, a statistical analysis was conducted to study the chemical space of this database and the main structural characteristics of the molecules stored on it, with exploratory purposes.

3.1. MATERIALS AND METHODS

EDCs DataBank was created from scratch. Therefore any website or database template was utilized, and it was developed using mainly open-source programs, under a Microsoft Windows Operating System environment.

3.1.1. Building of the internal library of EDCs

The names of the EDCs, as well as their CAS numbers, were obtained from the EU list of potential endocrine disruptors (http://www.mst.dk/English/Chemicals/endocrine_disruptors/the_EU_list_of_potential_endocrine_disruptors/) and the TEDX list (http://endocrinedisruption.org/endocrine-disruption/tedx-list-of-potential-endocrine-disruptors/overview), since the molecules registered in these repositories have been cited at least once as endocrine disruptors in scientific articles.

The coordinates and structural information of the EDCs with three-dimensional structures available in PubChem[2] were downloaded in 2D- and 3D-sdf formats. They were used for building an internal chemical library in Instant JChem (http://www.chemaxon.com), which was then saved as a single sdf file containing all the three-dimensional structures of the EDCs, as well as basic information of each one (Figure 3.1). The 3D-sdf files were also translated to mol2, pdb and pdbqt format files employing the Sybyl X-2.0 program package (Tripos, St. Louis, MO), Open Babel[3] and AutoDock tools (http://autodock.scripps.edu/resources/adt), respectively.

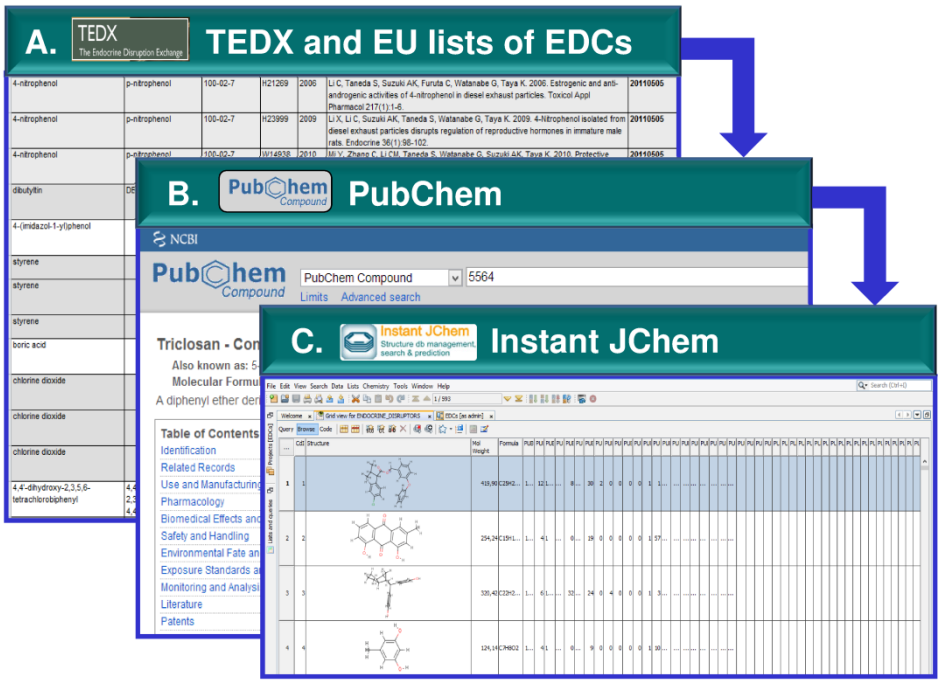


Figure 3.1. Workflow of the building of the internal library of EDCs, using (A) the TEDX list and EU list of potential endocrine disruptors, (B) PubChem[4] and (C) Instant JChem (http://www.chemaxon.com).

3.1.2. External links, physico-chemical and toxicological information

The internal database was enriched with a wide variety of information for each molecule, obtained from different databases and search engines, such as: PubChem[4], a database of chemical compounds and their activities against biological assays; TOXNET (Toxicology Data Network; http://toxnet.nlm.nih.gov/)[5], a group of databases covering different issues regarding the safety of chemicals and drugs; ACToR[6], a database with structural, toxicological and bioactivity information of environmental toxics derived from more than 150 sources of information including ToxCast[7]; Fable (fast automated biomedical literature extraction; http://fable.chop.edu), a tool

for finding human genes associated with a keyword and biomedical literature extraction; and PubMed (http://www.ncbi.nlm.nih.gov/pubmed).

EDCs DataBank has two links to PubMed, the first one is for a general search of articles containing the name of the compound, and the second one further includes 40 keywords related with endocrine disruption in the PubMed query as a filter to present only the articles associated with endocrine disruption. These terms were previously selected using the text mining tool PubMed PubReMiner (http://hgserver2.amc.nl/cgi-bin/miner/miner2.cgi). An example of a query search generated using this link for 3,5-dichlorobiphenyl is presented below:

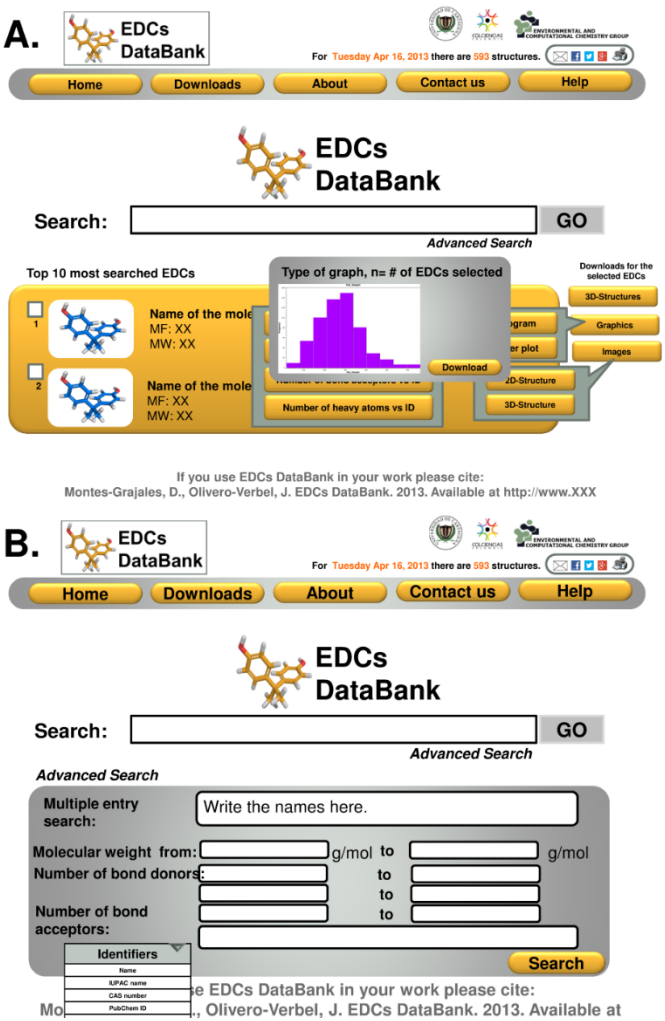
"3,5-dichlorobiphenyl" and ("androgen" or "androgenic" or "adrenergic" or "antiadrenergic" or "antiandrogen" or "antiandrogenic" or "antiestrogenic" or "antithyroid" or "aromatase" or "catecholamines" or "EDCs" or "endocrine" or "ERalpha" or "estradiol" or "estrogen" or "estrogenic" or "estrogenicity" or "fertility" or "glucocorticoid" or "hormone" or "proestrogenic" or "progesterone" or "progesteronic" or "reproductive" or "semen" or "sperm" or "spermatogenesis" or "steroids" or "thyroid" or "urogenital" or "xenoestrogen" or "xenoestrogenic" or "xenoestrogens" or "imposex" or "cancer" or "ovary" or "pregnancy" or "testosterone" or "uterus" or "tumor").

3.1.3. Graphics

High quality images of the 2D- and 3D-structures of each compound were prepared in Marvin Sketch (http://www.chemaxon.com) and PyMol (http://www.pymol.org/), using as input the obtained 2D- and 3D-sdf files. The colors were configured by atom type in order to facilitate their reading. In addition, the interactive tool for molecular visualization, Jmol (http://jmol.sourceforge.net/index.en.html/), was included on the EDCs DataSheets. These visual resources are useful for obtaining a general overview of the geometry of the molecules. In addition, icons of the different sources of EDCs were generated as a user-friendly tool that enables a quick identification of the potential places for greatest risk of exposure, helping to remember them for educational purposes.

3.1.4. Data schemes and database implementation

An initial design of the website of EDCs DataBank was generated (Figure 3.2). This was based on a list of requirements, subsequently analyzed, in terms of viability, and optimized during the programming process. Besides the technical aspects of a database for molecules, some of the selected features were directed to develop a user-friendly environment, such as readable font, appropriated contrasts and illustrations, as well as interactive images.



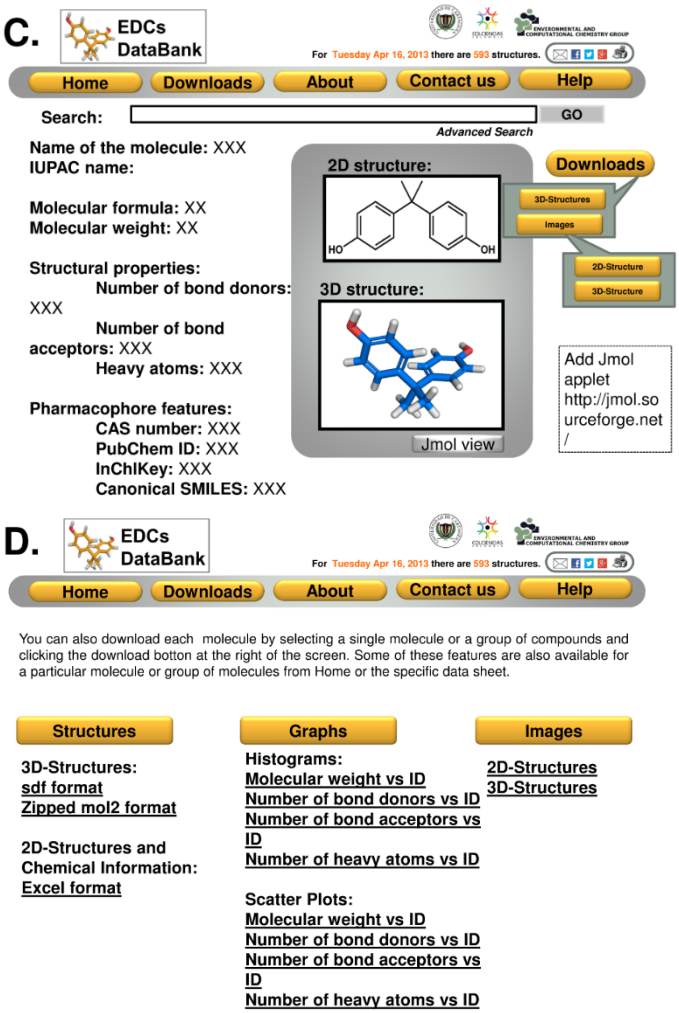


Figure 3.2. Initial design of the EDCs DataBank. (A) home, (B) advanced search, (C) EDCs datasheet and (D) downloads page.

In order to offer a structure for the EDCs DataSheets -pages containing the information for each EDC- in the web development and database management, the available data were organized in the following sections: name, synonyms, source, identifiers, structural properties, pharmacophore features, downloads, similarity search, toxicological information, external links and keywords, as shown in Table 3.1.

Table 3.1. Structure of the information contained in the EDCs DataSheets.

Field	Description
Name	The most common name for each molecule. These were obtained from the TEDX List (http://www.endocrinedisruption.com/endocrine.TEDXList.overview.php) and the EU List of Potential Endocrine Disruptors (http://www.mst.dk/English/Chemicals/endocrine_disruptors/the_EU_list_of_potential_endocrine_disruptors/).
Synonyms	Alternative names given to the compound (source: PubChem[4]).
Source	Exposure sources for each EDCs available on the internet.
Identifiers	IUPAC name, CAS number, PubChem ID, InChiKey and Canonical SMILES (simplified molecular input line entry specification; sources: TEDX List and PubChem[4]).
Structural properties	It includes molecular formula and molecular weight (source: PubChem[4]).
Pharmacophore features	Number of bond donors, bond acceptors and atoms different from hydrogen -heavy atoms- (source: PubChem[4]).
Downloads	2D-structure (2D-sdf file) and 3D-structures (3D-sdf file, mol2, pdb and pdbqt). Source: PubChem[4]; the formats mol2, pdb and pdbqt were obtaining by translation of the original sdf files.
Similarity search	It searches for similar molecules into the database. Source: The matrix was generated using the PubChem Score Matrix Service (http://pubchem.ncbi.nlm.nih.gov//score_matrix/score_matrix.cgi).
Toxicological Information	This is a resource that presents the available toxicological information in the databases ACToR[6], TEDX List, and those included TOX-NET[5], such as the Hazardous Substances Data Bank (HSDB), Chemical Carcinogenesis Research Information System (CCRIS), Genetic Toxicology Data Bank (GENE-TOX), Comparative Toxicogenomics Database (CTD), Integrated Risk Information System (IRIS), International Toxicity Estimates for Risk (ITER) and TEDX List.
External links	These are links to other databases and search engines, such as PubChem[4], PubMed (http://www.ncbi.nlm.nih.gov/pubmed), PhysProp (http://esc.syrres.com/fatepointer/search.asp) and Fable (http://fable.chop.edu).
Keywords	Words related to each EDCs. The words with the largest letters represent a major association. The keywords were obtained in the text mining tool LigerCat[8].

The EDCs DataBank website was developed from scratch using HTML (hypertext markup language), CSS (cascading style sheets), JQuery and JavaScript in a NotePad++ editor. The database was built on MySQL (my

structured query language), the most popular open source database, and then interconnected with the web environment using the PHP languages (PHP hypertext preprocessor) (Figure 3.3).

```
<?php include 'includes/header.php' ?>

        <div class="content">
            <div class="search">
                <?php include "includes/search.php"; ?>
            </div>
            <input type="hidden" value="<?php $img_var ?>" name="img_search"
            />

            <div id="list_wrapper">
                <div class="list_img_row">
                    <div class="img_holder">
                        <a href="result.php?img_var=acaricide"><img src=
                        "img/source_img/acaricide.png"></a>
                    </div>
                    <div class="img_holder">
                        <a href="result.php?img_var=analytical_chemistry"
                        ><img src="img/source_img/analytical_chemistry.png"
                        ></a>
                    </div>
                    <div class="img_holder">
                        <a href="result.php?img_var=antimicrobial"><img src=
                        "img/source_img/antimicrobial.png"></a>
                    </div>
                    <div class="img_holder">
                        <a href="result.php?img_var=antiseptics"><img src=
                        "img/source_img/antiseptics.png"></a>
                    </div>
```

Figure 3.3. Fragment of the PHP code of the EDCs DataBank website.

he architecture of EDCs DataBank is shown in Figure 3.4.

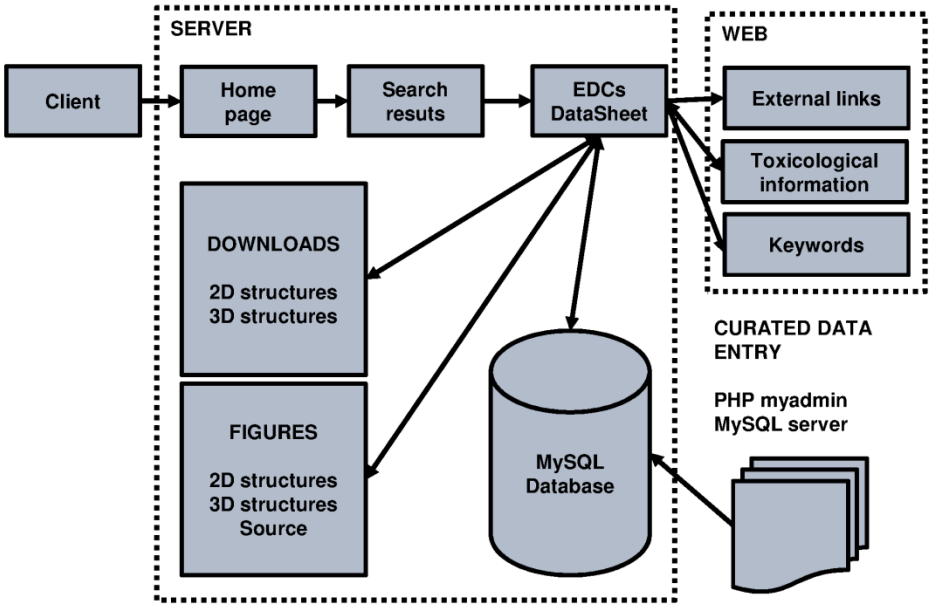


Figure 3.4. EDCs DataBank architecture.

3.1.5. Quality assurance, validation, completeness and curation

All the information has been hand cured in order to reduce errors as much as possible. The fields of this database have been completed based on all available information. Most of the structural data contained in EDCs DataBank was obtained directly from PubChem[4] and managed through Instant JChem (http://www.chemaxon.com), for its direct inclusion in the MySQL database of EDCs DataBank. The theoretical models of the three-dimensional structures were also obtained directly from PubChem[4] (https://pubchem.ncbi.nlm.nih.gov/release3d.html), leading the file names as default, and translated to several formats using standardized protocols in order to avoid typing errors.

All the information contained in each EDC DataSheet was carefully revised. The whole texts of each datasheet were revised three times by a Ph.D. student and an undergraduate student, the correct functioning of the external links and toxicological information *iframes* was also checked. The concordance between the names, 2D-structures, 3D-structures and Jmol interactive molecular viewer (http://jmol.sourceforge.net/index.en.html/) was examined; and each 2D- and 3D-structure was downloaded and the geometry was checked using pymol in order to be sure that the 3D-structures contain three-dimensional features, expected connectivity and normal bond lengths. These revisions will be performed systematically and regularly in order to keep the database updated and curated. Users are also encouraged to suggest improvements to the database through the "questions and comments" form, located in the *Contact us* option of the main menu.

The website of EDCs DataBanks is hosted as a subdomain of the University of Cartagena website (http://edcs.unicartagena.edu.co/), which contributes to its permanency and continuous development by the students of the Ph.D. Program in Environmental Toxicology of this University, working in the research line of endocrine disruptors and environmental pollutants. Additional information and new molecules will be included in the database, in accordance with the information available and the new discoveries in this field.

3.1.6. Chemical structural space and frequency distribution

A total of 1,600 descriptors were calculated in E-Dragon[9] for the 615 molecules contained in EDCs DataBank. The molecular descriptors were then ranked according to their Shannon entropy values, previously calculated in R version 3.0.2 (http://www.r-project.org/) using the library "entropy"[10]. The top 300 molecular descriptors where retained for principal component analysis (PCA). PCA is one of the most-widely used multivariate exploratory techniques for data classification and pattern recognition; it allows the detection of internal relationships between characteristics of a set of objects, thus enabling a drastic reduction of the dimensionality of the original raw data[11]. This analysis was carried out in R (http://www.r-project.org/), using the eigenvectors ("loadings"), the scores ("scores"), and the percent of

variation explained ("summary"). The first 3 principal components (PCs) were employed to compare the chemical structure space, which is defined by a cuboid in the three-dimensional space of the first 3 principal components[10]. Additional statistical calculations were also performed in JChem for Excel, to represent the frequency distribution of the data according to the number of bond donors, bond acceptors, heavy atoms (atoms different from hydrogen) and molecular weight.

3.2. RESULTS

3.2.1. Building of the internal library

The internal library was successfully developed. Currently EDCs DataBank (Figure 3.5) contains 615 molecules. This database allows the download of the whole 2D- and 3D- structures stored on it, in several formats (mol2, pdb, pdbqt, sdf). This can be executed from the download section of each EDC DataSheet or from the *download* option of the main menu. This possibility contributes to an easy use of these compounds in docking, QSAR and virtual screening studies, as well as other *in silico* approaches.

Figure 3.5. Logo of EDCs DataBank.

3.2.2. External links, physico-chemical and toxicological information

The EDCs belonging to EDCs DataBank include pesticides, flavonoids and medicinal plants, as well as compounds used in cosmetics, plastics and pharmaceutical industry, among others. Many of them also correspond to environmental pollutants, or are components of everyday goods such as household and personal care products. A broad diversity of data regarding each compound is available in the DataSheet of each molecule, organized in the following sections: name, synonyms, source, identifiers, structural properties, pharmacophore features, downloads, similarity search and keywords.

In addition, this database is hyperlinked to other important databases in the field of toxicology and literature search such as PubMed, PubChem[4], ACToR[6], LigerCat[8], PhysProp and TOXNET -This comprises the databases: HSDB (Hazardous Substances Data Bank), CCRIS (Chemical Carcinogenesis Research Information System), GENE-TOX (Genetic Toxicology Data Bank), CTD (Comparative Toxicogenomics Database), IRIS (Integrated Risk Information System) and ITER (International Toxicity Estimates for Risk)-. These links are present as a section on toxicological information, external links and keywords on the datasheet of each compound. The availability of hyperlinks regarding the toxicological information and physical properties depends upon the existence of the information in the source databases. The number of molecules in EDCs DataBank with available data in those fields is presented in Figure 3.6.

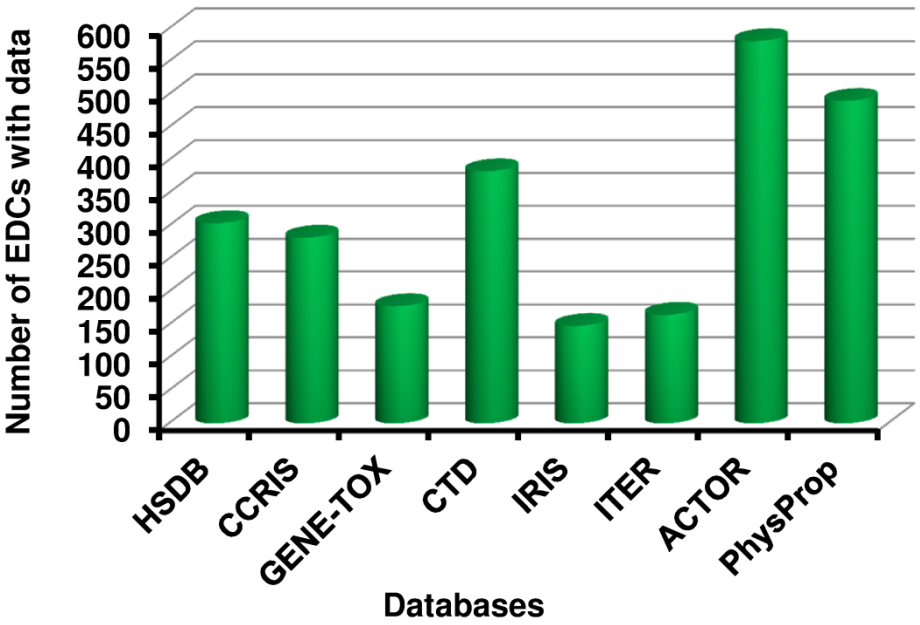


Figure 3.6. Number of molecules with toxicological and physicochemical data available in EDCs DataBank.

3.2.3. Graphics

Each EDC has images of its 2D- and 3D-structures in high-quality as well as an interactive Jmol applet, which allows the interaction with the three-dimensional structure and simple calculations by right-clicking on the image (an updated version of Java is required for its use). The 51 icons regarding the different sources of EDCs are also available as a search tool in the home page and for information in the EDCs DataSheets.

3.2.4. Data schemes and database implementation

3.2.4.1. EDCs DataBank description. EDCs DataBank is a unique database with three-dimensional structures of EDCs for virtual screening. It includes compounds utilized in pesticides, cosmetics, gasoline and food, health care and household products, among others. It is suitable for virtual screening because each molecule is presented in several formats (mol2, pdb, pdbqt, sdf), allowing the use of different computer aided tools for their analysis. Consequently, EDCs DataBank is a valuable repository to study the interactions of these compounds with macromolecules involved in several diseases, such as breast cancer, obesity and diabetes. The databank also benefits the community and students, since it contains possible sources of exposure to EDCs, toxicological information, and links to different databases in a user-friendly environment.

3.2.4.2. Website map. At the top of the page is the main menu of the EDCs DataBank. The user can find the search bar and access the advanced search. The website of the database possesses a *download* page in which the user can obtain the 3D structures of all molecules stored in EDCs DataBank in different formats, as well as images regarding statistical information of all compounds. This page also contains a questions and comments form in the *contact us* button, in which the user can communicate their inquiries and suggestions. The website includes an *about* and *help* links with detailed information regarding EDCs DataBank and its use. The map of the website is shown in Figure 3.7.

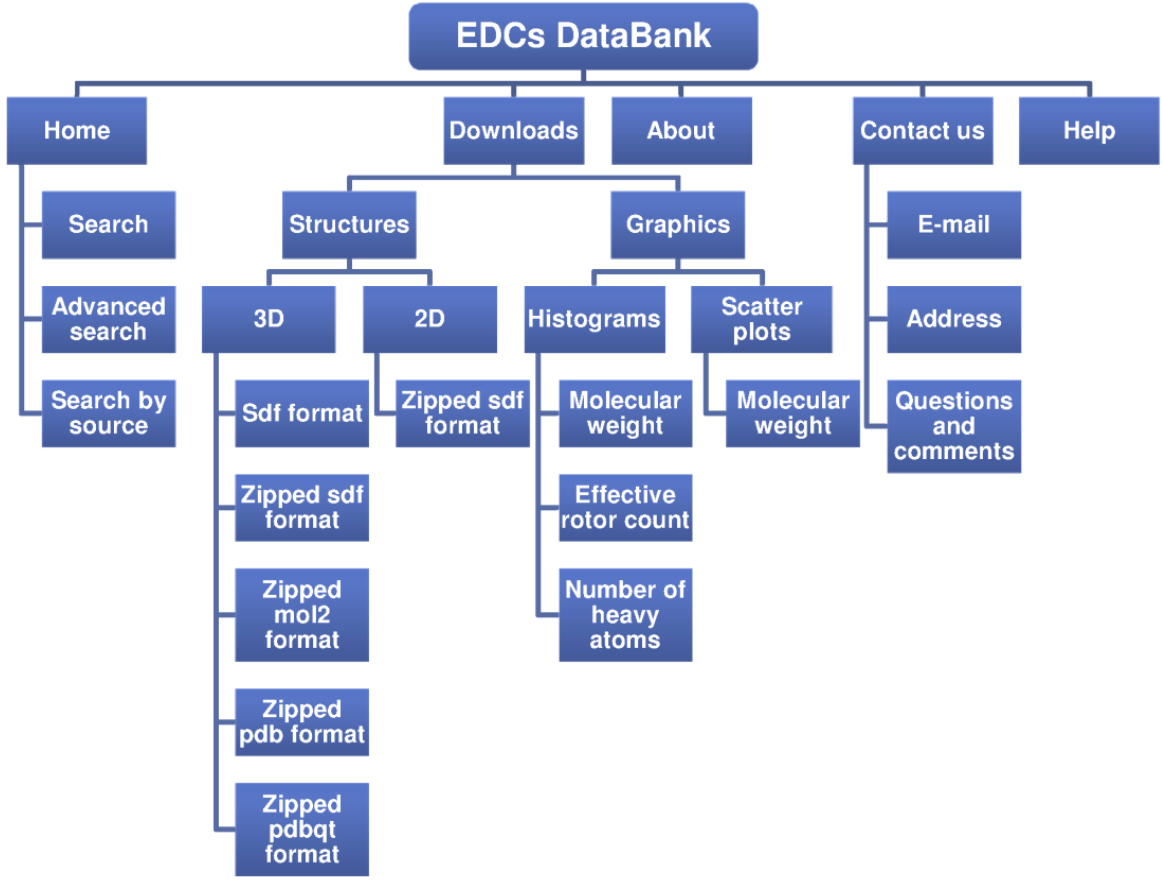


Figure 3.7. Map of EDCs DataBank website.

3.2.4.3. Search and advanced search. This database is fully searchable. The general searcher (Figure 3.8A), located at the top of the website home page, allows the user to seek expressions found in the category name, synonyms and the IUPAC name. The searcher considers many options to call a molecule, facilitating the searching and identification of the compounds. The advanced search (Figure 3.8B) generates a query according to the search parameter selected by the users. Compounds may be searched in a range of molecular weight, number of bond donors, number of heavy atoms (atoms different from hydrogen), number of bond acceptors, and a multiple entry search. The search by source (Figure 3.8C), found at the home page below the search box, consists of 51 icons representing the possible sources of exposure to endocrine disruptors. This tool helps to make our database more user-friendly and useful to the general public, by showing the compounds that may be found on certain elements or products to which people could be exposed to.

3.2.4.4. EDCs DataSheet. EDCs DataBank is a non-redundant database built on MySQL. The identifier InChIKey has been used as a primary key, in order to make it not redundant. In addition, given the current state of information and the difficulty of handling it, especially in terms of codes of molecules, we decided not to incorporate a new identifier of molecules for our database. Therefore, we use the existing and widely recognized PubChem compound identifier (CID)[4], which is shown in identifiers and utilized to name the 2D- and 3D-structure files and images in the different formats offered by this database.

All the information stored on the MySQL database is presented in a user-friendly web environment in the EDCs DataSheets (Figure 3.9). The website has been developed utilizing HTML, PHP, CSS, JQuery and JavaScript languages using an appropriately graphic design in order to simplify the processes of finding and reading data.

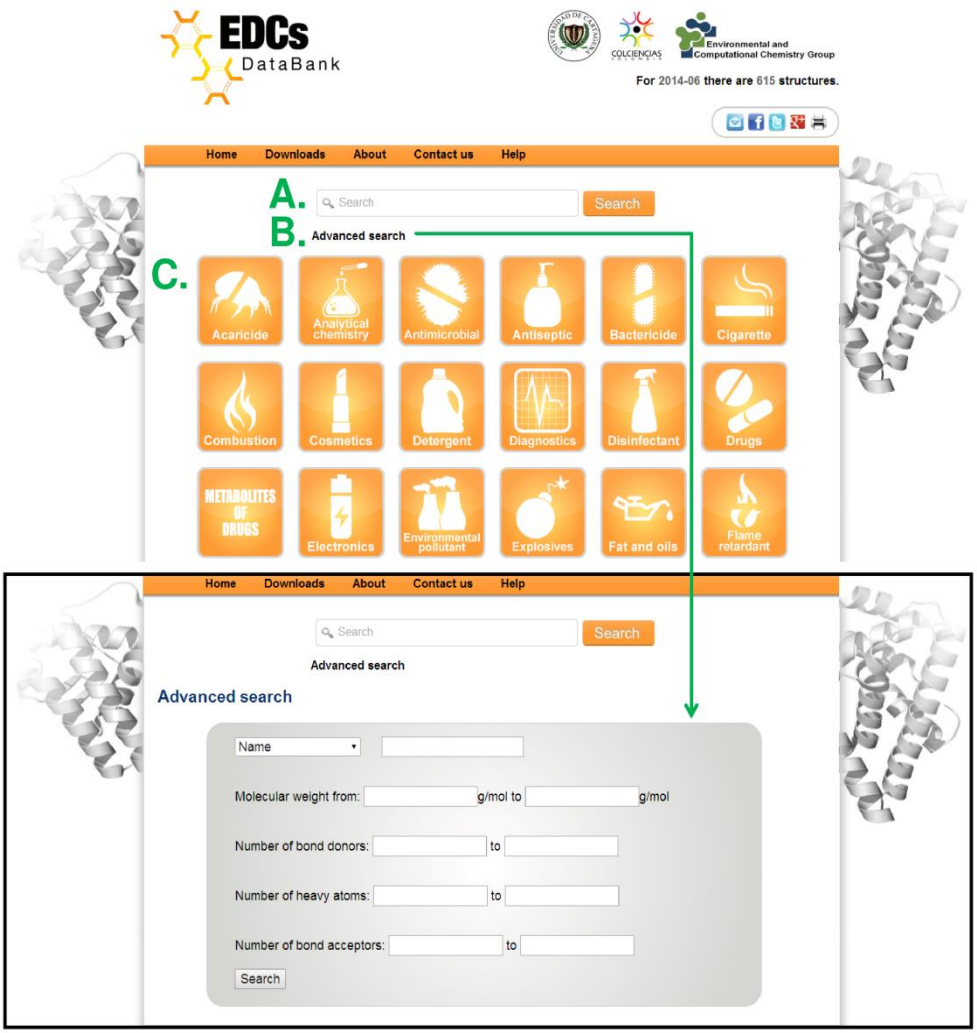
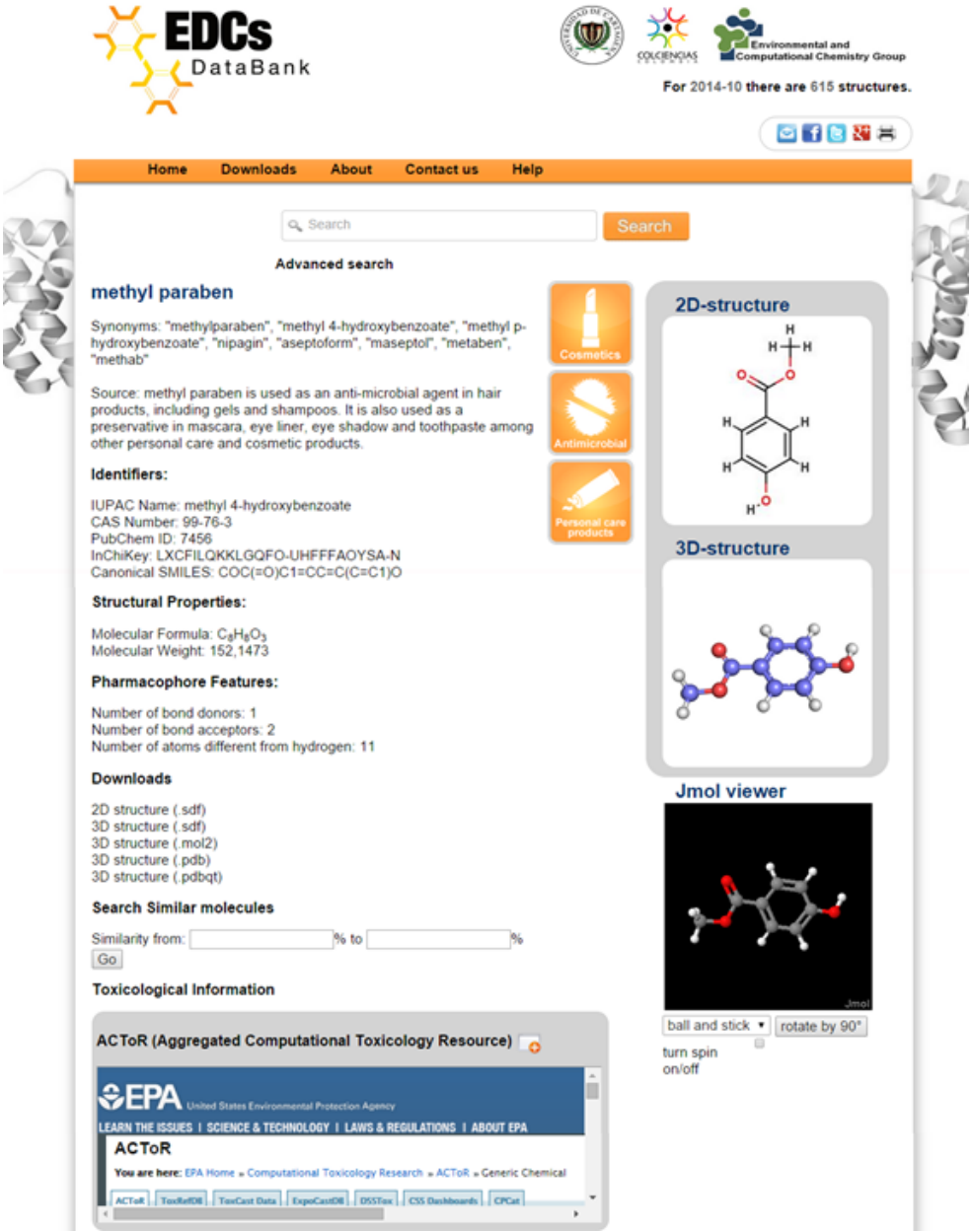


Figure 3.8. Screenshot of search of (A) search, (B) advanced search and (C) source.


EDCs DataBank
Environmental and Computational Chemistry Group
For 2014-10 there are 615 structures.
Home
Downloads
About
Contact us
Help
Search
Search
Advanced search
methyl paraben
Synonyms: "methylparaben", "methyl 4-hydroxybenzoate", "methyl p-hydroxybenzoate", "nipagin", "aseptoform", "maseptol", "metaben", "methab"
Source: methyl paraben is used as an anti-microbial agent in hair products, including gels and shampoos. It is also used as a preservative in mascara, eye liner, eye shadow and toothpaste among other personal care and cosmetic products.
Cosmetics
Antimicrobial
Personal care products
Identifiers:
IUPAC Name: methyl 4-hydroxybenzoate
CAS Number: 99-76-3
PubChem ID: 7456
InChiKey: LXCFILQKKLGQFO-UHFFFAOYSA-N
Canonical SMILES: COC(=O)C1=CC=C(C=C1)O
Structural Properties:
Molecular Formula: C8H8O3
Molecular Weight: 152,1473
Pharmacophore Features:
Number of bond donors: 1
Number of bond acceptors: 2
Number of atoms different from hydrogen: 11
Downloads
2D structure (.sdf)
3D structure (.sdf)
3D structure (.mol2)
3D structure (.pdb)
3D structure (.pdbqt)
Search Similar molecules
Similarity from: % to %
Go
Toxicological Information
ACToR (Aggregated Computational Toxicology Resource)
EPA United States Environmental Protection Agency
LEARN THE ISSUES | SCIENCE & TECHNOLOGY | LAWS & REGULATIONS | ABOUT EPA
ACToR
You are here: EPA Home » Computational Toxicology Research » ACToR » Generic Chemical
2D-structure
3D-structure
Jmol viewer
Jmol
ball and stick
rotate by 90°
turn spin on/off

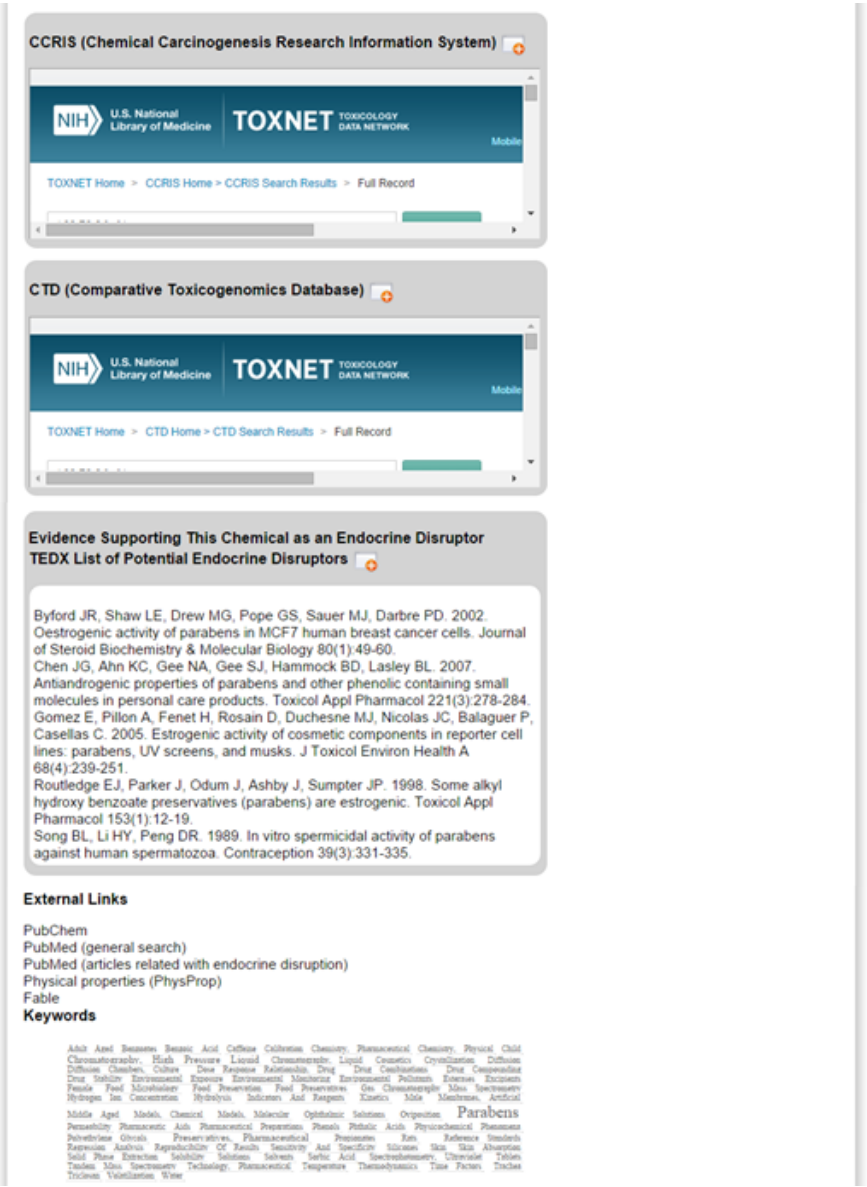


Figure 3.9. EDCs DataSheet screenshot for methyl-paraben.

EDCs DataSheets show the information for each single compound selected by the user. The data is presented in 10 main categories: name, synonyms, source, identifiers, structural properties, pharmacophore features, downloads, search similar structures, toxicological information, external links and keywords, which are accompanied by the source icons, 2D- and 3D-structure figures and an interactive three-dimensional view of each compound by Jmol (http://jmol.sourceforge.net/index.en.html/), which allows to perform simple calculations such as van der Waal surface, distances, angles, dihedral angles, among others. Another interesting tool in the EDCs DataSheet is the searcher for similar molecules. This tool was created by employing a similarity matrix calculated in the PubChem Score Matrix Service (https://pubchem.ncbi.nlm.nih.gov/score_matrix/score_matrix-help.html). The 2D similarity (substructure keys) between molecules was used as scoring method, so that the client only has to include the range of similitude percentage for the search.

Currently, EDCs DataBank is the unique database of endocrine disruptors with three dimensional structures available for virtual screening, with a complete repository of information regarding EDCs and their molecular structures. Therefore, it will have an impact in the study of these compounds through bioinformatics and computational toxicology strategies.

3.2.5. Quality assurance, validation, completeness and curation

The data contained in EDCs DataBank has been revised and hand curated, diminishing the possible errors. All the fields of this database are fully complete with exception of external links and toxicological information that depend on the availability of the data in the source databases. In addition, this database uses PubChem as primary source for the three-dimensional structures, which is one of the most recognized and employed databases for getting chemical structures of low molecular weight compounds[4].

3.2.6. Chemical structural space and frequency distribution

The PCA of molecular descriptors obtained for all the molecules in the database showed that the 43.02% of the total variance is expressed by the

first three components. Component one comprises 27.81%, component two 8.06% and component three 7.14%. The analysis of the chemical space is presented in Figure 3.10, represented by the three-dimensional plot of the factors scores (Annex 3.1). This shows the structural diversity of the compounds stored in EDCs DataBank.

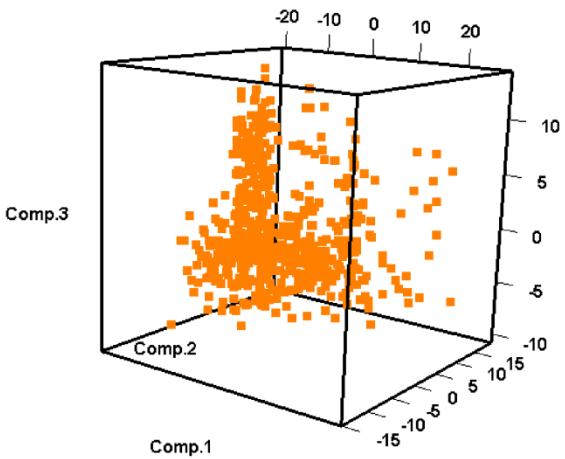


Figure 3.10. Chemical structure space of EDCs DataBank (Comp :Component).

The examination of the molecular descriptors eigenvectors (Figure 3.11; Annex 3.1) explains which of them account for each principal component. The first principal component is characterized by the predominance of GETAWAY descriptors (13.4%), which are descriptors calculated from the leverage matrix obtained by the centered atomic coordinates; followed by Eigenvalue-based indices (11.0%), topological descriptors calculated by the eigenvalues of a square matrix representing a molecular graph; BCUT descriptors (10.5%), molecular descriptors obtained from the positive and negative eigenvalues of the adjacency matrix, weighting the diagonal elements with atom weights; topological descriptors (10.3%), molecular descriptors obtained from molecular graph; edge adjacency indices (10.3%), topological molecular descriptors derived from the edge adjacency matrix, which encodes the

connectivity between graph edges; walk and path counts (9.5%), molecular descriptors obtained from the molecular graph, counting paths, walks and self-returning walks of different lengths and constitutional descriptors (8.6%), OD-descriptors, independent from molecular connectivity and conformations. The second descriptor is predominantly defined by WHIM descriptors (29.6%), molecular descriptors obtained as statistical indices of the atoms projected onto the 3 principal components obtained from weighted covariance matrices of the atomic coordinates, and the third one by GETAWAY descriptors (17.7%).

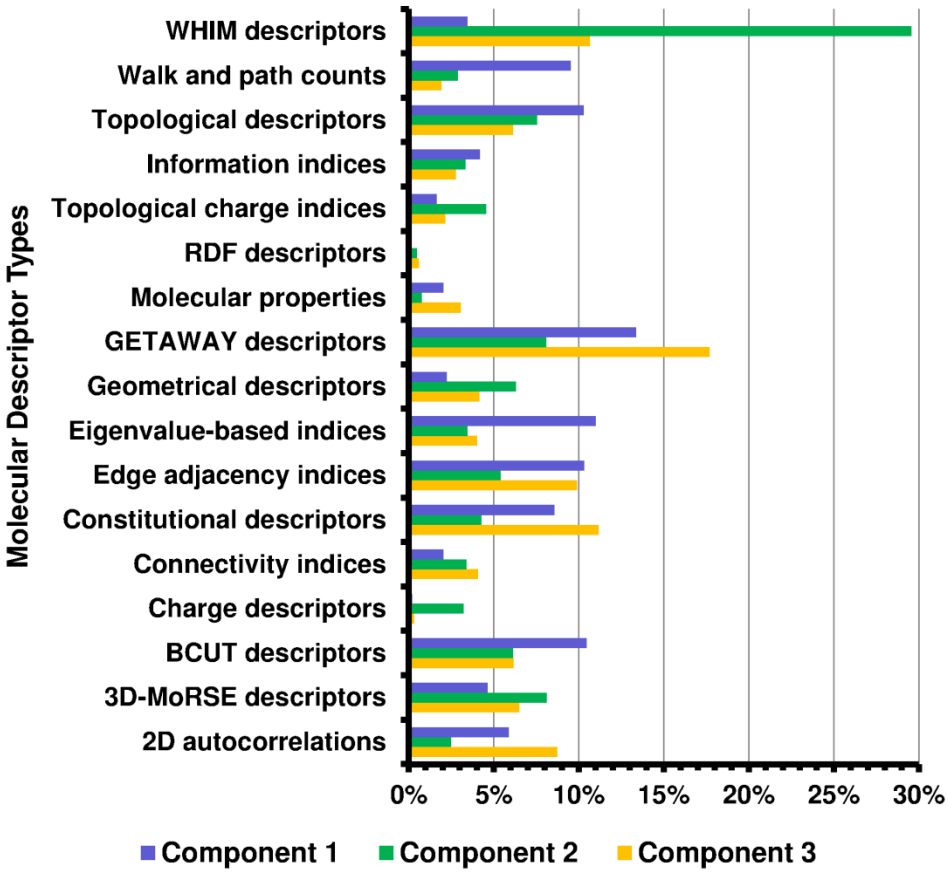


Figure 3.11. Statistical distribution of molecular descriptors in the three principal components.

Frequency distribution of structural features (Figure 3.12.) showed that the 80.16% of the molecules stored in EDCs DataBank have 0-1 bond donors. The 72.04% of the compounds have 0-2 bond acceptors, and the 54.15% of them have 14-20 heavy atoms (atoms different from hydrogen). The molecular weight of the majority of the molecules is in the range from 100-400 g/mol.

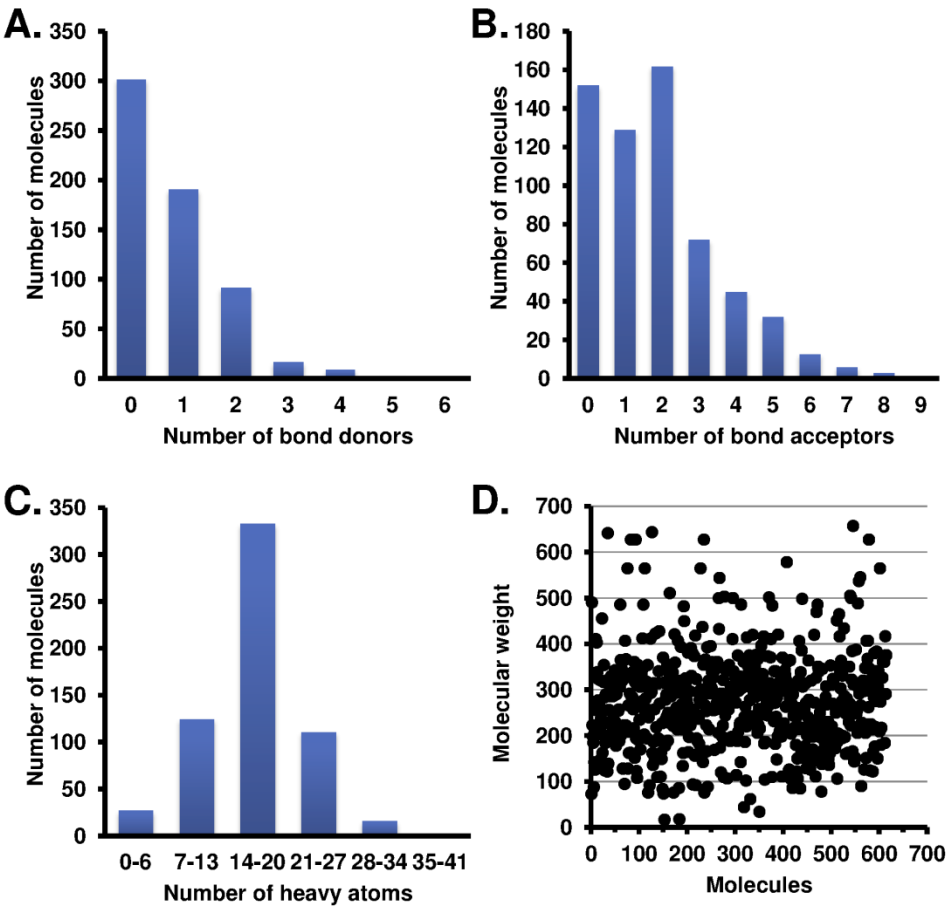


Figure 3.12. Statistical distribution of (A) number of bond donors, (B) number of bond acceptors, (C) number of heavy atoms and (D) scatter plot of molecular weight.

Therefore, molecules belonging to EDCs DataBank are usually of low molecular weight, with a low number of bond acceptors and bond donors. Consequently, these molecules have properties that facilitate their absorption by the body. The expected interactions of these molecules are mostly hydrophobic in nature.

3.3. DISCUSSION

Historically, the public, scientific and regulatory concern regarding the potential adverse health effects of exposure to EDCs were initiated in the early 1990s[12]. Since then, numerous efforts have been made to compile information related to these compounds. In 1988, before the term was coined "endocrine disruptor", Dr. Theo Colborn began the search for compounds with potential to interfere with the endocrine system. These compounds were consolidated in the TEDX list of potential endocrine disruptors (http://endocrinedisruption.org/endocrine-disruption/tedx-list-of-potential-endocrine-disruptors/overview). Another interesting project started in mid-1990s, was EDKB (Established Knowledge Base for Endocrine Disrupting Chemicals), which consists of a database with information on the chemical structure, name and experimental data, containing more than 3257 records and about 1800 EDCs[13]. In 2010, Nashev et al. created an internal database of endocrine disruptors which was used to perform virtual screening studies with protein 17ß dehydrogenase type 3[14]. Later, in 2012, the database was employed for the identification of xenobiotics as potential disruptors of corticosteroid action[15]. Other databases in this field are RBD (Development of the Receptor Database), which can assist with the problem of endocrine disruptors, presenting information about steroid hormone receptors to which endocrine disruptors could join[16]; EDID (Endocrine Disrupting Chemicals–Diet Interaction Database) has information about the possible interactions between diet and endocrine disruptors, allowing the search for scientific papers, showing this information by entering words found in the article title, keywords, and/or publication year[17]; EDPSD (Endocrine Disruptor Priority

Settings Database) is a useful database which ranks priority chemical parameters for screening and testing endocrine disruptors (http://www.epa.gov/endo/pubs/prioritysetting/finalarch.htm); and EADB (Estrogenic Activity Database), which contains information of estrogenic activity for 8212 chemicals in *in vitro* and *in vivo* assays of different species[10]. However, there was not a database of EDCs with three-dimensional structures available for virtual screening. Therefore, EDCs DataBank is a valuable repository to study the interactions of these compounds with macromolecules involved in several diseases, such as breast cancer, obesity and diabetes, among others. In addition, it was created in a very user-friendly environment that makes it helpful for community information regarding EDCs exposure in everyday products and their eventual toxicity impact in environmental and human health.

3.4. CONCLUSIONS

EDCs DataBank is the unique database with three-dimensional structures of EDCs for virtual screening, freely available on http://edcs.unicartagena.edu.co. It is also suitable for its use with other computational approaches as all molecular structures stored in it can be downloaded in several formats. Therefore, it is a valuable repository to study the molecular basis of the interaction between these chemicals and macromolecules involved in diverse pathologies. In addition, this database has been developed in a user-friendly environment and provides a vast amount of structural and toxicological information of each compound that can be used for research, academia and general population.

REFERENCES

1. (a) Markey, C. M.; Rubin, B. S.; Soto, A. M.; Sonnenschein, C., Endocrine disruptors: from Wingspread to environmental developmental biology. J. Steroid Biochem. Mol. Biol. **2002,** 83 (1–5), 235-244; (b) Mezcua, M.; Martínez-Uroz, M.; Gómez-Ramos, M.; Gómez, M.; Navas, J.; Fernández-Alba, A., Analysis of synthetic endocrine-disrupting chemicals in food: A review. Talanta **2012,** 100, 90-106; (c) Muncke, J., Endocrine disrupting chemicals and other substances of concern in food contact materials: An updated review of exposure, effect and risk assessment. J. Steroid Biochem. Mol. Biol. **2011,** 127 (1–2), 118-127.
2. Stojić, N.; Erić, S.; Kuzmanovski, I., Prediction of toxicity and data exploratory analysis of estrogen-active endocrine disruptors using counter-propagation artificial neural networks. J. Mol. Graph. Model. **2010,** 29 (3), 450-460.
3. O'Boyle, N.; Banck, M.; James, C.; Morley, C.; Vandermeersch, T.; Hutchison, G., Open Babel: An open chemical toolbox. J. Cheminform. **2011,** 3 (1), 33.
4. Bolton, E. E.; Wang, Y.; Thiessen, P. A.; Bryant, S. H., Chapter 12 PubChem: Integrated Platform of Small Molecules and Biological Activities. In Annu. Rep. Comput. Chem., Ralph, A. W.; David, C. S., Eds. Elsevier: 2008; Vol. Volume 4, pp 217-241.
5. Wexler, P., TOXNET: An evolving web resource for toxicology and environmental health information. Toxicology **2001,** 157 (1–2), 3-10.
6. Judson, R.; Richard, A.; Dix, D.; Houck, K.; Elloumi, F.; Martin, M.; Cathey, T.; Transue, T. R.; Spencer, R.; Wolf, M., ACToR—aggregated computational toxicology resource. Toxicol. Appl. Pharmacol. **2008,** 233 (1), 7-13.
7. Dix, D. J.; Houck, K. A.; Martin, M. T.; Richard, A. M.; Setzer, R. W.; Kavlock, R. J., The ToxCast program for prioritizing toxicity testing of environmental chemicals. Toxicol. Sci. **2007,** 95 (1), 5-12.
8. Sarkar, I. N.; Schenk, R.; Miller, H.; Norton, C. N. In LigerCat: using "MeSH clouds" from journal, article, or gene citations to facilitate the identification of relevant biomedical literature, AMIA Annual Symposium Proceedings, American Medical Informatics Association: 2009; p 563.
9. Tetko, I. V., Computing chemistry on the web. Drug Discov. Today **2005,** 10 (22), 1497-1500.
10. Shen, J.; Xu, L.; Fang, H.; Richard, A. M.; Bray, J. D.; Judson, R. S.; Zhou, G.; Colatsky, T. J.; Aungst, J. L.; Teng, C., EADB: an estrogenic activity database for assessing potential endocrine activity. Toxicol. Sci. **2013,** 135 (2), 277-291.
11. Lauria, A.; Ippolito, M.; Almerico, A. M., Principal component analysis on molecular descriptors as an alternative point of view in the search of new Hsp90 inhibitors. Comp. Biol. Chem. **2009,** 33 (5), 386-390.
12. Safe, S. H., Endocrine disruptors and human health--is there a problem? An update. Environ. Health. Perspect. **2000,** 108 (6), 487.
13. Ding, D.; Xu, L.; Fang, H.; Hong, H.; Perkins, R.; Harris, S.; Bearden, E. D.; Shi, L.; Tong, W., The EDKB: an established knowledge base for endocrine disrupting chemicals. BMC bioinformatics **2010,** 11 (Suppl 6), S5.

14. Nashev, L. G.; Schuster, D.; Laggner, C.; Sodha, S.; Langer, T.; Wolber, G.; Odermatt, A., The UV-filter benzophenone-1 inhibits 17ß-hydroxysteroid dehydrogenase type 3: Virtual screening as a strategy to identify potential endocrine disrupting chemicals. Biochemical pharmacology **2010,** 79 (8), 1189-1199.
15. Nashev, L. G.; Vuorinen, A.; Praxmarer, L.; Chantong, B.; Cereghetti, D.; Winiger, R.; Schuster, D.; Odermatt, A., Virtual Screening as a Strategy for the Identification of Xenobiotics Disrupting Corticosteroid Action. PLoS ONE **2012,** 7 (10), e46958.
16. Nakata, K.; Takai, T.; Kaminuma, T., Development of the receptor database (RDB): application to the endocrine disruptor problem. Bioinformatics **1999,** 15 (7), 544-552.
17. Baldi, F.; Mantovani, A., A new database for food safety: EDID (endocrine disrupting chemicals–diet interaction database). Reprod. Toxicol. **2008,** 26 (1), 57.

CHAPTER 4

Identification of new targets for BPA

BPA is an industrial pollutant used as an important monomer precursor for the production of various polycarbonate, polysulphonate plastics and epoxy resins[1]. It has been commercialized for more than 50 years[2], and as early as 1936, it was shown to have estrogenic activity *in vivo* when it was administered to rats by subcutaneous injection[3]. The frequency of human exposure to BPA has steadily augmented in recent years, as a result of the increased use of both polycarbonate plastics and epoxy resins in the coatings of food cans and plastic water bottles[4].

Some of the known mechanisms of BPA as endocrine disruptor include the modulation of the PGR gene expression in the hypothalamus, the binding of the androgen receptor as antagonist, the stimulation of testosterone production and release of prolactin[5].

In rats, perinatal exposure to BPA may change the prostate epigenome and thereby potentially predisposes the prostate gland to abnormal growth and carcinogenesis[6]. It can also cause altered DNA methylation patterns in other cell signaling genes, suggesting BPA may exert its effects through epigenetic mechanisms[4]. In addition, it can also be transferred across the human placenta where it binds to both nuclear and membrane ER[7] producing effects in subsequent generations.

The exposure to this xenoestrogen has been associated to numerous health conditions[7], such as diabetes, obesity, reproductive disorders, cardiovascular diseases and different cancer types, among others[8]. These could be mediated for its interaction with different cell signaling pathways. Therefore, in this

chapter, an inverse virtual screening study is presented to predict new protein targets for BPA, which constitute the second part of this thesis.

4.1. MATERIALS AND METHODS

4.1.1. Materials

All *in silico* calculations were performed using Dell precision T7400 Mini-tower workstation with two Intel® Xeon® processors X5492, 12 M cache, 3.40 GHz, 1600 MHz, 32 GB RAM, running on Red Hat Enterprise Linux 5 operating system.

4.1.2. Text-mining

A text-mining step was carried out in order to select proteins that exhibited connection with BPA in scientific reports. These proteins were recognized through literature reviews using a wide variety of academic databases and search engines available online such as Ali Baba (http://alibaba.informatik.hu-berlin.de/), PubMed PubReMiner (http://bioinfo.amc.uva.nl/human-genetics/pubreminer/), IHop, Information Hyperlinked over Proteins[4] and Chilibot[9], utilizing the expression: "protein and bisphenol A" to carry out the search. An association mining graph was generated using the terms found by the text mining, as well as the number of hits shown in PubMed (http://www.ncbi.nlm.nih.gov.ezproxy.unal.edu.co/pubmed/) when each word and "bisphenol A" were used as inputs.

4.1.3. Preparation of proteins structures

The coordinates of the BPA-linked proteins chosen by the data mining were downloaded from PDB[10] in pdb format. Some of the proteins do not have a crystallographic structure, therefore their modeled structures were downloaded from SWISS-MODEL and ModBase through the UniProtKB webpage (http://www.uniprot.org/). Then, the three-dimensional structures of the proteins were opened with Sybyl 8.1. program package (Tripos, St. Louis, MO), and prepared using the structure preparation tool of the same software;

all ions, water molecules and other substructures were removed from the coordinate file before docking[11]. The protein structures were pre-analyzed and prepared for the docking runs using the biopolymer structure preparation tool with default settings as implemented in the Sybyl program package (Tripos, St. Louis, MO)[12]. The resultant geometry of each protein was subsequently optimized by the Powell method utilizing the Kollman United force field, AMBER charges, dielectric constant 1.0, NB cutoff 8.0, maximum interactions 100 and termination gradient 0.001 kcal/mol [13] (Figure 4.1).

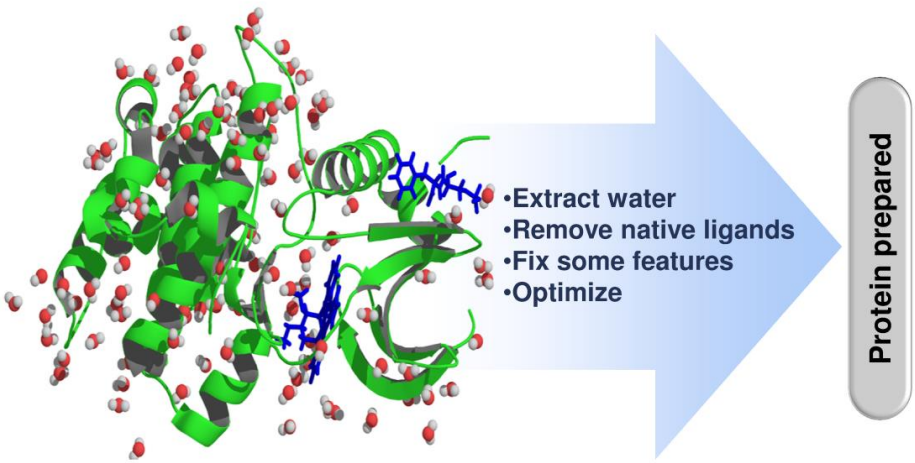


Figure 4.1. Protein preparation workflow.

The minimized pdb-formatted file was directly read into AutoDock Tools of MGLTools[14], which was employed to prepare the input pdbqt file required by AutoDock Vina, and to set the size and center to the grid box. Kollman charges and polar hydrogen atoms were added to the three-dimensional structures of the proteins using the same software[15].

4.1.4. Ligand structure

BPA structure was drawn and optimized by the DFT (density functional theory) method at the B3LYP/6-31G level in Gaussian 03 program package[16]. After optimization, the output file was translated to pdb with Open Babel[17] and to pdbqt using AutoDockTools[14].

4.1.5. Inverse virtual screening protocol

Docking coordinates were determined through a grid box in AutoDockTools (http://autodock.scripps.edu/resources/adt) utilizing a blind docking strategy with spacing of 0.357 Å and center on macromolecule[18] in order to include all the possible binding sites for BPA. The docking study was performed utilizing AutoDock Vina 1.0 molecular docking and virtual screening program[19] running on Linux; energy range of 1.5, number of modes of 20, and an exhaustiveness of 25 were employed. Resulting binding affinities were then utilized for ranking the proteins as BPA targets.

4.1.6. Refinement docking experiments

Refinement docking experiments with repetitions of 100 runs were performed on complexes presenting high affinity scores (lower than -8.0 kcal/mol) to obtain more accurate results. This was achieved by applying a blind docking strategy in AutoDock Vina[19]. The parameters utilized for this step were energy range of 1.5, number of modes of 50, and an exhaustiveness of 100.

4.1.7. Conformational analysis

LigandScout 3.0 program[20] was used to determine the interactions existing for the protein-BPA complexes with greatest affinity score values. The interaction cutoff threshold of the pdb interpretation settings was set to 7 Å. This threshold defines a sphere (in Å) around the ligand. The atoms of the protein enclosed inside the sphere were considered to be possibly involved in the interactions. All of the remaining settings were maintained as the default[21]. In addition, the predicted binding mode and the hydrogen bond interactions were explored with Molegro Molecular Viewer Software 2.2.0 (http://www.molegro.com/mmv-product.php).

4.1.8. Validation

In order to validate docking data obtained for BPA with the presumable targets, and to evaluate the capability of the docking process to locate the ligand in the same binding site reported for the experimental structures, *in silico* docking calculations were performed using the three-dimensional crystallographic structures of the human estrogen-related receptor gamma

(ESRRG) complexed with BPA (PDB: 2E2R and 2P7G). The BPA and any other substructures, as well as all water molecules, were removed from the downloaded pdb file employing Sybyl 8.1. software package (Tripos, St. Louis, MO). The resultant BPA and protein structures were optimized and docked employing the same protocols used during the *in silico* screening and the refinement docking experiments. The obtained affinity binding values of these complexes were used for comparison purposes. Ideally, complexes depicting binding affinities similar to those generated for the experimental structures are likely to occur at a molecular level. In addition, alignments of the protein and BPA structures before and after optimization and docking were performed by Sybyl 8.1. software package (Tripos, St. Louis, MO), to examine the corresponding docking pose predicted by AutoDock Vina[19] with respect to the experimental one.

The consistency of the docking procedures used for obtaining theoretical binding affinities was validated by utilizing experimental data from the literature (ToxCast database)[22]. Accordingly, AutoDock Vina binding affinities were calculated for 20 different proteins, including ER-alpha and ESRRG, which have been extensively reported as BPA targets. Protein preparation and optimization, as well as the docking procedure were carried out exactly as described before in the inverse virtual screening protocol, performing ten runs for each protein-BPA complex. Subsequently, a correlation between the average of AutoDock Vina-calculated affinity scores (kcal/mol) and the logarithm of the experimental activating concentration 50% (AC_{50}) obtained from the literature, was measured by GraphPad InStat software (3.05 version, 2000).

4.2. RESULTS

4.2.1. Text-mining

Based on text mining data, 271 proteins were chosen as commonly reported to have a role in different biochemical processes for which BPA has been directly or indirectly involved. A text mining graph (Figure 4.2) shows the network for

BPA and the number of hits found in PubMed. Accordingly, a total of 32 nuclear receptors, 15 circadian clock proteins, 71 proteins involved in insulin-signaling, 78 serum-related proteins, 36 breast cancer proteins and 39 proteins belonging to diverse biochemical pathways were selected for inverse virtual screening.

4.2.2. Inverse virtual screening with AutoDock Vina

An inverse virtual screening was carried out through docking protocols for the selected sets of proteins and BPA (Annexes 4.1-4.6) using AutoDock Vina[19]. The best docking affinity values (lower than -8.0 kcal/mol) in the group of nuclear receptors (Annex 4.1) were found for the complexes formed by BPA and ESRRG, constitutive androstane receptor (CAR), thyroid hormone receptor alpha (EAR-7), ER and retinoic acid receptor beta (RARB), with affinity scores of -9.9, -8.9, -8.8, -8.3 and -8.1 kcal/mol, respectively. The circadian clock proteins (Annex 4.1.), dual specificity protein kinases CLK4, CLK1 and CLK2 exhibited the best affinity scores for BPA, with values of -9.5, -9.0 and -9.0 kcal/mol, respectively. These proteins are biochemically unrelated to the ER. The protein involved in insulin-signaling (Annex 4.1.) with the best affinity score was protein kinase C theta (PKCθ) with a value of -8.6 kcal/mol.

4.2.3. Refinement docking experiments

BPA target proteins with high affinity for the plasticizer (lower than -8.0 kcal/mol), as suggested by the inverse virtual screening, were submitted to refinement docking experiments, obtaining similar results (Table 4.1.), and in all cases, the standard deviation for the 100 runs was less than one. As expected, the ligand–protein complex with the highest binding affinity was ESRRG, with an average of -9.9 ± 0.0 kcal/mol. Moreover, the process reaffirms the idea that some non-estrogen receptor related proteins may possess theoretical binding sites for BPA. In particular, dual specificity protein kinases CLK4, CLK1, CLK2 with docking affinity values of -9.5 ± 0.0, -9.0 ± 0.0 and -9.0 ± 0.0 kcal/mol, respectively, showed promising *in silico* affinities for BPA, deserving further investigation.

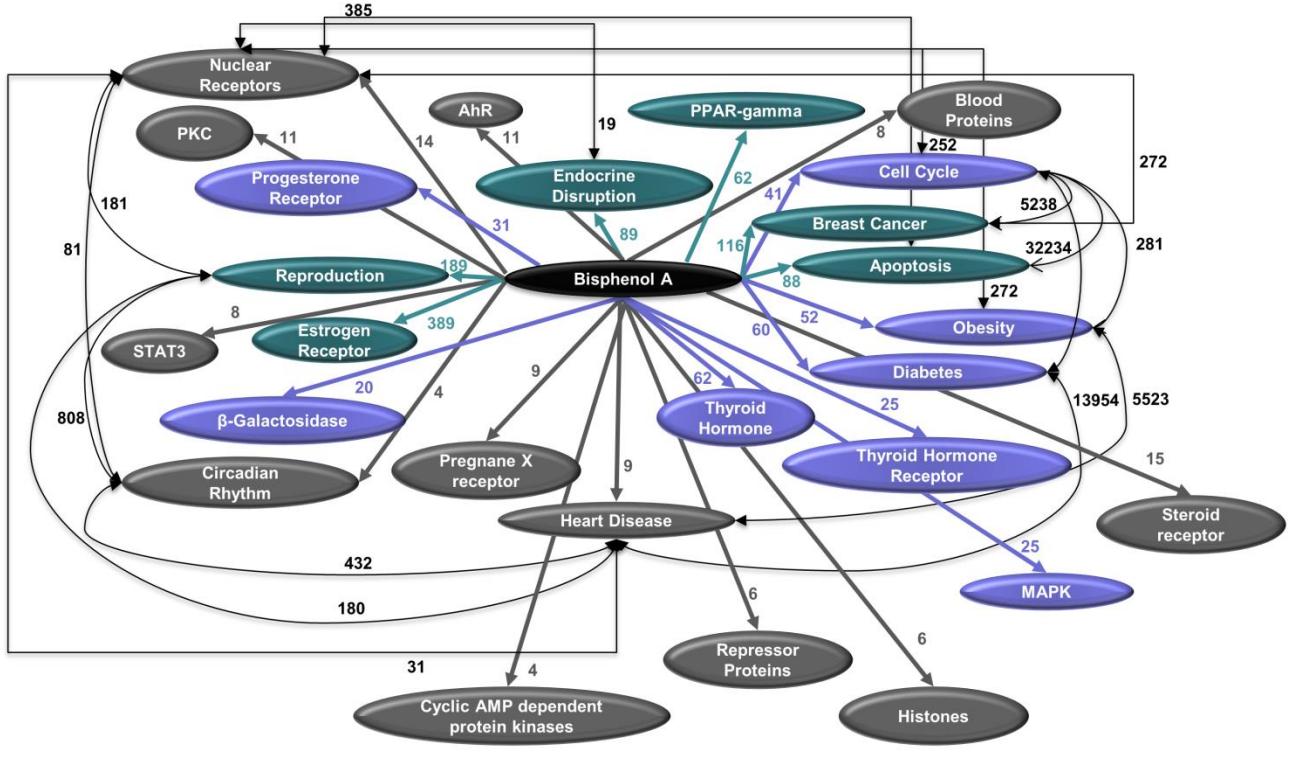


Figure 4.2. Text-mining graph based on the PubMed search of the root keyword "bisphenol A", indicating the number of hits found in PubMed for this and each term found; updated August 27, 2012.

Table 4.1. Results of the refinement docking experiments with AutoDock Vina.

Target name	PDB/UniProtKB	Affinity (kcal/mol)
Nuclear receptors		
ESRRG	2E2R/P62508	-9.9 ± 0.0
RARB	1XAP/P10826	-8.0 ± 0.0
EAR-7	2H79/P10827	-8.7 ± 0.0
ER-alpha	1XPC/P03372	-8.3 ± 0.0
RXR-alpha	3KWY/P19793	-8.2 ± 0.0
CAR	1XVP/Q14994	-8.9 ± 0.0
Circadian clock		
CLK1	2VAG/P49759	-9.1 ± 0.0
CLK2	3NR9/P49760	-9.0 ± 0.0
CLK4	MODBASE/Q9HAZ1	-9.5 ± 0.0
Insulin signaling pathway		
PKC-alpha	3IW4/P17252	-7.8 ± 0.1
PKC-theta	1XJD/Q04759	-8.6 ± 0.0
Serum proteins		
R-PTP-epsilon	2JJD/P23469	-8.0 ± 0.0
ADORA2A	3EML/P29274	-8.6 ± 0.0
MMP-13	830C/P45452	-8.0 ± 0.0
TIMP-3	3CKI/P35625	-8.0 ± 0.8
Breast cancer		
MMP-8	1I76/P22894	-7.6 ± 0.0
TOP2A	1ZXM/P11388	-8.1 ± 0.0
Other proteins		
SHBG	1F5F/P04278	-8.2 ± 0.0
RXR-Beta	1UHL/P28702	-8.1 ± 0.0

4.2.4. Conformational analysis

Results from LigandScout 3.0[19] revealed (Annex 4.1) the more commonly found interactions in protein/BPA complexes correspond to hydrophobic, as well as, hydrogen bond donor and bond acceptor types. Using an interaction cutoff threshold of 7 Å in LigandScout 3.0[20], contact residues participating in the interaction between BPA and CLK4 were Leu293a, Val322a, Phe239a, Val223a, Phe170a, Ala187a, Leu165a (hydrophobic interactions), Leu242a (hydrogen bond acceptor), Asn291a (hydrogen bond donor) and Asp323a

(aromatic ring interaction) (Figure 4.3). Similarly, six residues were found in the interaction between CLK2 and BPA. These were Ala191c, Phe243c, Val227c, Phe174c (hydrophobic interaction, Leu246c (hydrogen bond acceptor) and Asn295c (hydrogen bond donor) (Figure 4.4); and the contact residues that participated in the interaction between CLK1 and BPA were Phe172a, Lys191a, Val175a, Val225a, Val324a, Phe241a, Ala189a, Leu295a (hydrophobic interactions), Glu292a, Asn293a (hydrogen bond donor) and Leu244a (hydrogen bond acceptor) (Figure 4.5).

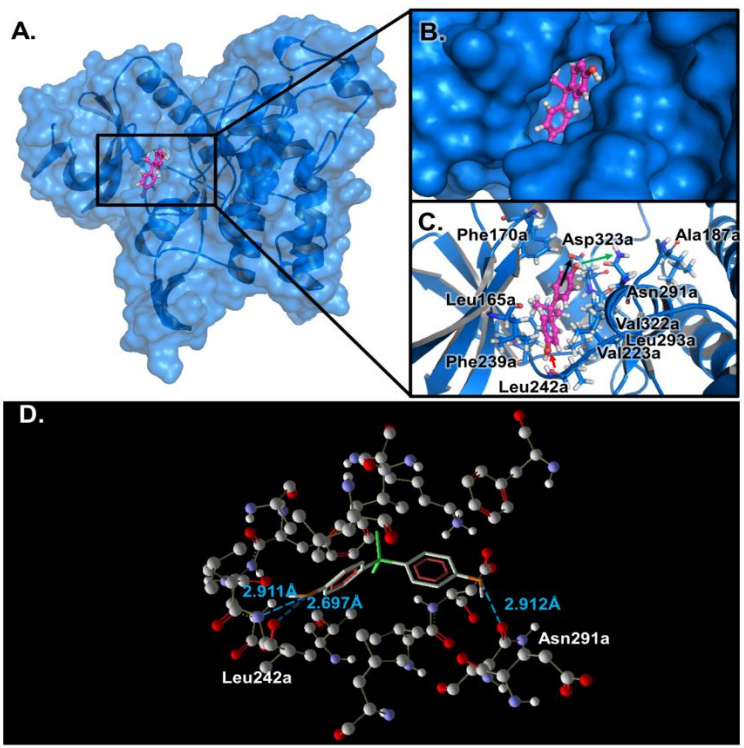


Figure 4.3. Three-dimensional view of the (A) CLK4/BPA complex, (B) binding site, (C) contact residues and interactions predicted by LigandScout 3.0, and (D) hydrogen bonds observed using Molegro molecular viewer software 2.2.0. The red arrows show the hydrogen-bond acceptor features, the green arrows represent hydrogen-bond donor features, the black arrows indicate aromatic ring interactions and the blue lines represent hydrogen bonds.

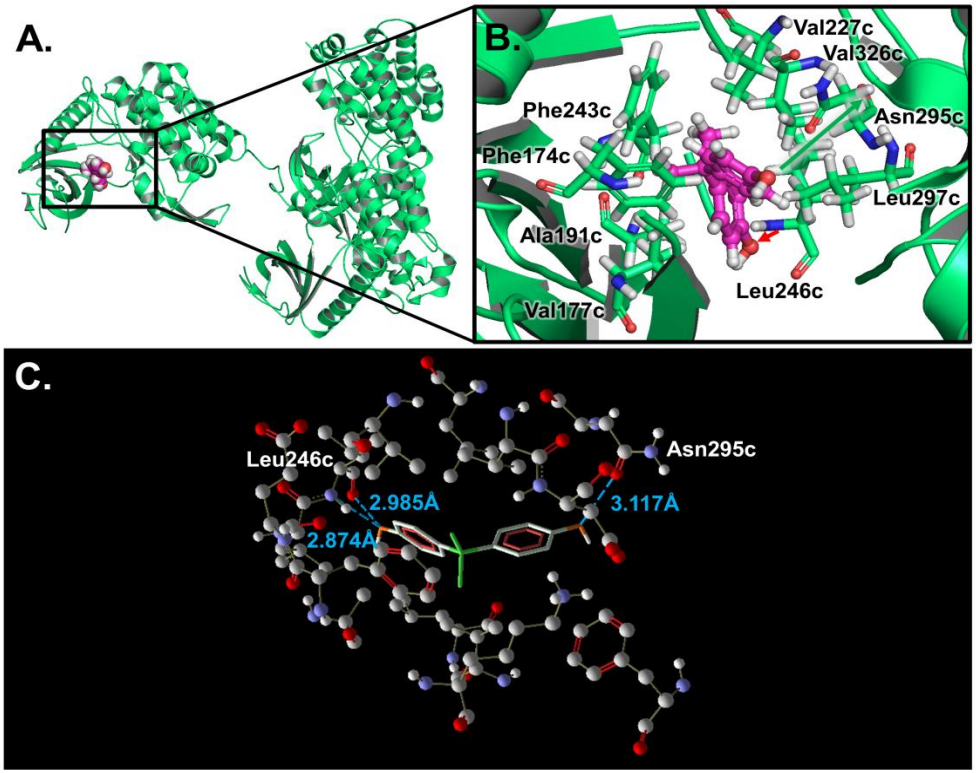


Figure 4.4. Three-dimensional view of the (A) CLK2/BPA complex, (B) contact resides at the binding site and interactions predicted by LigandScout 3.0, and C. hydrogen bonds observed using Molegro molecular viewer software 2.2.0. The red arrows show the hydrogen-bond acceptor features, the green arrows represent hydrogen-bond donor features and the blue lines indicate hydrogen bonds.

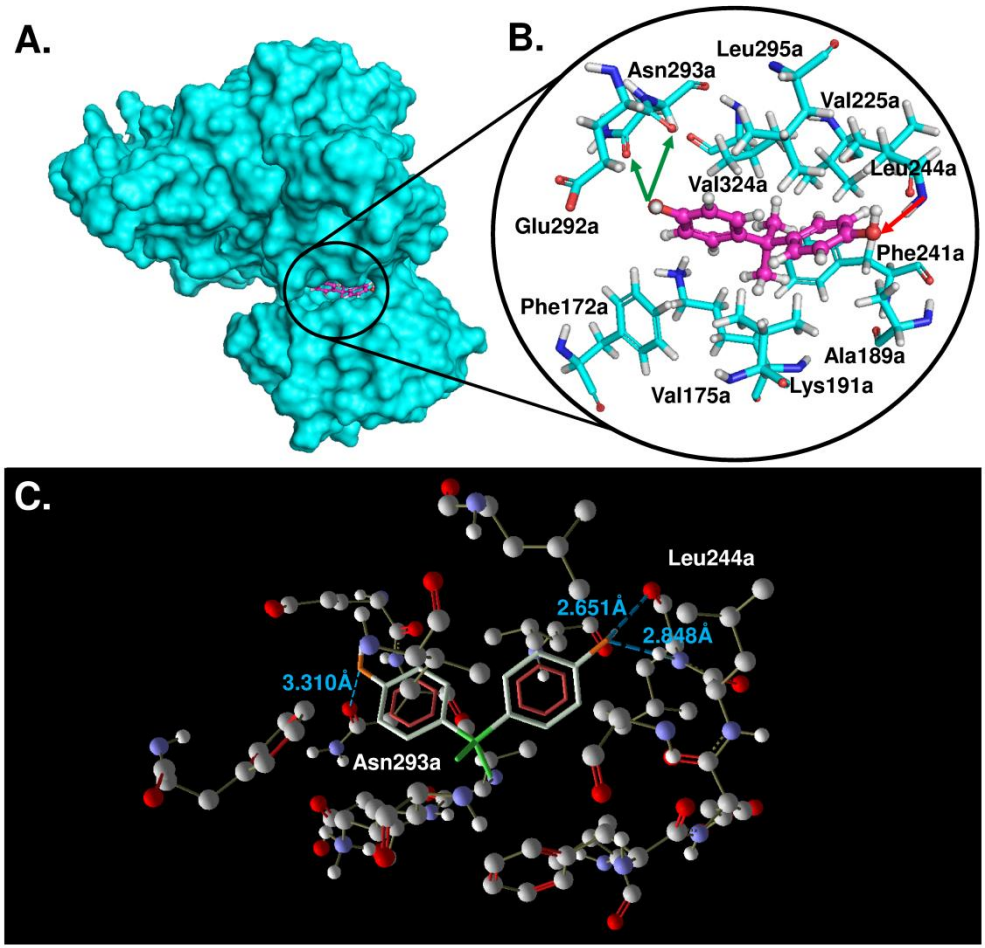


Figure 4.5. Three-dimensional view of the (A) CLK1/BPA complex, (B) contact resides at the binding site and interactions predicted by LigandScout 3.0, and (C) hydrogen bonds observed using Molegro molecular viewer software 2.2.0. The red arrows show the hydrogen-bond acceptor features, the green arrows represent hydrogen-bond donor features and the blue lines indicate hydrogen bonds.

4.2.5. Docking validation

The validation protocol showed that BPA structures resulting from the docking procedures always selected the same site reported by crystallographic data from PDB. The root mean square deviation (RMSD) values for the crystallographic-determined and the docking-resultant poses of the BPA

structure were 0.39 ± 0.07 Å and 0.44 ± 0.07 Å for the proteins under the PDB accession numbers 2E2R and 2P7G, respectively (Figure 4.6). The docking affinity scores obtained for these complexes were -9.9 kcal/mol (ESRRG/BPA; PDB ID: 2E2R) and -9.5 (ESRRG/BPA; PDB ID: 2P7G), similar to those values found from text-mining for the proteins that exhibited the best affinity scores, CLK4 (-9.5 kcal/mol), CLK1 (-9.1 kcal/mol) and CLK 2 (-9.0 kcal/mol).

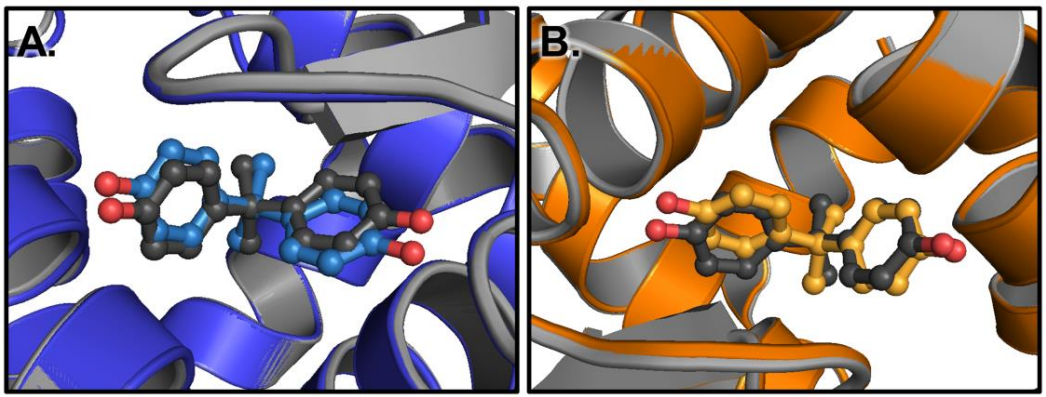


Figure 4.6. Superposition between the crystallographic structures of the complexes ESRRG/BPA with the resultant docking posses of BPA with the optimized structure of (A) ESRRG (PDB: 2E2R) and (B) ESRRG (PDB: 2P7G). Crystallographic structures are represented in gray.

The second step of the validation was performed by calculating the correlation between the docking results and experimental AC_{50} data for different proteins with BPA. These proteins were selected because of having experimental information, regardless of whether they had a high or low activity. Calculated binding affinity and experimental AC_{50} values, as well as Log (AC_{50}) are presented in Table 4.2. The Pearson's correlation coefficient between calculated binding affinity (AutoDock Vina Affinity) and the logarithm of the experimental values of AC_{50} of BPA with 20 different proteins was R = 0.762 with a P-value of 0.0001 (Figure 4.7).

Table 4.2. In silico binding and experimental AC50 and affinities for 20 proteins with BPA.

Name	PDB/UniProtKB	Affinity (Kcal/mol)	AC_{50} (μM)	Log (AC_{50})
ER-alpha	1XPC/P03372	-8.1 ± 0.7	1.040 [22]	0.01703
ER-beta	1QKM/Q92731	-6.9 ± 0.2	1.320 [22]	0.12057
ESRRG	2E2R/P62508	-9.9 ± 0.1	0.013 [22]	-1.88606
MMP-9	1GKC/P14780	-6.4 ± 0.0	40.0 [ToxCast™: BSK_KF3CT_MMP9_down]	1.60206
MMP-1	1HFC/P03956	-6.9 ± 0.7	40.0 [ToxCast™: BSK_BE3C_MMP1_up]	1.60206
RAR-alpha	3KMR/P10276	-7.2 ± 0.6	15.0 [ToxCast™: ATG_RARa_TRANS]	1.17609
AR	3L3X/P10275	-6.2 ± 0.4	73.0 [ToxCast™: NCGC_AR_Antagonist]	1.86332
PPAR-gamma	3LMP/P37231	-7.1 ± 0.1	27.0 [ToxCast™: ATG_PPARg_TRANS]	1.43136
TGFB1	1KLA/P01137	-5.7 ± 0.0	4.44 [ToxCast™: BSK_BE3C_TGFb1_down]	0.64738
VCAM1	1VSC/P19320	-7.0 ± 0.0	4.44 [ToxCast™: BSK_hDFCGF_VCAM1_down]	0.64738
CD69	3CCK/Q07108	-8.2 ± 0.0	1.48 [ToxCast™: BSK_SAg_CD69_up]	0.17026
CYP2C19	4GQS/P33261	-7.7 ± 0.2	6.57 [ToxCast™: NVS_ADME_hCYP2C19_Activator]	0.81756
CYP1A1	4I8V/P04798	-6.7 ± 0.2	9.62 [ToxCast™: NVS_ADME_hCYP1A1]	0.98318
CYP3A4	4K9W/P08684	-8.5±0.0	3.53 [ToxCast™: CLZD_CYP3A4_48]	0.54777
CCL2	1DON/P13500	-6.3 ± 0.0	4.44 [ToxCastTM: BSK_LPS_MPC1_down; BSK_SAg_MCP1_down]	0.64738
CXCL10	1O7Y/P02778	-7.0 ± 0.1	4.44 [ToxCast™: BSK_BE3C_IP10_down]	0.64738
NR3C1	3K22/P04150	-7.6 ± 0.0	10.5 [ToxCast™: NVS_NR_hGR]	1.02119
CD40	3QD6/P25942	-7.9 ± 0.3	4.44 [ToxCast™: BSK_SAg_CD40_down]	0.64738
CAR (NR1I3)	1XVP/Q14994	-8.9 ± 0.0	0.11[ToxCast™: NVS_NR_hCAR]	-0.95468
RORC	3L0L/P51449	-7.5 ± 0.1	0.891 [ToxCast™: X]	-0.05012

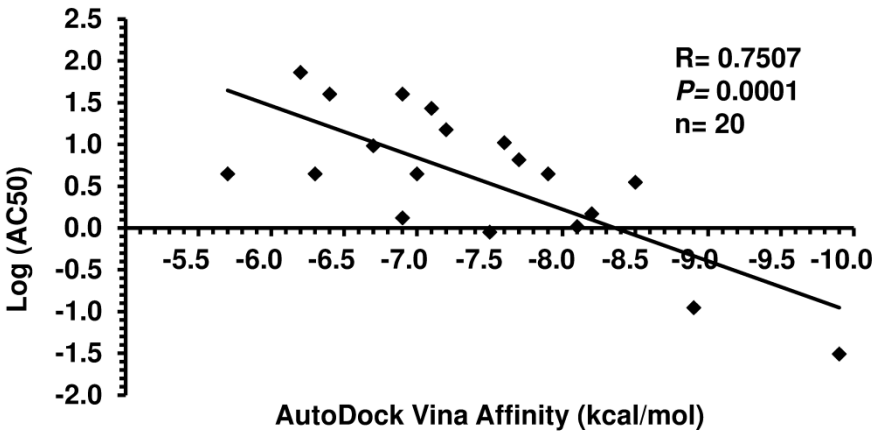


Figure 4.7. Calculated docking affinities (kcal/mol) of BPA for 20 proteins vs. the logarithm of the experimental AC_{50}.

4.3. DISCUSSION

The *in silico* docking calculations performed on BPA and target proteins, found by text-mining to have relationships with the plasticizer, revealed that several have the capacity to interact with this chemical *in silico*. This may suggest BPA could be responsible for mediating other biochemical responses besides the endocrine disruption.

Highest affinity scores for protein complexes non-related to the ER were found for three of the four members of the Cdc2-like kinase (CLK) family, which is conserved throughout the evolution[23], and it has been proposed to alter the function of the spliceosome by phosphorylating serine-arginine-rich (SR) proteins within the spliceosome assembly[24]. CLK1 and CLK4 kinase activity modulates alternative splicing events by phosphorylating some SR proteins[25]. The spliceosome regulates the processing, or splicing, of pre-mRNAs, yielding mature protein-encoding mRNAs. Many human genes express more than one mRNA *via* alternative splicing, leading to protein

diversity and it is well known that the misregulation of alternative splicing is involved in the pathogenesis of cancer and other diseases[24]. Although the predisposition to different types of cancer has been associated with BPA[26], mechanisms involving these proteins have not yet been considered. The theoretical approach used in this study revealed BPA may interact with proteins of the CLK family. Therefore, the evaluation of the role of these potential targets in the development of cancer and other diseases in organisms exposed to BPA is recommended.

The fact that circadian clock proteins could eventually work as BPA targets is quite interesting. CLK2 is regulated by feeding/fasting cycles in the liver, and upon activation by insulin/Akt signaling, it phosphorylates the SR domain on the peroxisome proliferator-activated receptor-γ coactivator (PGC-1α), causing potent repression of gluconeogenic gene expression and hepatic glucose output that leads to hypoglycemia[27]. Interestingly, the induction of hypoglycemia with an increase in the plasma insulin levels has been found in female rats and male mice acutely exposed to BPA[28]. Besides CLK proteins, several others seemed to be suitable targets for BPA. Although the theoretical affinity of these proteins was lower than those obtained for the CLK members, the values were better than those registered for the ER/BPA complex (-8.3 kcal/mol), increasing the chances that these proteins could be weak targets for BPA. These proteins include the nuclear receptor CAR, which may be involved in the development of certain diseases, including obesity, metabolic syndrome, diabetes[27] and cancer[27]. EAR-7, which has been recognized as a target for BPA, is an anti-thyroid hormonal endocrine disrupter that has shown an apparent antagonistic effect toward this receptor reducing their mRNA levels[29] . PKCθ, a pivotal protein in the insulin signaling pathway, is also important in T-cell activation[30], skeletal muscle signal transduction, and in neuronal differentiation. In T-cell receptor activation pathways, PKCθ serves as a positive regulator of cell survival and its inhibition results in p53-independent G1 cell-cycle arrest in various cancer cells[31]. ADORA2A, one of the serum proteins with good theoretical affinity for BPA, is localized in several neurons and has been shown to modulate the neurotransmission of the γ-aminobutyric acid (GABA), acetylcholine, and glutamate, contributing to motor behavior. Actually, its agonists have been

reported to inhibit locomotor activity and induce catalepsy in rodents, and in contrast, its antagonists prevent motor disturbances[32]. Another protein that exhibited good affinity *in silico* for BPA was TIMP-3. This protein is essential for normal physiology and has been reported to be down-regulated by BPA exposure[33]. Several authors suggest its participation in a number of pathological disturbances such as cancer, arthritis and cardiovascular diseases[34].

Although experimental verification must be performed to assess the likelihood that BPA could be targeting several of the proteins suggested here, it is quite promising to discover that the validation performed on existing BPA-binding proteins was successful to detect high ligand binding affinities (lower than -9.0 kcal/mol) for such complexes (ESRRG/BPA). Almost as important, was the fact that the *in silico* validation suggested BPA acquires a binding pose on ESRRG, which agrees with the crystallographic structure, and a very good correlation between predicted binding affinities by AutoDock Vina and the experimental activity data; obtaining the greatest affinity scores for those compounds with the lower AC_{50} and the poorest AutoDock Vina binding affinity for those with high values of AC_{50}. In addition, the BPA molecule shows drug-like features because of its size and chemical characteristics accordingly with the Lipinski's rule of five[35]. It may be of special interest for the possible development of new drugs to take BPA as a template, taking into account the number of proteins that may be interacting with it, and the multiple types of interactions that they could establish with the receptors, including hydrogen bonds, hydrophobic interactions and aromatic ring interactions.

In short, several proteins involved in important diseases such as obesity, metabolic syndrome, diabetes, motor disturbances, arthritis, cardiovascular diseases, and cancer[31-32, 34, 36], may be considered putative targets for BPA. This is interesting due to the fact that BPA exposure has been associated with several of these conditions in humans[37]. Therefore, BPA may be interacting with these proteins generating biochemical responses that could alter cellular functions predisposing the organism to the apparition of these pathologies.

4.4. CONCLUSIONS

This study suggests BPA may have some other targets different from ER, and it eventually could influence mechanisms of biochemical processes through the interaction with these proteins.

REFERENCES

1. Haighton, L. A.; Hlywka, J. J.; Doull, J.; Kroes, R.; Lynch, B. S.; Munro, I. C., An evaluation of the possible carcinogenicity of bisphenol A to humans. *Regul. Toxicol. Pharmacol.* **2002,** *35* (2), 238-254.
2. Brugnera, M. F.; Rajeshwar, K.; Cardoso, J. C.; Zanoni, M. V. B., Bisphenol A removal from wastewater using self-organized TIO2 nanotubular array electrodes. *Chemosphere* **2010,** *78* (5), 569-575.
3. Sakuma, S.; Nakanishi, M.; Morinaga, K.; Fujitake, M.; Wada, S.-i.; Fujimoto, Y., Bisphenol A 3,4-quinone induces the conversion of xanthine dehydrogenase into oxidase *in vitro*. *Food Chem. Toxicol.* **2010,** *48* (8–9), 2217-2222.
4. Avissar-Whiting, M.; Veiga, K. R.; Uhl, K. M.; Maccani, M. A.; Gagne, L. A.; Moen, E. L.; Marsit, C. J., Bisphenol A exposure leads to specific microRNA alterations in placental cells. *Reprod. Toxicol.* **2010,** *29* (4), 401-406.
5. Berger, R. G.; Shaw, J.; deCatanzaro, D., Impact of acute bisphenol-A exposure upon intrauterine implantation of fertilized ova and urinary levels of progesterone and 17ß-estradiol. *Reprod. Toxicol.* **2008,** *26* (2), 94-99.
6. Bredhult, C.; Sahlin, L.; Olovsson, M., Gene expression analysis of human endometrial endothelial cells exposed to Bisphenol A. *Reprod. Toxicol.* **2009,** *28* (1), 18-25.
7. Balakrishnan, A.; Stearns, A. T.; Rounds, J.; Irani, J.; Giuffrida, M.; Rhoads, D. B.; Ashley, S. W.; Tavakkolizadeh, A., Diurnal rhythmicity in glucose uptake is mediated by temporal periodicity in the expression of the sodium-glucose cotransporter (SGLT1). *Surgery* **2008,** *143* (6), 813-818.
8. Rezg, R.; El-Fazaa, S.; Gharbi, N.; Mornagui, B., Bisphenol A and human chronic diseases: Current evidences, possible mechanisms, and future perspectives. *Environ. Int.* **2014,** *64*, 83-90.
9. Chen, H.; Sharp, B., Content-rich biological network constructed by mining PubMed abstracts. *BMC Bioinformatics* **2004,** *5* (1), 147.
10. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The protein data bank. *Nucleic Acids Res.* **2000,** *28* (1), 235-242.
11. Hetényi, C.; van der Spoel, D., Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Sci.* **2002,** *11* (7), 1729-1737.
12. Sarsam, S. W.; Nutt, D. R.; Strohfeldt, K.; Watson, K. A., Titanocene anticancer complexes and their binding mode of action to human serum albumin: A computational study. *Metallomics* **2011,** *3* (2), 152-161.

13. Maldonado-Rojas, W.; Olivero-Verbel, J.; Ortega-Zuñiga, C., Searching of protein targets for alpha lipoic acid. *J. Braz. Chem. Soc.* **2011,** *22*, 2250-2259.
14. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009,** *30* (16), 2785-2791.
15. Lim, S. V.; Rahman, M. B.; Tejo, B., Structure-based and ligand-based virtual screening of novel methyltransferase inhibitors of the dengue virus. *BMC Bioinformatics* **2011,** *12* (Suppl 13), S24.
16. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Laham, A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A., Gaussian 03, Revision C.02. 2003.
17. Guha, R.; Howard, M. T.; Hutchison, G. R.; Murray-Rust, P.; Rzepa, H.; Steinbeck, C.; Wegner, J.; Willighagen, E. L., The Blue ObeliskInteroperability in Chemical Informatics. *J. Chem. Inf. Model.* **2006,** *46* (3), 991-998.
18. Ranjan, N.; Andreasen, K. F.; Kumar, S.; Hyde-Volpe, D.; Arya, D. P., Aminoglycoside Binding to Oxytricha nova Telomeric DNA. *Biochemistry* **2010,** *49* (45), 9891-9903.
19. Trott, O.; Olson, A. J., AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010,** *31* (2), 455-461.
20. Wolber, G.; Langer, T., LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. *J. Chem. Inf. Model.* **2004,** *45* (1), 160-169.
21. Durdagi, S.; Duff, H. J.; Noskov, S. Y., Combined Receptor and Ligand-Based Approach to the Universal Pharmacophore Model Development for Studies of Drug Blockade to the hERG1 Pore Domain. *J. Chem. Inf. Model.* **2011,** *51* (2), 463-474.
22. Takayanagi, S.; Tokunaga, T.; Liu, X.; Okada, H.; Matsushima, A.; Shimohigashi, Y., Endocrine disruptor bisphenol A strongly binds to human estrogen-related receptor γ (ERRγ) with high constitutive activity. *Toxicol. lett.* **2006,** *167* (2), 95-105.
23. Nashev, L. G.; Schuster, D.; Laggner, C.; Sodha, S.; Langer, T.; Wolber, G.; Odermatt, A., The UV-filter benzophenone-1 inhibits 17β-hydroxysteroid dehydrogenase type 3: Virtual screening as a strategy to identify potential endocrine disrupting chemicals. *Biochem. Pharmacol.* **2010,** *79* (8), 1189-1199.

24. Rosenthal, A. S.; Tanega, C.; Shen, M.; Mott, B. T.; Bougie, J. M.; Nguyen, D.-T.; Misteli, T.; Auld, D. S.; Maloney, D. J.; Thomas, C. J., Potent and selective small molecule inhibitors of specific isoforms of Cdc2-like kinases (Clk) and dual specificity tyrosine-phosphorylation-regulated kinases (Dyrk). *Bioorg. Med. Chem. Lett.* **2011,** *21* (10), 3152-3158.
25. Ogawa, Y.; Hagiwara, M., Challenges to congenital genetic disorders with "RNA-targeting" chemical compounds. *Pharmacol. Ther.* **2012,** *134* (3), 298-305.
26. Huang, Z.; Chen, G.; Shi, P., Emodin-induced apoptosis in human breast cancer BCap-37 cells through the mitochondrial signaling pathway. *Arch. Pharm. Res.* **2008,** *31* (6), 742-748.
27. Rodgers, J. T.; Haas, W.; Gygi, S. P.; Puigserver, P., Cdc2-like Kinase 2 Is an Insulin-Regulated Suppressor of Hepatic Gluconeogenesis. *Cell Metab.* **2010,** *11* (1), 23-34.
28. (a) Neel, B. A.; Sargis, R. M., The paradox of progress: environmental disruption of metabolism and the diabetes epidemic. *Diabetes* **2011,** *60* (7), 1838-1848; (b) Jubendradass, R.; D'Cruz, S. C.; Mathur, P. P., Short-term exposure to nonylphenol induces pancreatic oxidative stress and alters liver glucose metabolism in adult female rats. *J. Biochem. Mol. Toxicol.* **2011,** *25* (2), 77-83.
29. Iwamuro, S.; Yamada, M.; Kato, M.; Kikuyama, S., Effects of bisphenol A on thyroid hormone-dependent up-regulation of thyroid hormone receptor α and ß and down-regulation of retinoid X receptor γ in Xenopus tail culture. *Life Sci.* **2006,** *79* (23), 2165-2171.
30. Lee, H.-S.; Kim, Y.-D.; Na, B.-R.; Kim, H.-R.; Choi, E.-J.; Han, W.-C.; Choi, H.-K.; Lee, S.-H.; Jun, C.-D., Phytocomponent p-Hydroxycinnamic acid inhibits T-cell activation by modulation of protein kinase C-θ-dependent pathway. *Int. Immunopharmacol.* **2012,** *12* (1), 131-138.
31. Chu, P.-Y.; Hsu, N. C.-H.; Tai, H.-C.; Yeh, C.-M.; Lin, S.-H.; Hou, M.-F.; Yeh, K.-T., High nuclear protein kinase Cθ expression may correlate with disease recurrence and poor survival in oral squamous cell carcinoma. *Hum. Pathol.* **2012,** *43* (2), 276-281.
32. Matasi, J. J.; Caldwell, J. P.; Hao, J.; Neustadt, B.; Arik, L.; Foster, C. J.; Lachowicz, J.; Tulshian, D. B., The discovery and synthesis of novel adenosine receptor (A 2A) antagonists. *Bioorg. Med. Chem. Lett.* **2005,** *15* (5), 1333-1336.
33. Fukushima, A.; Funabashi, T.; Kawaguchi, M.; Mitsushima, D.; Kimura, F., Bisphenol A induces transforming growth factor-ß3 mRNA in the preoptic area: A cDNA expression array and Northern blot study. *Neurosci. Lett.* **2007,** *411* (1), 81-85.
34. Fogarasi, M.; Janssen, A.; Weber, B. H. F.; Stöhr, H., Molecular dissection of TIMP3 mutation S156C associated with Sorsby fundus dystrophy. *Matrix Biol.* **2008,** *27* (5), 381-392.
35. Keller, T. H.; Pichota, A.; Yin, Z., A practical view of 'druggability'. *Curr. Opin. Chem. Biol.* **2006,** *10* (4), 357-361.
36. (a) Chen, Y.; Tang, Y.; Guo, C.; Wang, J.; Boral, D.; Nie, D., Nuclear receptors in the multidrug resistance through the regulation of drug-metabolizing enzymes and drug transporters. *Biochem. Pharmacol.* **2012,** *83* (8), 1112-1126; (b) Dring, A. M.; Anderson, L.

E.; Qamar, S.; Stoner, M. A., Rational quantitative structure–activity relationship (RQSAR) screen for PXR and CAR isoform-specific nuclear receptor ligands. *Chem. Biol. Interact.* **2010,** *188* (3), 512-525.

37. Huang, Y. Q.; Wong, C. K. C.; Zheng, J. S.; Bouwman, H.; Barra, R.; Wahlström, B.; Neretin, L.; Wong, M. H., Bisphenol A (BPA) in China: a review of sources, environmental levels, and potential human health impacts. *Environ. Int.* **2012,** *42*, 91-99.

CHAPTER 5

Urban EDCs targeting breast cancer

Breast cancer is a multifactorial disease, including chronic exposure to chemicals, such as EDCs that have been widely documented to increase the risk of this disease[1] as well as other reproductive disorders and different cancer types[2]. In this chapter a vHTS is presented to predict the protein-ligand interaction between proteins involved in the development of this disease and EDCs. The *in vitro* validation of the interactions between one of the predicted complexes, SHBG/BPA, was also carried out.

The *in silico* approach is useful to address the topic of EDCs, due to the growing need to screen a large number of chemicals in order to determine their adverse biological activity[3], as a tool for prioritizing chemicals for *in vitro* screening[4] and whole animal tests[5]. Historically, computational toxicology has been most commonly used in ecological risk assessment or safety evaluation of pharmaceutical drugs[6], but recently the advances in the determination of protein structures and high-performance computing has refocused attention on the virtual screening and automated docking methods of environmental chemicals[7], as this strategy represent a valid and fast option to analyze a large number of molecules, focusing on their interaction with critical disease pathways[8], in very short times[9].

5.1. MATERIALS AND METHODS

A four-step approach has been used to identify EDCs with potential to target breast cancer proteins. This includes the selection of the proteins by data mining tools, vHTS, evaluation of *in silico* interactions and an experimental validation by spectroscopic methods.

5.1.1. Text mining

Proteomics studies of breast cancer have provided relevant information regarding the identification of proteins of interest in the diagnosis and treatment of this disease[10], even though there is not a consensus list or repository with this information. Therefore, a meta-analysis was performed to identify proteins and genes associated with breast cancer in PubMed (http://www.ncbi.nlm.nih.gov/pubmed/), using the text mining tool Fable (http://fable.chop.edu/), employing the key words "breast cancer". The results were ranked according to their number of hits, and a cutoff of ≥100 references related to breast cancer in PubMed was used. This methodology was carried out in order to include the proteins more frequently cited with a role in breast cancer. Some important nuclear receptors and cell cycle targets were also included, as they have been reported to be deregulated in human breast cancer[11].

5.1.2. High-throughput virtual screening

Molecular docking simulations were carried out to find EDCs with the potential to target breast cancer proteins *in silico*. Calculations were run on Linux RHEL6 for IBM Power, utilizing a System X rack with Power 730 processors. AutoDock Vina[12], an open-source program for molecular docking and virtual screening, was used to assess the binding affinity of each EDC/protein complex, as this software has been largely employed and formally validated for target identification of these kind of molecules, having a high prediction performance when compared to experimental data[13], not only regarding the binding affinity but also the crystallographic binding modes[14]. A total of 305 EDCs from urban sources were downloaded from EDCs DataBank[15] and docked with the selected breast cancer proteins, using the same protocol reported in our previous work for inverse virtual screening[11b]. The crystallographic coordinates of the proteins, with resolution suitable for docking studies (in average, 2.0 Å), were downloaded from PDB[16] in pdb text format file which were then employed for preprocessing in Sybyl-X 2.0 program package (Tripos, St. Louis, MO). All ions, water molecules, and other substructures were removed from the coordinated files, and the biopolymer structure preparation tool was used for analyzing and fixing the

3D-structures with default settings. Optimization of the proteins was also carried out by the same software package employing the Powell method, Kollman united and Kollman all atoms force fields, AMBER charges, dielectric constant 1.0, NB cutoff 8.0, maximum interactions 100 and termination gradient 0.001 kcal/mol. The resultant pdb file was then submitted to AutoDock Tools[17] for preparing the grid parameters and the required pdbqt files for the docking studies. Kollman charges and polar hydrogen atoms were added to the three-dimensional structures of the proteins, the grid was centered in the macromolecule and the size adjusted to include the whole protein surface with a spacing of 0.357 Å. The structures of the EDCs were directly downloaded from EDCs DataBank[15], and used for virtual screening in Autodock Vina[12]. Each ligand/protein pair was docked by triplicate and the best affinity scores of the single runs were then used to calculate the averages and the results ranked according to these values.

The EDCs were separated according to their source of exposure in four groups: dioxins and related molecules, plastics and other types of polymers, everyday products and miscellaneous. The results ranked and the best protein/ligand complexes of each category were then selected to determine their non-covalent interactions by computer aided simulations. In addition, a hierarchical clustering, which is a powerful tool to blindly explore proteomic data[18], was performed in R (http://www.r-project.org/) for each group, using the "heatmap.2" function of the gplot library[19]. The color key of the heat map was selected using the RColorBrewer package of R, to present in red the protein-ligand pairs with strong and moderate docking affinity (-15.0 to -8.0 kcal/mol), in white those with weak affinity (-8.0 to -7.0 kcal/mol) and in blue the proteins-ligand pairs that are not likely to interact according to the simulation (-7.0 to 5.0 kcal/mol). As a result, patterns in the behavior of EDCs against breast cancer proteins and vice versa were identified. The employed affinity cutoff (< -8.0 kcal/mol) was the same used in the protocol of our previous studies with AutoDock Vina[12], which allowed a good correlation between *in silico* binding affinity and experimental data[14, 20].

5.1.3. Evaluation of protein-ligand interactions

An *in silico* approach was employed to evaluate the contact residues participating in the protein-ligand interaction in selected complexes formed between breast cancer related proteins and EDCs. This was achieved using LigandScout 3.1[21]. The best docking pose of the EDC resulting from the docking with the target protein by AutoDock Vina[12] was isolated in AutoDock Tools[17] and merged with the optimized file of the protein structure in pdb format by Sybyl X-2.0 (Tripos, St. Louis, MO). This file was then used as input in LigandScout, the parameters utilized for the pharmacophore features and interacting residues were those established by default in the program.

In order to validate our protocol, a protein-ligand pair, was selected for studying its interactions *in silico* and *in vitro*. The criteria used to pick the complex that exhibited an affinity score near the cutoff (around -8.0 kcal/mol), a common EDC with generalized exposure in the population and a protein with certain impact in breast cancer.

5.1.4. Experimental validation of the existence of protein-ligand interactions for a predicted EDC/protein complex

The protein ligand interaction between BPA, one of the most common EDCs used in plastics[22], and the steroid transporter protein SHBG was analyzed by LigandScout software[21], using the same protocol described above; and the validation of the binding and conformational changes of this complex was achieved by spectroscopic methods[23], employing MST and CD. The protein was obtained by recombinant DNA technology and BPA (≥99%; 239658-50G) was purchased from Sigma Aldrich.

5.1.5. Protein expression and purification

The gene for human SHBG (PDB: 1F5F) (obtained from Eurofins, Ebersberg) was inserted into the blank plasmid pET15-MHL and harbored by *E. coli* DH5α cells (Figure 5.1). Then, one of the recombinant colonies was chosen to growth the pET15MHL-SHBG plasmid, as it appears in Figure 5.2. The plasmid was isolated from the Mini-prep kit/Invitrogen following the instructions of the supplier and transformed into chemically competent *E. coli*

C41(DE3), with a similar procedure presented in Figure 5.1 but using this strain. Transformed cells were plated onto LB agar containing 100 µg/mL of ampicillin, and incubated overnight at 37 °C. In order to get the starting solution, a single colony was used to inoculate 100 mL of LB/ampicillin for overnight growth, with a similar procedure showed in Figure 5.2 but using the recombinant colony *E. coli* C41(DE3)/pET15MHL-SHBG. Then, 20 mL of the starting solution was used to inoculate 400 mL of LB/ampicillin medium. The culture was incubated at 37 °C (180 rpm) until the A_{600} was 0.6–0.8 at which point protein expression was induced by the addition of isopropyl 1-thio-ß-d-galactopyranoside (IPTG) to a final concentration of 0.5 mM.

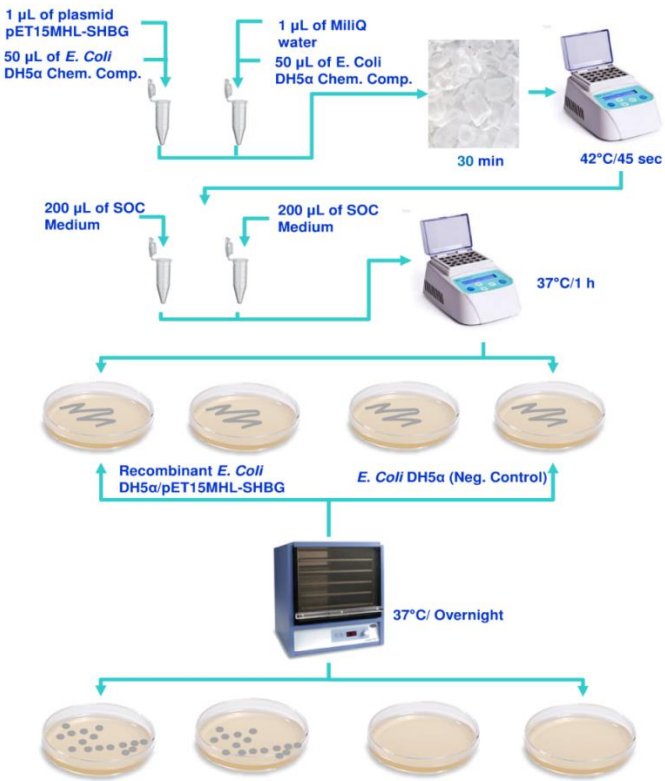


Figure 5.1. Workflow for the E. coli DH5a transformation with pET15MHL-SHBG.

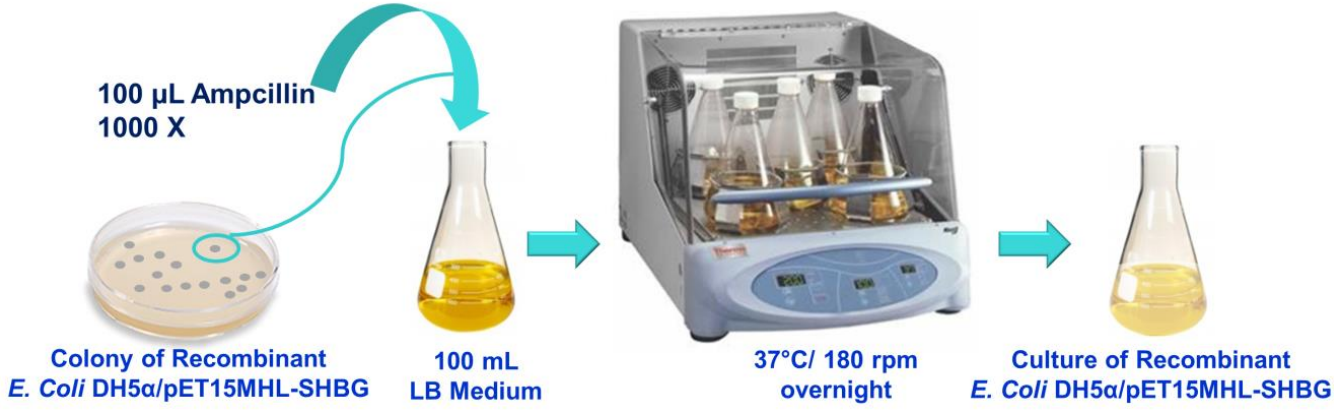


Figure 5.2. Recombinant plasmid growth on E. coli DH5α/pET15MHL-SHBG.

After overnight incubation at 20 °C[24], the induced cells were harvested by centrifugation, the pellet was resuspended in cold PBS (phosphate-buffered saline) buffer and washed twice, resuspended again in 15 mL of cold lysis buffer (150 mM sodium chloride, 1.0% Triton X-100, 50 mM Tris, pH 8.0), maintained in agitation 30 min (4°C), and centrifuged (20 min, 12,000 rpm). This pellet was discarded and the supernatant was placed on ice, and stored at -20 °C until protein purification (http://www.abcam.com/index.html?pageconfig=resource&rid=11379).

Magnetic beads, Dynabeads His-Tag Isolation and Pulldown (Life technologies), were used for protein purification according to the protocol provided by the supplier (http://tools.lifetechnologies.com/content/sfs/manuals/DynabeadsHisTagIsolationPulldown_man.pdf). A dialysis procedure was also employed to remove small molecules using a slide-A-Lyzer Dialysis Cassette, 20K molecular weight cutoff (MWCO; Thermo scientific) according to the protocol provided by the supplier.

Then, the protein purity was checked by SDS-Page, using a 4-12% NuPAGE Bis-Tris gel (Invitrogen, UK) and SeeBlue Plus2 prestained protein standard (Invitrogen, UK)[25]. During the protein expression, this technique was also applied to monitor the presence and quantity of the recombinant protein, as well as a western blot using a monoclonal mouse anti-His tag antibody employing an standard protocol[26]. The chemiluminescence signal was detected by using a Typhoon Trio Imager (GE Healthcare Life Sciences).

5.1.6. Protein storage

In order to avoid frequent freezing and thawing, the protein sample was divided into aliquots, stored frozen at -20°C, and kept at physiological pH (PBS buffer pH 7.4).

5.1.7. Microscale Thermophoresis

With the purpose of evaluating the likelihood of a possible functionality of BPA binding on SHBG, a qualitative MST analysis was carried out. This method reflects the directed movement of particles in a microscopic temperature gradient, and enables the analysis of molecular interactions in solution at microliter scale[27], with low sample consumption[28].

The SHBG protein was labeled using the blue fluorescent dye NT-495-NHS (NanoTemper Technologies) and incubated for 1 hour at room temperature. In order to separate the free dye from the protein, a superdex 200 10/300GL column and an AKTA PURE FPLC (fast protein liquid chromatography) system (GE Healthcare) were utilized with PBS buffer pH 7.4 at 0.5 mL/s flow rate. Fractions of 0.25 mL were collected with a fraction collector F9-R (GE Healthcare). The labeled SHBG was incubated for 5 min at room temperature with different concentrations of BPA in PBS containing 5% ethanol as vehicle, testing 1:2 concentrations from 0.03 to 1000 nM (1 μM) of BPA. The procedure to make the dilutions was carried out as shown in Figure 5.3.

The parameters used for the MST calculations were: Led color blue, cap from 1 to 16, MST power 40 and Led power 40. The samples (3–5 μL) were loaded into hydrophilic glass capillaries (Monolith NT Capillaries) and the thermophoresis analysis was performed (LED 40%, IR laser 20%) using a NanoTemper Monolith NT.115 (NanoTemper Technologies) instrument. The capillary used for the MST analysis with BPA and SHBG was the hydrophilic one, as this generated the expected Gaussian curve.

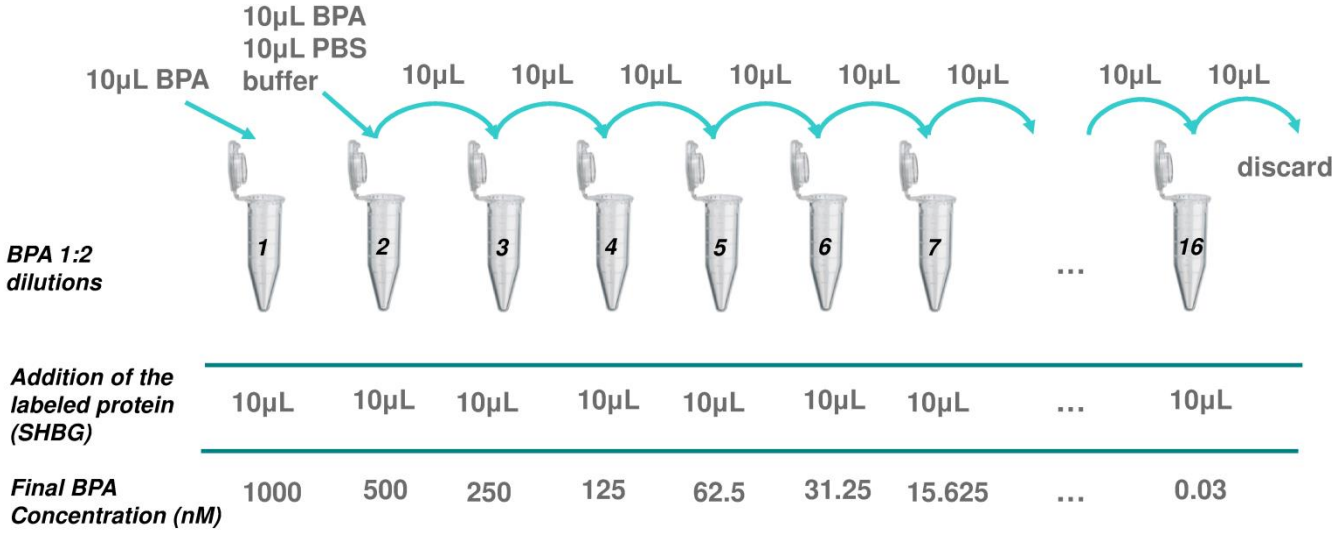


Figure 5.3. MST sample preparation.

5.1.8. Circular Dichroism

The spectra of SHBG (1 µM in Tris pH 7.4) was analyzed to characterize the non-bonded folding state of the recombinant protein, by comparison with the information provided in PDB[16]. The changes in the secondary structure of the protein after BPA binding were accessed by recording the CD spectra of SHBG with two final concentrations of BPA (1.5×10^{-5} and 4.5×10^{-5} M) at time zero and after overnight incubation, the stock solution consisted of 30 mM BPA in absolute ethanol. The dilutions were prepared adding 0.2 and 0.6 µL of the BPA stock solution to 400 µL of SHBG (1 µM in Tris pH 7.4), respectively. The Tris buffer was used as background and subtracted from all runs.

The CD spectra were recorded using the applied photophysics Chriscan, wavelength (nm): 190-350; step: 0.5; time per point (s): 0.5 repeats: 5, temperature: 25°C and auto-subtraction of the background. All the spectra were recorded in both delta A and mdeg units, with 5 repetitions, using a quartz cuvette of 10 mm path length, volume: 0.7 mL, inside: 2 mm; spectral range 190-2500 nm; 4 transparent windows and dimensions 45 mm x 12.5 mm x 12.5 mm, in a nitrogen atmosphere[29]. The data was recorded using the programs Prodata viewer, Chirascan and APLData Converter.

The circular dichroism spectra were saved as ASCII files (.kin file format), employing the Pro-data Chriscan software, and utilized as input files in DichroWeb[30] for the secondary structure determination. The parameters used were: file format=free, input units=milidegrees, initial wavelength=190, final wavelength=350, wavelength step=0.5, lowest datapoint to use in the analysis=190, analysis program=CONTIN, reference set=SMP 180 (optimized for 190-240 nm), optional scaling factor=1 and output units=delta epsilon.

A statistical analysis of the results of the CD was performed by GraphPad Prism 6.01 (GraphPad Software Inc., San Diego, CA, USA). A non-parametric multiple comparison test Kruskal-Wallis was carried out using the raw data of the spectra recorded for SHBG and SHBG/BPA. A post-hoc Dunn's multiple comparison test of the row data based on Kruskal-Wallis analysis was also applied to compare the medians of the individual groups. Statistical significance was accepted at $p \leq 0.05$.

5.2. RESULTS

5.2.1. Text mining

Based on the data mining performed on Fable, 294 genes/proteins were found to be related to breast cancer (≥100 citations in PubMed) (Annex 5.1), and a total of 133 were suitable for docking studies and employed for further analysis. This selection was complemented by adding proteins recently discovered as important in breast cancer, as well as with nuclear receptors and cell cycle regulation proteins. Finally, a total of 189 proteins were selected for the vHTS with EDCs.

5.2.2. High-throughput virtual screening

A total of 305 EDCs were docked with the proteins related to breast cancer selected in the previous step. The EDCs were classified according to the potential sources of exposures in four groups: dioxins and related molecules, plastics and other types of polymers, everyday products and miscellaneous. Results were ranked according to the affinity scores obtained in AutoDock Vina[12]. Clustered heat maps of the affinity scores of EDCs against breast cancer proteins are presented for each category as supplementary material (Annexes 5.2-5.5).

The virtual screening allowed the identification of protein-ligand pairs with high affinity, candidates to be tested *in vitro* and *in vivo* for breast cancer in further studies. In the following sections, the results in each group of EDCs, according to the exposure source, are presented.

5.2.3. Dioxins and related molecules

The main cluster in the heat map of the virtual screening of breast cancer proteins with dioxins and related molecules (Annex 5.1 and 5.2) indicates that these interact mostly with a group of proteins, including catalase (CAT); RARB; SHBG; PGR; phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform (PIK3CA); nuclear receptor ROR-gamma (RORC); nuclear receptor ROR-alpha (RORA); adiponectin (ADIPOQ); cytochrome P450 3A4 (CYP3A4); phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit gamma isoform (PIK3CG) and apolipoprotein D (APOD). Therefore, the underlying mechanisms by which these EDCs elicit their effects could be associated to signaling pathways involved in cell proliferation and oxidative stress, among others.

The complexes that presented the highest affinity score in this group are CAT/1,3,7,8-tetrachlorodibenzo-p-dioxin (Figure 5.4), and 2,3,4,7-tetrachlorodibenzofuran with the protein cytochrome P450 1A2 (CYP1A2, Figure 5.5). The top ranking of the complexes formed by EDCs in this group with breast cancer proteins is showed in Table 5.1.

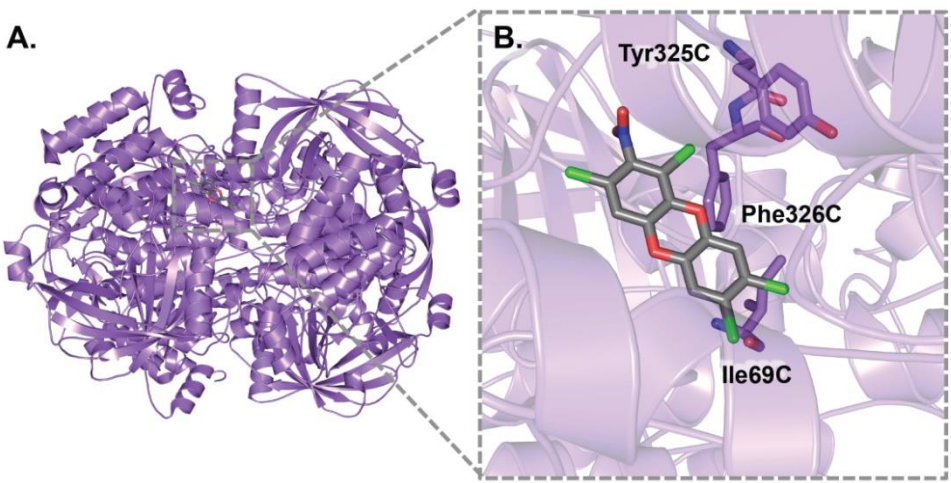


Figure 5.4. Three-dimensional view of the (A) CAT/1,3,7,8-tetrachlorodibenzo-p-dioxin, (B) showing the binding site and interactions predicted by LigandScout 3.1.

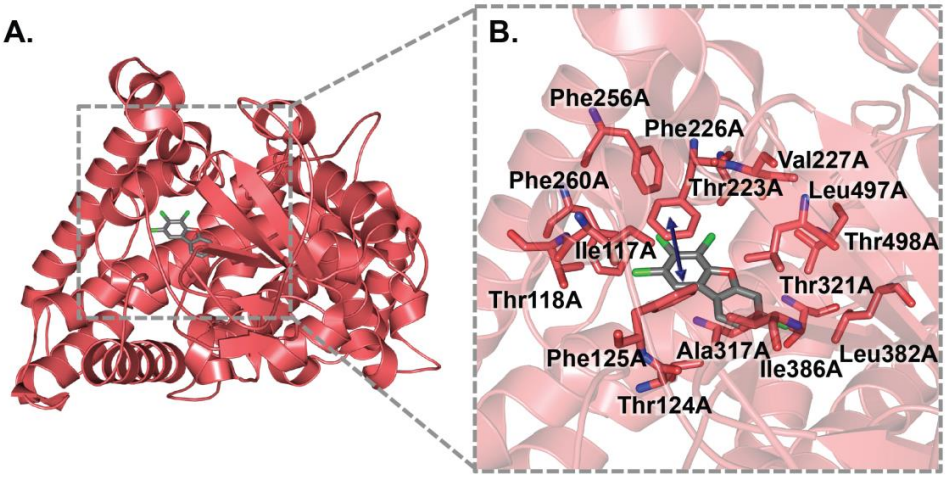


Figure 5.5. Three-dimensional view of the (A) CYP1A2/2,3,4,7-tetrachlorodibenzofuran, (B) showing the binding site and interactions predicted by LigandScout 3.1. The blue arrows represent aromatic ring interactions.

Table 5.1. Dioxins and related molecules with best affinity scores (<-10.0 kcal/mol) after docking with breast cancer associated proteins.

Short name	PDB	EDCs	CID	Affinity (kcal/mol)
CAT	1DGF	1,3,7,8-tetrachlorodibenzo-p-dioxin	149104	-10.3
CYP1A2	2HI4	2,3,4,7-tetrachlorodibenzofuran	55111	-10.2

5.2.4. Plastics and other types of polymers

The docking simulation showed that several EDCs in this category have the potential to bind breast cancer proteins (Annexes 5.1. and 5.3). However, the derivatives of the well-known plasticizer BPA presented the best affinities, being located the top of the ranking table (Table 5.2).

Table 5.2. EDCs in plastics and other types of polymers with best affinity scores (<-10.0 kcal/mol) after docking with breast cancer associated proteins.

Short name	PDB	EDCs	CID	Affinity (kcal/mol)
SRC	2H8H	bisphenol M	3292100	-10.8
ESRRG	2E2R	bisphenol AF	73864	-10.8
RXRB	1H9U	bisphenol M	3292100	-10.7
VDR	1IE9	bisphenol A dimethacrylate	76739	-10.3
VDR	1IE9	bisphenol M	3292100	-10.3
SHBG	1F5F	bisphenol M	3292100	-10.3
RARB	4DM6	bisphenol M	3292100	-10.2
ESRRG	2E2R	bisphenol B	66166	-10.2
CYP1A2	2HI4	dihydroxymethoxychlor olefin	84677	-10.1
HBA1	1A01	2,4,6-triphenyl-1-hexene	45356241	-10.1
CYP1A2	2HI4	diphenyl-p-phenylenediamine	6319	-10.1

The chemical structures of BPA and its analogs are presented in Figure 5.6, as well as their potential targets in breast cancer according to the vHTS, and a functional protein association network generated by STRING[31].

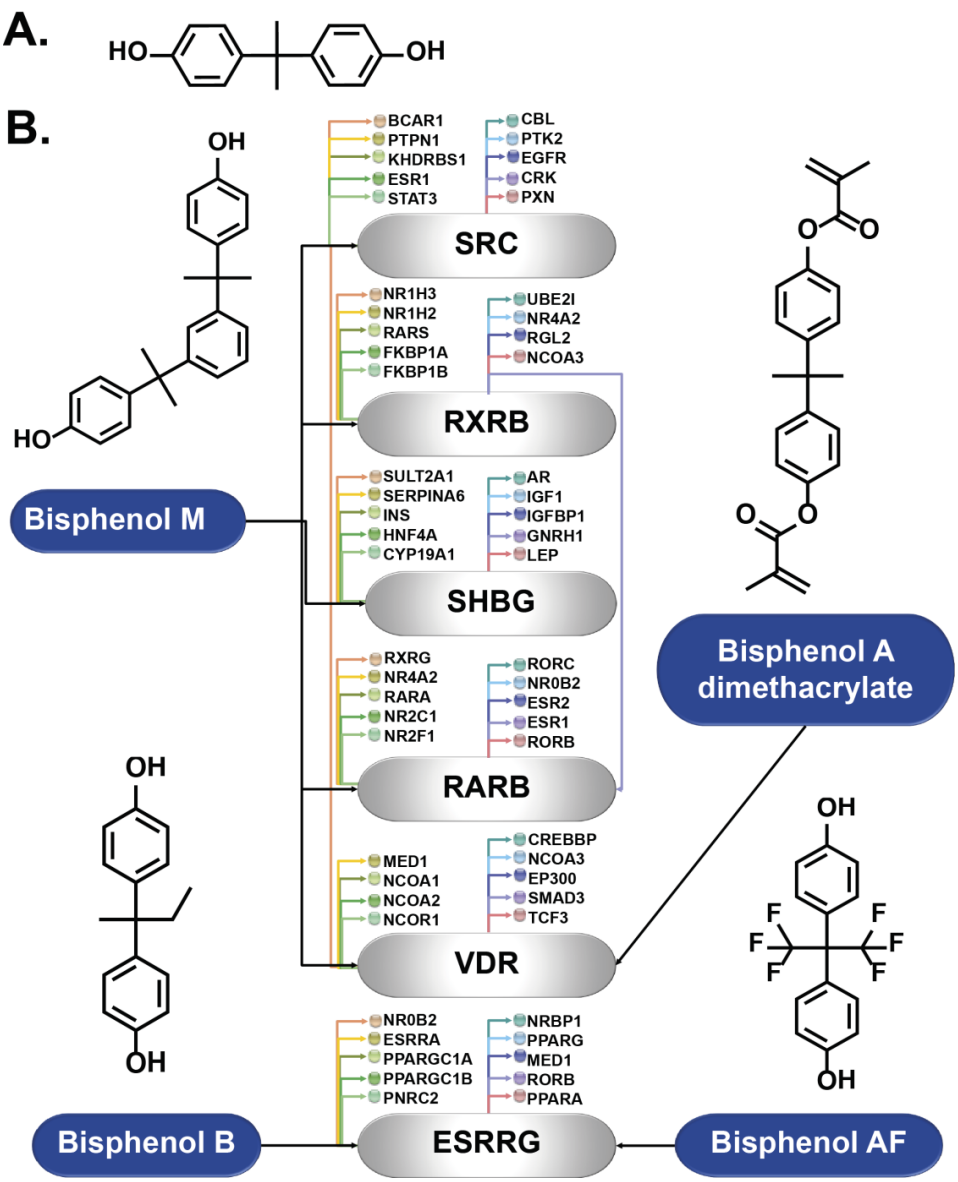


Figure 5.6. (A) Chemical structure of BPA and (B) its analogs showing the interactions with breast cancer proteins according to docking studies, and the analysis of the protein-protein interactions by STRING v.10[31].

The two complexes with the best affinity scores were proto-oncogene tyrosine-protein kinase Src (SRC)/bisphenol M and ESRRG/bisphenol AF (-10.8 kcal/mol). This is an interesting finding, as bisphenol AF (BPAF), is considered a new bisphenol analogue used as raw material in plastic industry; however, little is known about its occurrence in the environment and the potential associated risk[32]. The three-dimensional view of these complexes and their interactions are presented in Figures 5.7-5.8 Most of the interactions were hydrophobic in nature; however there are some hydrogen bond donor features that suggest the presence of hydrogen bonds between the ligand and the surrounding residues.

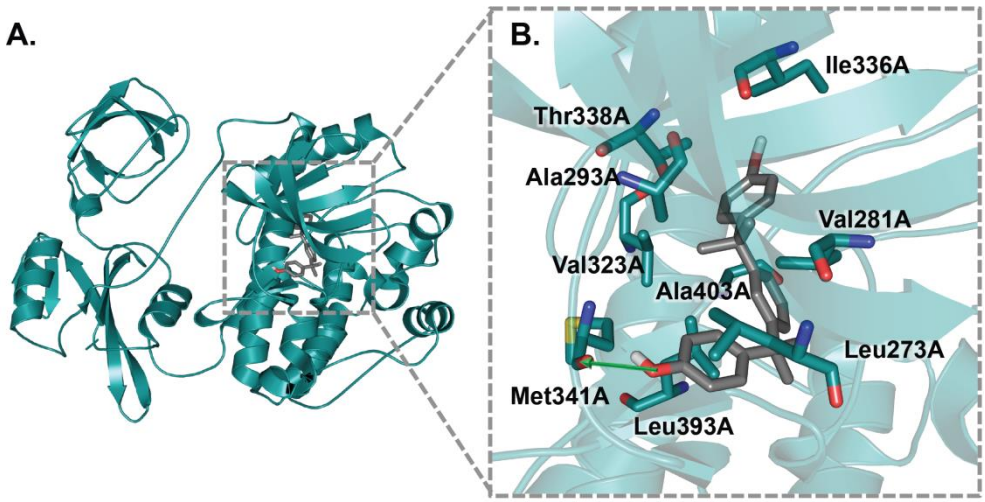


Figure 5.7. Three-dimensional view of the (A) SRC/bisphenol M complex, (B) showing the binding site and interactions predicted by LigandScout 3.1. The green arrows represent hydrogen-bond donor features.

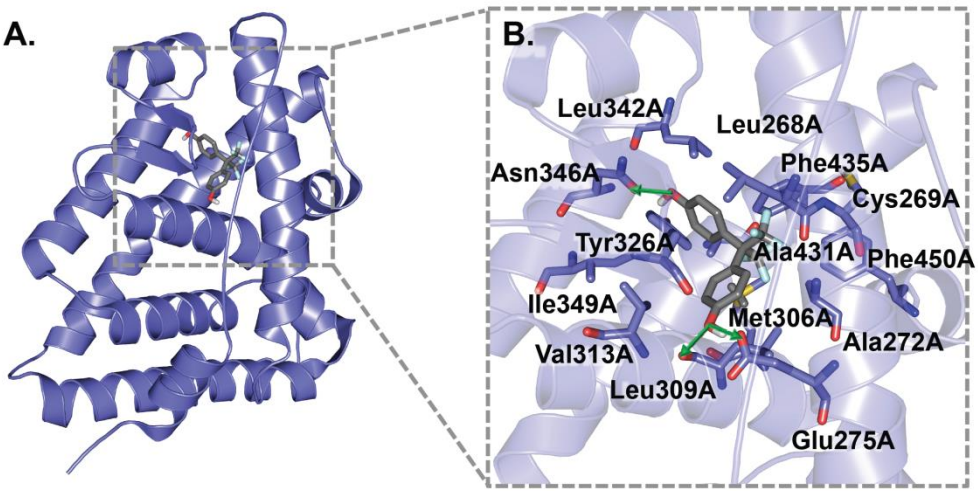


Figure 5.8. Three-dimensional view of the (A) ESRRG/bisphenol AF complex, (B) showing the binding site and interactions predicted by LigandScout 3.1. The green arrows represent hydrogen-bond donor features.

5.2.5. Everyday products

A broad range of EDCs presents in cosmetics, household products, drugs and personal care products, among others, exhibited good affinity for proteins involved in breast cancer (Annexes 5.1 and 5.4). The complexes with the best affinity scores in this group were the PFOS with the hormone transporter SHBG (Figure 5.9), and the RARB/AHTN complex (Figure 5.10).

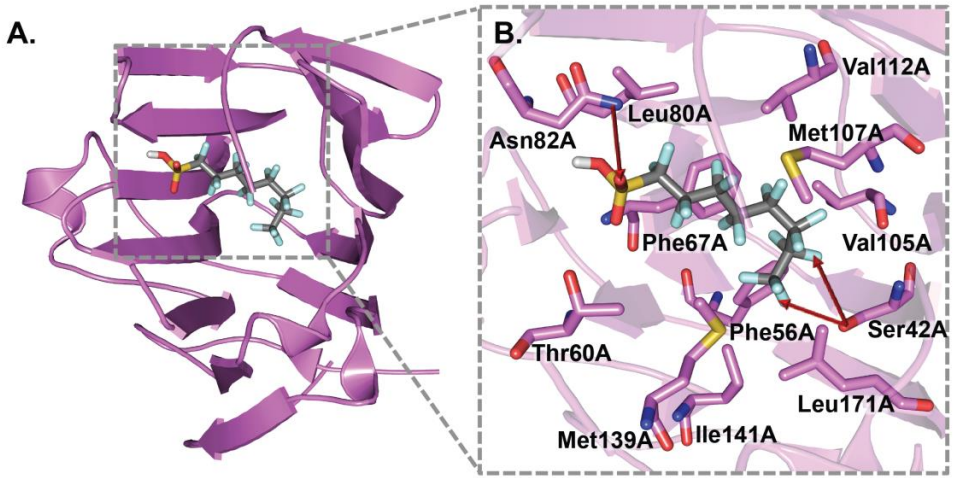


Figure 5.9. Three-dimensional view of the (A) SHBG/PFOS complex, (B) showing the binding site and interactions predicted by LigandScout 3.1. The red arrows represent hydrogen-bond acceptor features.

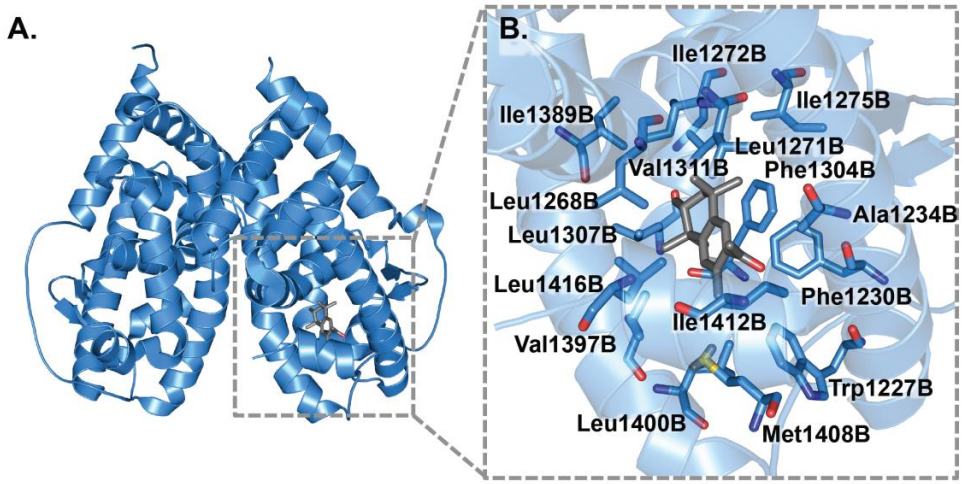


Figure 5.10. Three-dimensional view of the (A) RARB/AHTN complex, (B) showing the binding site and interactions predicted by LigandScout 3.1.

Other complexes that also exhibited good affinity scores are presented in Table 5.3.

Table 5.3. EDCs in everyday products with best affinity scores (<-10.0 kcal/mol) after docking with breast cancer associated proteins.

Short name	PDB	EDCs	CID	Keywords	Affinity (kcal/mol)
SHBG	1F5F	perfluorooctane sulfonic acid	74483	Household products	-10.4
RARB	4DM6	AHTN	89440	Cosmetics	-10.4
GSTP1	3N9J	Emodin	3220	Drugs	-10.2
VDR	1IE9	3-(4-methylbenzylidene)camphor	6434217	Cosmetics	-10.2
SHBG	1F5F	3-(4-methylbenzylidene)camphor	6434217	Cosmetics	-10.1
SERPINB5	1WZ9	Triclocarban	7547	Antimicrobial, personal care products, household products	-10.1
CAT	1DGF	Emodin	3220	Drugs	-10.1

5.2.6. Miscellaneous

The molecules that exhibited the best affinities in this category were mostly related to combustion, some of them are 3-hydroxy-benzo(a)pyrene, (benzo(a)pyrene and 6-hydroxychrysene (Annex 5.1 and 5.5). The best affinity score was predicted for the complex CYP1A2/benzo(a)pyrene (Figure 5.11).

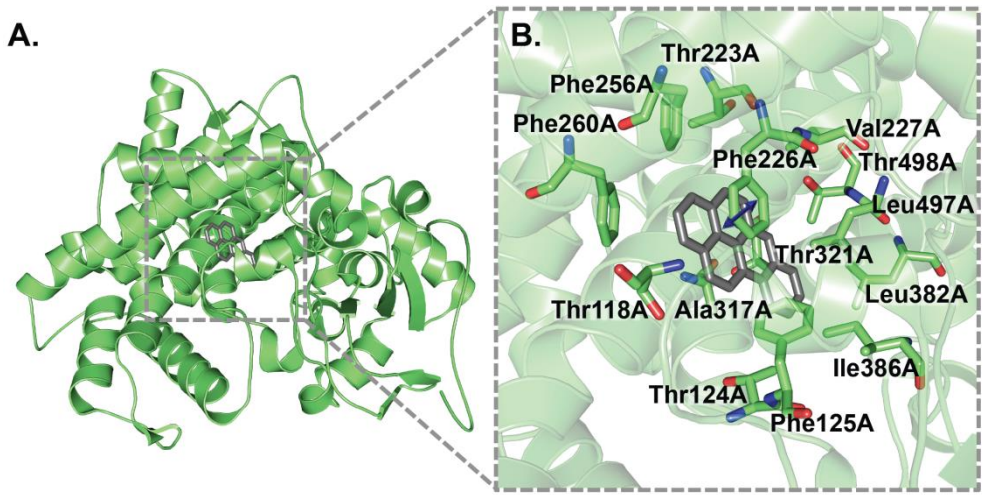


Figure 5.11. Three-dimensional view of the (A) CYP1A2/benzo(a)pyrene complex, (B) showing the binding site and interactions predicted by LigandScout 3.1. The blue arrows represent aromatic ring interactions.

Other complexes with EDCs from different sources also exhibited good affinity scores (Table 5.4). However, most of them were polycyclic aromatic hydrocarbons (PAHs).

Table 5.4. Miscellaneous EDCs with best affinity scores (<-10.0 kcal/mol) after docking with breast cancer associated proteins.

Short name	PDB	EDCs	CID	Keywords	Affinity (kcal/mol)
CYP1A2	2HI4	benzo(a)pyrene	2336	Combustion, wood, cigarette	-13.2
CYP1A2	2HI4	Benzanthrone	6697	Industrial pollutant, combustion	-13.0
CYP1A2	2HI4	6-hydroxychrysene	37766	Metabolite, combustion	-12.1
CYP1A2	2HI4	1-hydroxypyrene	21387	Metabolite, combustion	-12.0
MYLK4	2X4F	3-hydroxy-benzo(a)pyrene	25890	Combustion	-12.0
RARB	4DM6	benzo(a)pyrene	2336	Combustion, wood, cigarette	-12.0
MYLK4	2X4F	benzo(a)pyrene	2336	Combustion, wood, cigarette	-11.9
SHBG	1F5F	benzo(a)pyrene	2336	Combustion, wood, cigarette	-11.8
CYP1A2	2HI4	1,9-dimethylphenanthrene	34454	Environmental pollutant, cigarette	-11.8
CAT	1DGF	3-hydroxy-benzo(a)pyrene	25890	Combustion	-11.8
RARB	4DM6	Benzanthrone	6697	Industrial pollutant, combustion	-11.7
CAT	1DGF	benzo(a)pyrene	2336	Combustion, wood, cigarette	-11.6
CAT	1DGF	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-11.6
CHEK2	2W0J	benzo(a)pyrene	2336	Combustion, wood, cigarette	-11.5
CHEK2	2W0J	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-11.5
MYLK4	2X4F	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-11.5
SHBG	1F5F	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-11.4
CYP3A4	1TQN	benzo(a)pyrene	2336	Combustion, wood, cigarette	-11.4

Continuation of the Table 5.4.

Short name	PDB	EDCs	CID	Keywords	Affinity (kcal/mol)
CHEK2	2W0J	3-hydroxy-benzo(a)pyrene	25890	Combustion	-11.4
CYP3A4	1TQN	3-hydroxy-benzo(a)pyrene	25890	Combustion	-11.3
CYP3A4	1TQN	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-11.3
CYP2B6	3IBD	6-hydroxychrysene	37766	Metabolite, combustion	-11.3
SHBG	1F5F	3-hydroxy-benzo(a)pyrene	25890	Combustion	-11.3
PIK3CA	3HHM	benzo(a)pyrene	2336	Combustion, wood, cigarette	-11.2
ADIPOQ	4DOU	benzo(a)pyrene	2336	Combustion, wood, cigarette	-11.2
ADIPOQ	4DOU	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-11.2
PIK3CA	3HHM	3-hydroxy-benzo(a)pyrene	25890	Combustion	-11.2
ESR2	1QKM	benzo(a)pyrene	2336	Combustion, wood, cigarette	-11.1
CHEK2	2W0J	6-hydroxychrysene	37766	Metabolite, combustion	-11.1
ADIPOQ	4DOU	3-hydroxy-benzo(a)pyrene	25890	Combustion	-11.1
CYP1A2	2HI4	3-hydroxy-benzo(a)pyrene	25890	Combustion	-11.1
CDK2	1AQ1	3-hydroxy-benzo(a)pyrene	25890	Combustion	-11.0
SHBG	1F5F	6-hydroxychrysene	37766	Metabolite, combustion	-11.0
HPGDS	2VCQ	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-11.0
ADIPOQ	4DOU	6-hydroxychrysene	37766	Metabolite, combustion	-11.0
CDK2	1AQ1	benzo(a)pyrene	2336	Combustion, wood, cigarette	-11.0
CDK2	1AQ1	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.9
SRC	2H8H	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.9
RARG	2LBD	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.9
MYLK4	2X4F	6-hydroxychrysene	37766	Metabolite, combustion	-10.9

Continuation of the Table 5.4.

Short name	PDB	EDCs	CID	Keywords	Affinity (kcal/mol)
MYLK4	2X4F	Benzanthrone	6697	Industrial pollutant, combustion	-10.9
SHBG	1F5F	Benzanthrone	6697	Industrial pollutant, combustion	-10.8
NME1	1UCN	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.8
RXRG	2GL8	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.8
SRC	2H8H	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.8
SRC	2H8H	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.8
PIK3CA	3HHM	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.8
CYP2B6	3IBD	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.8
RARB	4DM6	1-hydroxypyrene	21387	Metabolite, combustion	-10.8
RXRB	1H9U	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.7
RXRB	1H9U	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.7
AR	2AM9	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.7
PPARD	2AWH	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.7
RXRA	2P1T	HHCB	91497	Fragrance	-10.7
MYLK4	2X4F	1-hydroxypyrene	21387	Metabolite, combustion	-10.7
STRADA/MO25	3GNI	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.7
RARB	4DM6	1,9-dimethylphenanthrene	34454	Environmental pollutant, cigarette	-10.7
HPGDS	2VCQ	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.6

Continuation of the Table 5.4.

Short name	PDB	EDCs	CID	Keywords	Affinity (kcal/mol)
HPGDS	2VCQ	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.6
AR	2AM9	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.6
APOD	2HZQ	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.6
APOD	2HZQ	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.6
RXRA	2P1T	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.6
CAMK2B	3BHH	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.6
RARB	4DM6	6-hydroxychrysene	37766	Metabolite, combustion	-10.6
GSTP1	3N9J	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.6
CAT	1DGF	6-hydroxychrysene	37766	Metabolite, combustion	-10.6
AR	2AM9	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.6
CAT	1DGF	Benzanthrone	6697	Industrial pollutant, combustion	-10.5
RARB	4DM6	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.5
NME1	1UCN	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.5
PPARD	2AWH	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.5
APOD	2HZQ	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.5
STRADA/MO25	3GNI	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.5
MAPK1	3I60	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.5
RPS6KB1	4L3J	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.5
RPS6KB1	4L3J	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.5
ESR2	1QKM	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.5

Continuation of the Table 5.4.

Short name	PDB	EDCs	CID	Keywords	Affinity (kcal/mol)
RPS6KB1	4L3J	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.5
CYP1A2	2HI4	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.4
NME1	1UCN	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.4
PPARD	2AWH	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.4
RXRG	2GL8	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.4
RPS6KA1	2WNT	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.4
PIK3CG	3MJW	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.4
GSTP1	3N9J	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.4
MET	1R0P	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.4
RELA	2O61	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.4
CAMK2B	3BHH	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.4
CAMK2B	3BHH	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.4
NQO1	1D4A	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.3
RORA	1N83	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.3
PGR	1SQN	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.3
RXRA	2P1T	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.3
RORC	3L0L	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.3
PIK3CG	3MJW	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.3
CXCR4	3ODU	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.3
SHBG	1F5F	1-hydroxypyrene	21387	Metabolite, combustion	-10.2
MET	1R0P	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.2
MET	1R0P	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.2

Continuation of the Table 5.4.

Short name	PDB	EDCs	CID	Keywords	Affinity (kcal/mol)
PGR	1SQN	benzo(a)pyrene	2336	Combustion, Wood, Cigarette	-10.2
PGR	1SQN	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.2
NR3C2	2AA2	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.2
AR	2AM9	6-hydroxychrysene	37766	Metabolite, Combustion	-10.2
PPARD	2AWH	6-hydroxychrysene	37766	Metabolite, Combustion	-10.2
MSH2	2O8B	benzo(a)pyrene	2336	Combustion, Wood, Cigarette	-10.2
RPS6KA1	2WNT	2-hydroxybenzo(a)pyrene	42027	Metabolite, Combustion	-10.2
FGFR2	3B2T	2-hydroxybenzo(a)pyrene	42027	Metabolite, Combustion	-10.2
ARF-BP1	3H1D	benzo(a)pyrene	2336	Combustion, Wood, Cigarette	-10.2
MAPK1	3I60	benzo(a)pyrene	2336	Combustion, Wood, Cigarette	-10.2
MAPK1	3I60	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.2
GSTP1	3N9J	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.2
CHKA	2CKO	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.2
RARG	2LBD	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.2
HPGDS	2VCQ	Aurin	5100	Analytical Chemistry	-10.2
RPS6KA1	2WNT	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.2
PIK3CA	3HHM	6-hydroxychrysene	37766	Metabolite, Combustion	-10.2
MTOR	4JSN	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.2
PIK3CA	3HHM	Benzanthrone	6697	Industrial Pollutant, Combustion	-10.1
CDK2	1AQ1	6-hydroxychrysene	37766	Metabolite, Combustion	-10.1
CAT	1DGF	1-hydroxypyrene	21387	Metabolite, Combustion	-10.1

Continuation of the Table 5.4.

Short name	PDB	EDCs	CID	Keywords	Affinity (kcal/mol)
SHBG	1F5F	1,9-dimethylphenanthrene	34454	Environmental pollutant, cigarette	-10.1
RORA	1N83	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.1
NR3C2	2AA2	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.1
NR3C2	2AA2	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.1
ESRRG	2E2R	3-monobromobisphenol A	656688	Metabolite, flame retardants	-10.1
MYLK4	2X4F	1,9-dimethylphenanthrene	34454	Environmental pollutant, cigarette	-10.1
ESR1	3ERT	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.1
RORC	3L0L	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.1
PPARG	3LMP	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.1
GSTP1	3N9J	Benzanthrone	6697	Industrial pollutant, combustion	-10.1
ADIPOQ	4DOU	Benzanthrone	6697	Industrial pollutant, combustion	-10.1
MTOR	4JSN	benzo(a)pyrene	2336	Combustion, wood, cigarette	-10.1
ESR2	1QKM	3-hydroxy-benzo(a)pyrene	25890	Combustion	-10.1
ADIPOQ	4DOU	bis(4-hydroxyphenyl)[(2-phenoxysulfonyl)phenyl]methane	130780	Analytical chemistry	-10.1
CXCR4	3ODU	2-hydroxybenzo(a)pyrene	42027	Metabolite, combustion	-10.1

5.2.7. BPA/SHBG interaction

The interaction of BPA or its derivatives and breast cancer associated proteins is of special concern, as humans are exposed to these chemicals through contact with different products that contain them, especially through foods that have been in contact with packaging materials that may release these pollutants[33]. Several authors have proposed these chemicals affect the function of multiple organs and increase the risk of breast cancer in mice[34]. In breast cancer, many proteins play a pivotal role in the pathogenesis of this disease. However, SHBG is a small and stable protein with high affinity for estrogens and androgens that has been found in breast tissue and cell lines through immunostaining[35], being the major and specific binding protein for testosterone and estradiol. Besides, decreased circulating levels of this protein have been observed in breast cancer patients possibly indicating higher bioavailable estrogens[36]. These were the main reasons we use to choose it to obtain it and to validate our docking predictions. Therefore, we selected the hormone transporter SHBG[37], as this exhibited a good affinity score in AutoDock Vina (-8.2 kcal/mol), to computationally assess their interactions and conformational changes by CD. Moreover, the data reported in the scientific literature suggests that human SHBG may transport some xenoestrogens into the plasma and modulate their bioavailability to cell tissues[38], which could disrupt the natural hormonal balance.

The conformational analysis performed in LigandScout 3.1[21] suggests that BPA interacts in the binding pocket of SHBG (Figure 5.12). Most of the interactions of BPA were hydrophobic (Met139A, Leu80A, Val112A), although one residue in the binding site exhibited hydrogen bond donor and acceptor features (Thr60A), and an aromatic ring interaction was predicted for BPA and one of the amino acids of the contact residues (Phe67A).

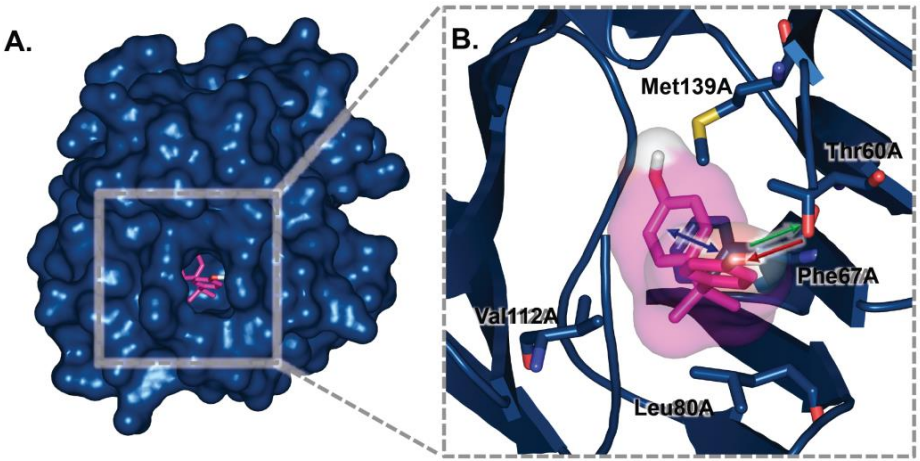


Figure 5.12. Three-dimensional view of the (A) SHBG/ BPA complex, (B) showing the binding site and interactions predicted by LigandScout 3.1. The green arrows represent hydrogen-bond donor features, the red arrows show the hydrogen-bond acceptor features, and the blue arrows indicate aromatic ring interactions.

The three-dimensional view of the complex also indicates the position of BPA deep in the binding site, which can contribute to an elevated affinity. It should be pointed out that conformational changes should have occurred to allow this docking pose.

5.2.8. Protein expression and purification

Recombinant SHBG was expressed and purified from *E. coli* C41(DE3) cells using Dynabeads His-Tag Isolation and Pulldown.

5.2.8.1. Recombinant plasmid design. The blank plasmid pET15-MHL (Figure 5.13) was selected as empty backbone for the insertion of the sequence that codify for the human protein SHBG in *E. coli*, as it has a bacterial resistance gene for ampicillin, HIS-tag (N-terminal on backbone) and a Lac operon for the induction of the expression by IPTG; this also have restriction sites compatible with the sequence of the protein gene. This plasmid was purchased to Addgene (plasmid # 26093).

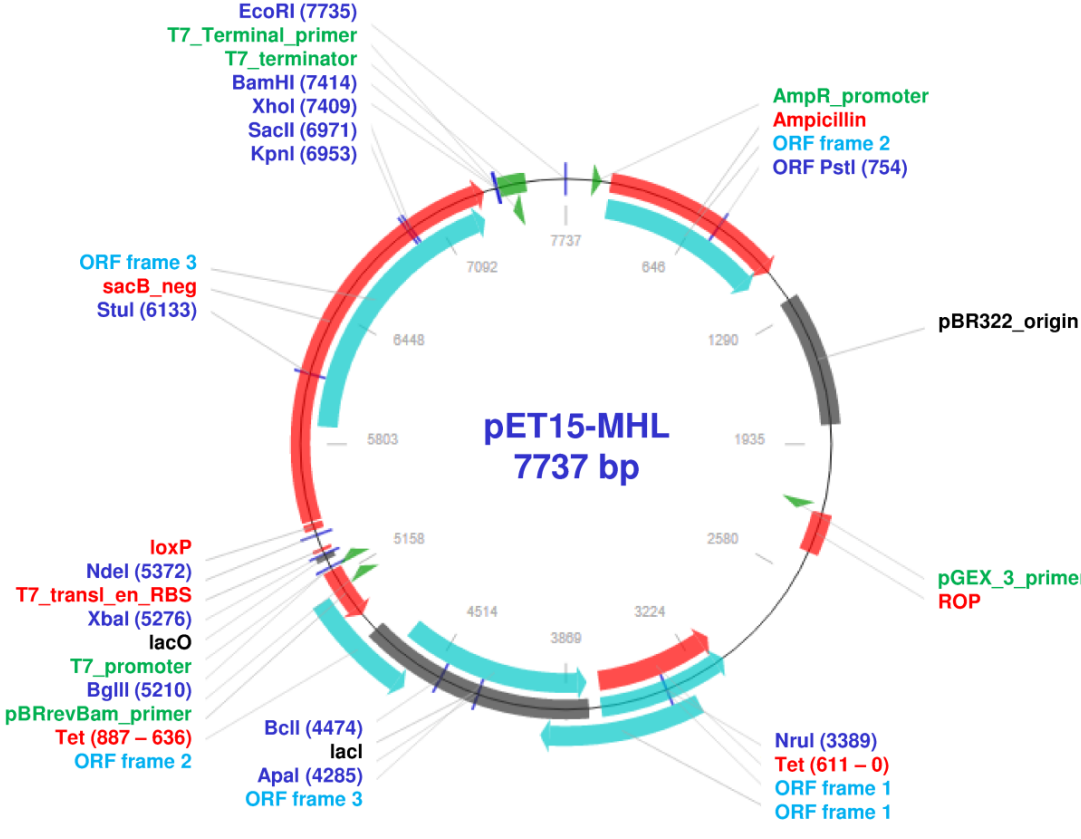


Figure 5.13. pET15-MHL plasmid map.

The recombinant plasmid, containing the SHBG sequence was synthetized by Eurofins, using the blank plasmid pET15-MHL. Based on the recombinant plasmid design (Annex 5.6), the restriction sites selected for the SHBG gene insertion were 5' NdeI and 3' XhoI, as their restriction sequences are not included in the SHBG gene. The gene sequence, as well as the *E. coli* codon usage is presented below.

- ***Synthesized gene sequence of SHBG***

catATG CTG CGT CCA GTC TTA CCG ACG CAA TCA GCG CAT GAT CCG CCA GCA GTG CAC CTT AGC AAT GGC CCT GGT CAG GAA CCC ATT GCG GTG ATG ACC TTC GAT CTG ACC AAG ATT ACG AAA ACC AGT AGC TCG TTT GAG GTT CGC ACA TGG GAT CCG GAA GGC GTC ATC TTC TAT GGG GAC ACT AAC CCG AAA GAC GAC TGG TTC ATG CTC GGG TTG CGT GAT GGT CGT CCG GAA ATT CAG CTT CAC AAC CAT TGG GCC CAA CTG ACC GTA GGC GCT GGT CCT CGC TTG GAC GAT GGA CGT TGG CAT CAG GTT GAG GTG AAA ATG GAA GGC GAT TCC GTG CTG CTG GAA GTG GAT GGA GAA GAA GTC CTG CGT CTG CGG CAA GTT TCG GGT CCC TTG ACG AGC AAA CGC CAC CCG ATT ATG CGG ATT GCA TTA GGC GGT CTG CTG TTT CCG GCC AGC AAT CTC CGC CTT CCG CTG GTA CCG GCA CTG GAT GGC TGT CTG CGC CGC GAT TCC TGG CTC GAC AAA CAG GCC GAG ATC TCT GCT TCT GCG CCA ACC TCG TTA CGC AGT TGC GAC GTT GAG TCA AAC CCT GGC ATC TTT CTG CCA CCG GGT ACT CAG GCG GAA TAActcgag

- ***Codon Usage***

Codon usage AAA 0.77 n/a 0.83 AAC 0.55 n/a 0.60 AAG 0.23 n/a 0.17 AAT 0.45 n/a 0.40 ACA 0.13 n/a 0.09 ACC 0.44 n/a 0.45 ACG 0.27 n/a 0.27 ACT 0.17 n/a 0.18 AGC 0.28 n/a 0.27 AGT 0.15 n/a 0.13 ATC 0.42 n/a 0.38 ATG 1.00 n/a 1.00 ATT 0.58 n/a 0.62 CAA 0.35 n/a 0.38 CAC 0.43 n/a 0.50 CAG 0.65 n/a 0.62 CAT 0.57 n/a 0.50 CCA 0.19 n/a 0.21 CCC 0.12 n/a 0.11 CCG 0.53 n/a 0.53 CCT 0.16 n/a 0.16 CGC 0.52 n/a 0.50 CGG 0.10 n/a 0.14 CGT 0.38 n/a 0.36 CTC 0.10 n/a 0.12 CTG 0.53 n/a 0.52 CTT 0.10 n/a 0.12 GAA 0.69 n/a 0.67 GAC 0.37 n/a 0.40 GAG 0.31 n/a 0.33 GAT 0.63 n/a 0.60 GCA 0.21 n/a 0.25 GCC 0.27 n/a 0.25 GCG 0.36 n/a 0.33 GCT 0.16 n/a 0.17 GGA 0.11 n/a 0.12 GGC 0.40 n/a 0.41 GGG 0.15 n/a 0.12 GGT 0.34 n/a 0.35 GTA 0.15 n/a 0.14 GTC 0.22 n/a 0.21 GTG 0.37 n/a 0.36 GTT 0.26 n/a 0.29 TAA 0.71 n/a 0.00 TAC 0.43 n/a 0.00 TAT 0.57 n/a 1.00 TCA 0.12 n/a 0.13 TCC 0.15 n/a 0.13 TCG 0.15 n/a 0.20 TCT 0.15 n/a 0.13 TGA 0.29 n/a 0.00 TGC 0.56 n/a 0.50 TGG 1.00 n/a 1.00 TGT 0.44 n/a 0.50 TTA 0.13 n/a 0.12 TTC 0.43 n/a 0.50 TTG 0.13 n/a 0.12 TTT 0.57 n/a 0.50

The codon usage in *E. coli* for the SHBG gene sequence was in general low. Therefore the expression of the protein was expected to be moderate.

5.2.8.2. SHBG protein expression. The protein SHBG was expressed using the C41 (DE3) strain of *E. coli*. In order to produce more plasmid, *E.coli* DH5α was transfected with the recombinant plasmid pET15-MHL/SHBG (Figure 5.14). In this procedure, recombinant bacteria colonies containing the plasmid were generated, for its growth in LB media for the posterior pET15-MHL/SHBG plasmid isolation.

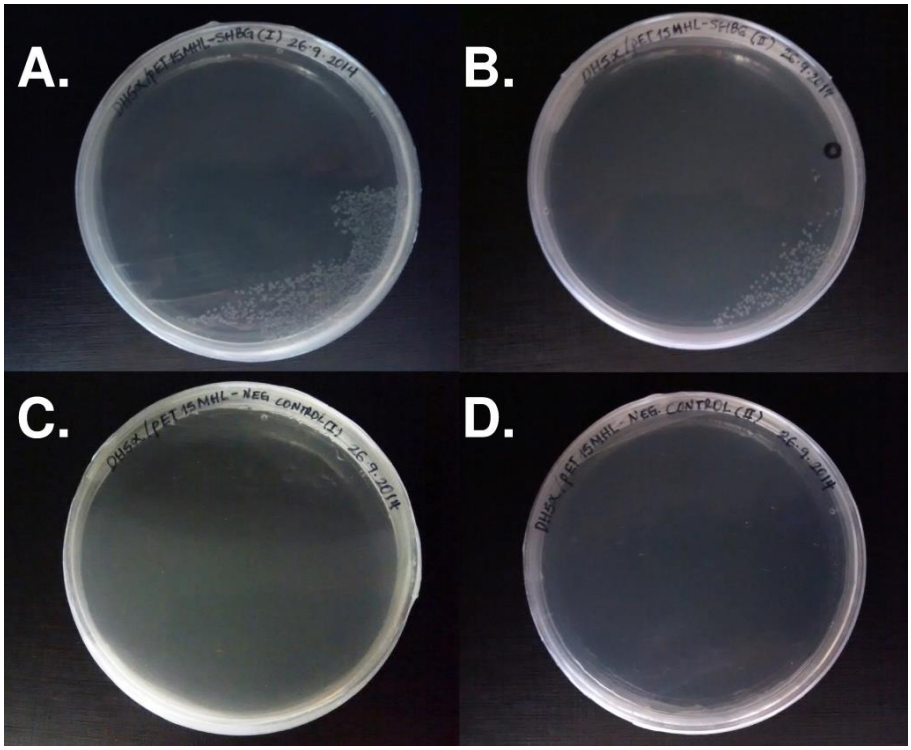


Figure 5.14. Agar plates with (A-B) recombinant E. coli DH5a/pET15MHL-SHBG and (C-D) negative control (E. coli DH5a without the recombinant plasmid).

The selected *E. coli* strain for the expression, C41(DE3), was then transfected with the plasmid (Figure 5.15). For verification of the successful transfection, they were growth in Agar plates containing the antibiotic ampicillin, therefore only the bacteria containing the plasmid survived. A negative control was also carried out in both cases, and no bacteria growth was observed on it.

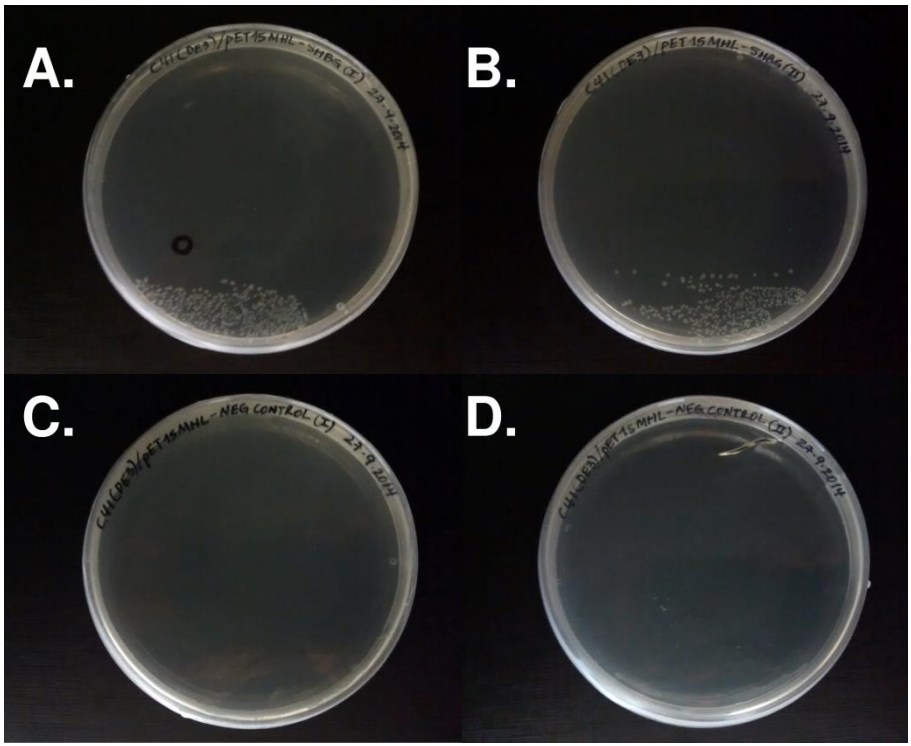


Figure 5.15. Agar plates with (A-B) recombinant E. coli C41(DE3)/pET15-MHL-SHBG and (C-D) negative control (E. coli C41(DE3) without the recombinant plasmid).

One colony of recombinant *E. coli* C41(DE3)/pET15MHL-SHBG was collected and grown in LB medium, used as a starter solution to inoculate the medium for protein expression, the required optical density 0.6±0.2 for induction was

reached 100 minutes after the bacteria inoculation. The optical density of the solution at different times is presented in Table 5.5.

Table 5.5. Optical density of the LB media after inoculation with recombinant E. coli C41(DE3).

Time since inoculation (min)	Absorbance (600 nm)	
	Control	Sample
60	(0.0321)	0.3219
90	(0.0322)	0.5166
100	(0.0324)	0.5782

Samples were collected each hour after induction with IPTG, and an electrophoresis gel was run, samples were then centrifuged to improve the concentration, however a clear distinction of the protein band was not perceived (Figure 5.16).

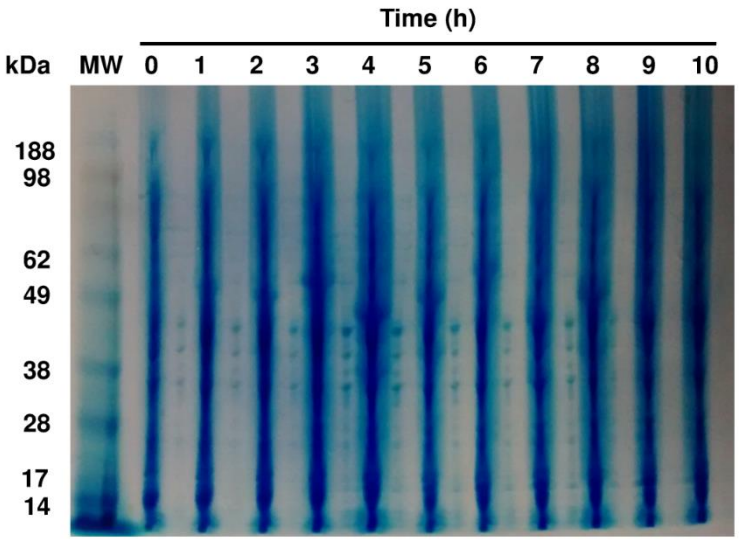


Figure 5.16. SDS-Page electrophoresis gel of the centrifuged bacteria culture in LB media at different hours, after induction with IPTG (MW: molecular weight, SeeBlue Plus2 prestained protein standard).

Therefore a western blot was required. This showed an increase in the protein quantity with the time (Figure 5.17). It was observed at a higher molecular weight than expected. However, it was similar to the reported for human SHBG western blot by Hong et al.[39].

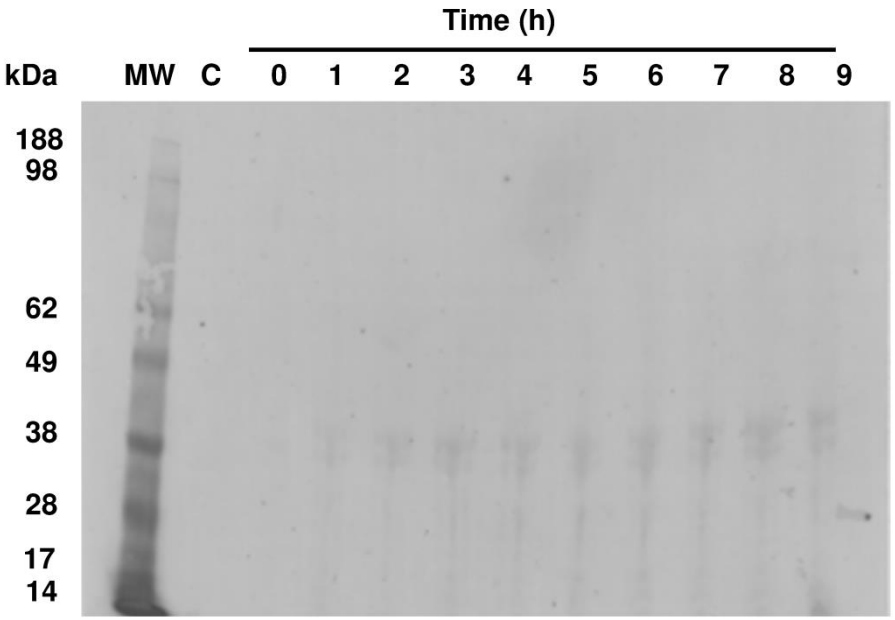


Figure 5.17. Western-blot of the centrifuged bacteria culture in LB media at different hours, after induction with IPTG (MW: molecular weight, SeeBlue Plus2 prestained protein standard; C: Control).

In order to set up the conditions of time and IPTG concentration, a new experiment was ran by collecting the samples at 12, 14 and 16h after induction with 0.1, 0.5 and 1mM IPTG, and an electrophoresis gel was carried out, using the pellet of the centrifuged samples. The suitable conditions for the protein expression were established as 14 hours and 0.5 mM IPTG, as there were no visual differences between the bands obtained for this concentration and 1mM IPTG in the same time.

5.2.8.3. Protein Purification. The protein expression was then carried out under these conditions. The pellet was submitted to a lysis (Lysis procedure= http://www.abcam.com/index.html?pageconfig=resourceandrid=11379), and concentration procedure before the protein purification with Dynabeads® His-Tag Isolation and Pulldown kit. A dialysis was performed after the protein purification to eliminate possible impurities bellow 20 kDa. After this procedure, the protein looked very clean (Figure 5.18) and the band was placed in the expected molecular weight (22.610 kDa).

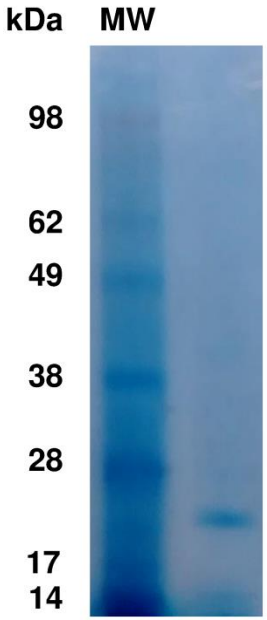


Figure 5.18. SDS-Page electrophoresis gel of the purified SHBG after dialysis (MW: molecular weight, SeeBlue Plus2 prestained protein standard).

5.2.9. Microscale thermophoresis

After FPLC purification, a clear separation between the protein and dye was observed in the spectra (Figure 5.19). However a small portion of the total protein was labeled and the fractions without free dye collected (75-77).

Therefore, this study was used with qualitative purposes, which was enough to evaluate the functionality of the recombinant protein to bind the ligand.

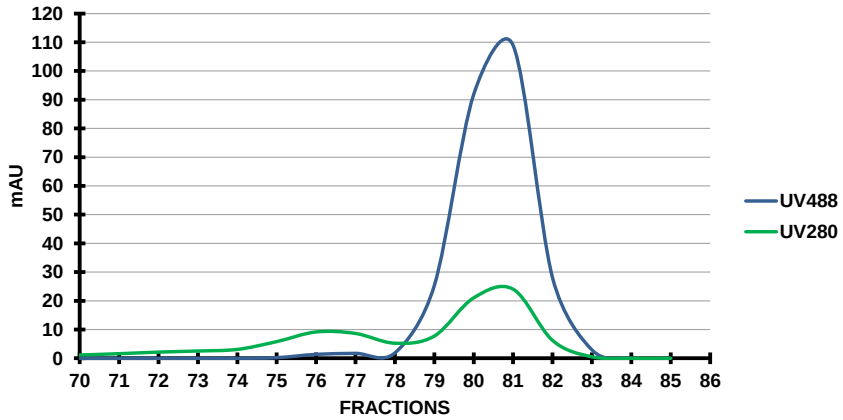


Figure 5.19. FPLC espectra of the labeled protein and dye separation.

The capillary scan showed that the hydrophilic one was suitable for the study due to it generated a Gaussian curve (Figure 5.20).

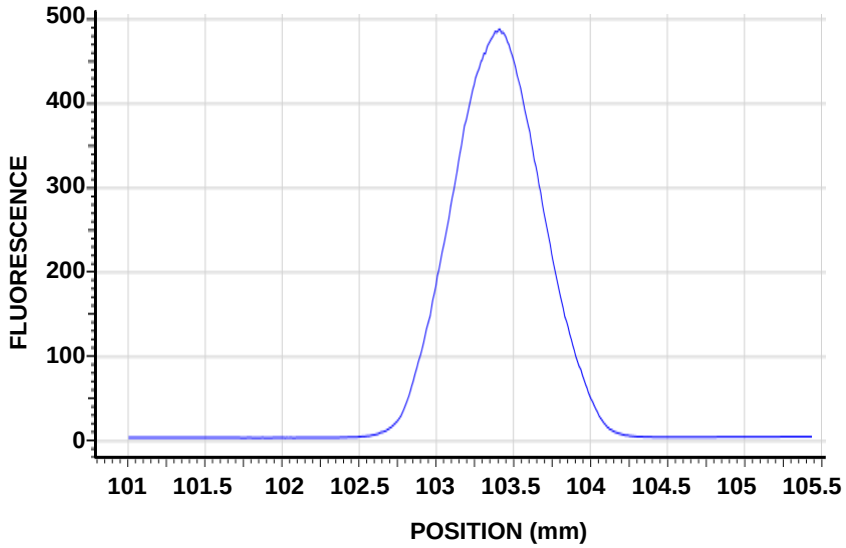


Figure 5.20. Cap scan of labeled SHBG with the hydrophilic capillary.

The curve of the MST analysis and the fluorescence data is presented in Figure 5.21. This shows that SHBG has the tendency to bind BPA at a concentration of approximately 1μM of BPA, although a saturation point was not reached. This could be the result of a low protein labelling with a high non-labeled protein ratio, which makes it more difficult to reach the saturation point as the total protein concentration was high in the sample.

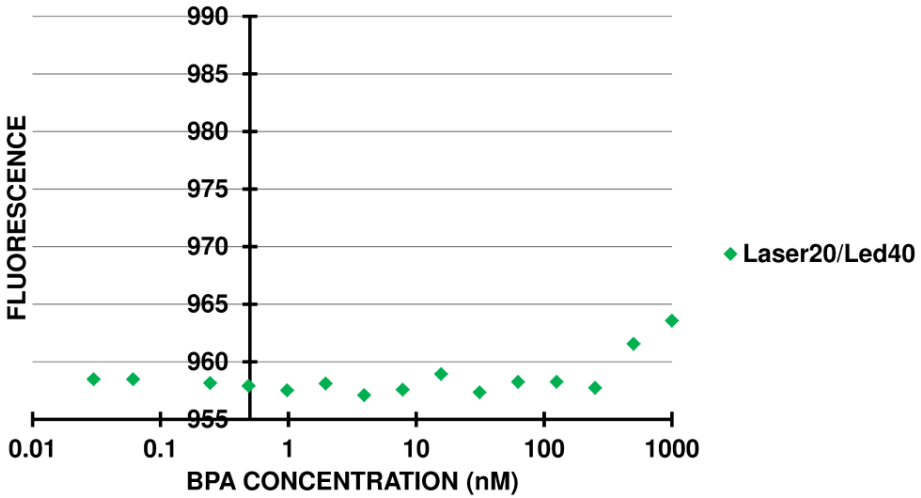


Figure 5.21. MST study of the protein-ligand interaction of fluorescently labeled SHBG and BPA.

5.2.10. Circular dichroism

The correct folding of the recombinant protein was assessed by CD. According to PDB[16], the secondary structure of SHBG (PDB ID: 1F5F) consist of 4% alpha helix and 40% beta sheets. This correlates well with the values obtained by circular dichroism of the SHBG dissolved in PBS at pH 7.4: helix=6.3% and beta sheets=40.6% (strand 1=26.3% and strand 2=14.3%). The results of the secondary structure analysis, according to the DichroWeb[30] software are presented in Figure 5.22.

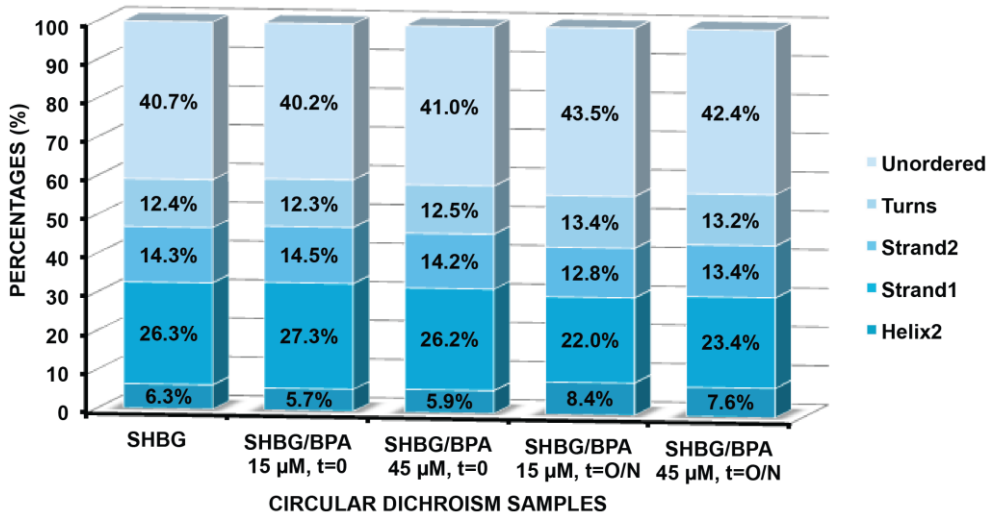


Figure 5.22. Changes in the secondary structure of SHBG after incubation with BPA at time zero (t=0) and overnight (t=O/N).

At time zero (t=0), no main differences were observed for the contents of helix or strands of the samples of SHBG with BPA 1.5 x 10^{-5} and 4.5 x 10^{-5} M. However, after overnight incubation (t=O/N), an increase in the helix content of the protein was observed, with a percentage increase of more than 20% in both cases. The CD spectra for 1.5 x 10^{-5} and 4.5 x 10^{-5} M of BPA at time zero and overnight are presented in Figure 5.23. The Dunn's test did not show statistically significant differences between the medians of the SHBG sample and the SHBG/BPA samples in time zero. However, a statistical difference was revealed for the comparison of each sample after overnight incubation with the protein sample, and the SHBG/BPA samples in time zero (Table 5.6).

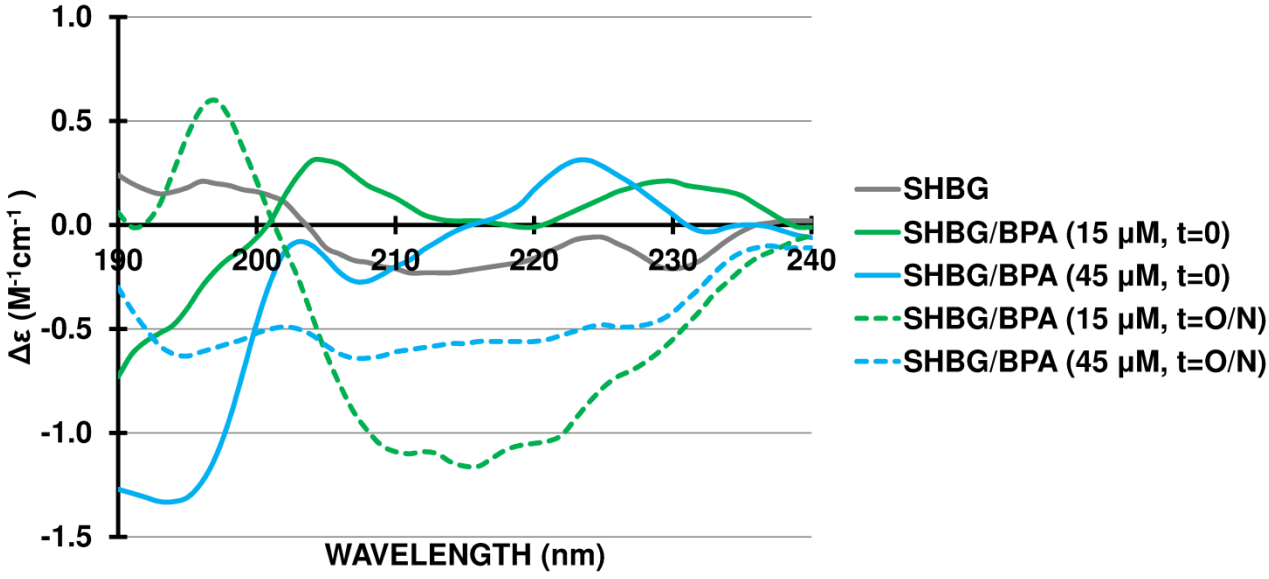


Figure 5.23. Processed circular dichroism spectra of SHBG and SHBG incubated with BPA, 1.5 x 10^{-5} and 4.5 x 10^{-5} M, at time zero (t=0) and overnight (t=O/N) by DichroWeb[30].

Table 5.6. Results of the statistical analysis.

Dunn's multiple comparisons test	Mean rank difference,	Significant
SHBG vs. SHBG/BPA (15µM, t=0)	-26,71	No
SHBG vs. SHBG/BPA (45µM, t=0)	7,735	No
SHBG vs. SHBG/BPA (15µM, t=O/N)	101,5	Yes
SHBG vs. SHBG/BPA (45µM, t=O/N)	78,19	Yes
SHBG/BPA (15µM, t=0) vs. SHBG/BPA (45µM, t=0)	34,44	No
SHBG/BPA (15µM, t=0) vs. SHBG/BPA (15µM, t=O/N)	128,2	Yes
SHBG/BPA (15µM, t=0) vs. SHBG/BPA (45µM, t=O/N)	104,9	Yes
SHBG/BPA (45µM, t=0) vs. SHBG/BPA (15µM, t=O/N)	93,78	Yes
SHBG/BPA (45µM, t=0) vs. SHBG/BPA (45µM, t=O/N)	70,45	Yes
SHBG/BPA (15µM, t=O/N) vs. SHBG/BPA (45µM, t=O/N)	-23,33	No

5.3. DISCUSSION

In this work, we studied the potential of EDCs from urban sources to target breast cancer proteins using an *in silico* approach. The macromolecules CAT, RARB, SHBG, PGR and PIK3CA presented the highest number of theoretical complexes with xenoestrogens (Affinity ≤ -8.0 kcal/mol). These are pivotal in breast cancer and could be modulating the response to EDCs exposure through metabolic ways. CAT is an antioxidant enzyme that helps to control oxidative stress and DNA damage associated to EDCs exposure and breast cancer development[40]; RARB modulates the proliferation of breast cancer cells by limiting the growth and promoting the apoptosis[41]; SHBG transports steroid hormones through the bloodstream and limits their free fraction, therefore its binding with xenoestrogens could increase the bioavailable estradiol and augment the risk of breast cancer[42]; PGR, a well-known marker that with the ER status is used in the immunohistochemically prognosis of this disease, except for the triple negative subtype, that comprises around 15% of the cases[43]; and PIK3CA has an important role in neoplasia, and aberrations in its pathway or in the gene that codifies it results in an increased risk of cancer[44].

Several of the tested small molecules exhibited high docking affinity in computer simulations and should be prioritized for *in vitro* and *in vivo* assays. In addition some of them showed to be frequent hitters as they act as promiscuous compounds binding different targets[45]. The visual inspection of the docking affinity heat maps (see Annexes 5.2-5.5) indicates that the highest density of protein-ligand pairs with potential to interact between them is presented for dioxins and related molecules. Therefore, this may be the principal source of EDCs interacting with the metabolic pathways involved in breast cancer. The number of potential targets was followed for the categories: miscellaneous, everyday products, and plastics and other types of polymers, respectively.

The compounds with the highest number of theoretical targets among the breast cancer proteins were the PAHs: 2-hydroxybenzo(a)pyrene, benzo(a)pyrene and 3-hydroxy-benzo(a)pyrene. In the group of plastics and other type of polymers, the most frequent hitter was bisphenol M, followed by the synthetic compound used in the production of thermally stable polyesters and polycarbonates, 4,4'-(octahydro-4,7-methano-5H-inden-5-ylidene) bisphenol, and the styrene trimer, 2,4,6-triphenyl-1-hexene, generated in the decomposition of plastics and released from containers to food[46]. For everyday products, the most promiscuous compound found was the PFOS. This is interesting, as EDCs can be interacting with many different pathways through protein-ligand interactions, making their mechanisms and effects more difficult to understand and control.

The two complexes with the highest affinity scores in the group of dioxins and related molecules were: CAT/1,3,7,8-tetrachlorodibenzo-p-dioxin and CYP1A2/2,3,4,7-tetrachlorodibenzofuran. However, other chemicals in this group also presented very good affinity (≤-9.5 kcal/mol) for CYP1A2 (See Annex 5.1), which is in agreement with the scientific literature indicating these types of compounds have been found to bind and induce CYP1A2[47]. The eventual binding of 1,3,7,8-tetrachlorodibenzo-p-dioxin to CAT may have an effect in the hydrogen peroxide binding an then in the oxidative stress response.

EDCs from plastics and other types of polymers constitute a special category due to the increased use of these materials in human life and the annual tendency to progressively augment their already high volumes of production[48]. One of the plasticizers that have received more attention in the last decade is the BPA, a monomer used in polycarbonate plastics and epoxy resins frequently found in food containers, which as a result of restrictive regulations has started to be replaced by other analogues[49]. Surprisingly, some of these derivatives (bisphenol M, bisphenol B, BPAF, dihydroxymethoxychlor olefin and bisphenol A dimethacrylate) are on the top of the complexes with highest docking affinity for breast cancer proteins *in silico*, with greater affinity than BPA; therefore we suggest them as priority compounds to be tested against this disease. The two best affinity scores in

this group were obtained for bisphenol M in complex with SRC, a protein that may promote the growth of tumor cells and is overexpressed in breast cancer[50]; and for BPAF with ESRRG, an orphan nuclear receptor widely implicated in the transcriptional regulation of energy homeostasis[51] that acts as tumor suppressor in several types of cancers[52]. Interestingly, BPAF has been found to promote breast cancer cell proliferation *in vitro*[53], and bisphenol M exhibited high predicted affinity (<-10.0 kcal/mol) for several proteins related to breast cancer.

A small number of compounds released from household products, cosmetics, drugs and personal care products were found to strongly bind breast cancer proteins *in silico* (<-10.0 kcal/mol). Some of them are: perfluorooctane sulfonic acid used in the manufacture of plastics, textiles, electronics, and many other industrial products[54]; AHTN found in cosmetics[55]; emodin, a medicine that has been found to suppress tumor growth[56]; 3-(4-methylbenzylidene)camphor employed as UV filter in cosmetics[57]; and triclocarban, frequently utilized in detergents and soaps as antimicrobial agent[58]. Nevertheless, there were many molecules in this group that presented moderate and weak binding that eventually could have a role in the homeostasis disruption in relation to the development or progression of this disease. On the other hand, several of the targeted proteins for these compounds are shared with the top list for plastics and related molecules, as well as with the proteins most frequently found in theoretical EDCs-breast cancer protein complexes, some of them are SHBG, RARB, CAT and vitamin D3 receptor (VDR).

In contrast, a large number of PAHs, derived from combustion, occupied the top ranking of complexes formed by miscellaneous EDCs. Some of these are the well-known carcinogen benzo(a)pyrene, benzanthrone, 6-hydroxychrysene and 1-hydroxypyrene. These types of molecules have been classically associated to breast cancer through genetic damage by DNA adduct formation[59], however they could also be acting throw protein-ligand interactions. In addition, other molecules from diverse sources also exhibited high affinity for breast cancer proteins such as: HHCB, used in fragrances[55]; the chemical indicator aurin[60]; and the flame retardant 3-monobromobisphenol A[61]; among others.

The computer simulations of the docking poses and protein-ligand interaction analysis of the top complexes in each category showed that the most frequent binding forces were due to hydrophobic, hydrogen bond donor, hydrogen bond acceptor and aromatic ring features of the ligands and contact residues in the binding site. The conformational analysis of the protein-ligand pair selected for *in vitro* studies, SHBG/BPA, showed that the predicted interacting residues (Met139A, Thr60A, Phe67A, Leu80A and Val112A), are found to be the same experimentally on the binding site of endogenous hormones and xenoestrogens, that binds this protein with high affinity[62]. Therefore, BPA may occupy the pocket for natural estrogens in the body, reducing the fraction of bound estrogen and increasing its bioavailability[63], which is considered a risk factor to develop this disease[35].

The recombinant human SHBG used in our spectroscopic analysis was obtained without any mutation, and according to the MST analysis, it was functional, and their folding was correct, as assessed by the circular dichroism spectra of the native protein without the ligand. The differences in the circular dichroism spectra of SHBG and SHBG with BPA may be the result of both BPA binding and the dimerization of the protein, as has been reported with steroid ligands that bind this protein with elevated affinity[64]. The main conformational change occurred after BPA binding was an increment of the alpha helix content after overnight incubation. This result has also been observed for the binding of this molecule to other proteins, such as human serum albumin[29]. The Dunn's test showed statistical differences in the spectra for overnight incubation with BPA, but not for the immediate CD recorded after BPA addition (t=0), suggesting that the reaction between BPA and SHBG, may be time-dependent. This is not surprising, as related studies regarding binding assays, used the same physiologic conditions of pH and temperature as in our protocol, as well as an incubation time of at least one hour[63a].

The results discussed in this section showed the plausibility of some EDCs to interact with proteins involved in signaling pathways regarding breast

cancer. Therefore further analysis in this field is needed to support regulatory actions, being of special concern the case of the new BPA analogs.

5.4. CONCLUSIONS

This study presents a virtual screening that helps to understand how endocrine disruptors could be plausible ligands for the breast cancer proteome. This approach provided important candidates to be evaluated *in vitro* against models of this disease, predicting protein patterns of proteins that are more commonly affected by EDCs. Many of the predicted hits have no scientific reports regarding its experimental evaluation; therefore further studies are required in this field. Computational studies are an important tool that can improve the velocity and efficiency of the evaluations of EDCs overcoming the known limitations regarding the high number of this compounds and diseases associated to them. Several endocrine disruptors have the theoretical capability to bind proteins related to breast cancer, some of them are more promiscuous being able to bind different targets. Therefore these are proposed as priority for experimental test. Moreover, BPA binds SHBG, showing a conformational change of the protein.

REFERENCES

1. Brody, J. T., Breast cancer risk in relation to occupations with exposure to carcinogens and endocrine disruptors: a Canadian case-control study. National Network on Environments and Women's Health: 2012.
2. (a) Hauser, R.; Skakkebaek, N. E.; Hass, U.; Toppari, J.; Juul, A.; Andersson, A. M.; Kortenkamp, A.; Heindel, J. J.; Trasande, L., Male Reproductive Disorders, Diseases, and Costs of Exposure to Endocrine-Disrupting Chemicals in the European Union. J. Clin. Endocrinol. Metab. **2015,** 100 (4), 1267-1277; (b) Bellanger, M.; Demeneix, B.; Grandjean, P.; Zoeller, R. T.; Trasande, L., Neurobehavioral Deficits, Diseases, and Associated Costs of Exposure to Endocrine-Disrupting Chemicals in the European Union. J. Clin. Endocrinol. Metab. **2015,** 100 (4), 1256-1266; (c) Bedia, C.; Dalmau, N.; Jaumot, J.; Tauler, R., Phenotypic malignant changes and untargeted lipidomic analysis of long-term exposed prostate cancer cells to endocrine disruptors. Environ. Res. **2015,** 140 (0), 18-31;

(d) Prins, G. S., Endocrine disruptors and prostate cancer risk. Endocr. Relat. Cancer **2008,** 15 (3), 649-656.

3. Stojić, N.; Erić, S.; Kuzmanovski, I., Prediction of toxicity and data exploratory analysis of estrogen-active endocrine disruptors using counter-propagation artificial neural networks. J. Mol. Graph. Model. **2010,** 29 (3), 450-460.
4. Erickson, B. E., EPA RETOOLS ENDOCRINE PROGRAM. Chem. Eng. News **2013,** 91 (11), 30-32.
5. Rotroff, D. M.; Martin, M. T.; Dix, D. J.; Filer, D. L.; Houck, K. A.; Knudsen, T. B.; Sipes, N. S.; Reif, D. M.; Xia, M.; Huang, R.; Judson, R. S., Predictive Endocrine Testing in the 21st Century Using in Vitro Assays of Estrogen Receptor Signaling Responses. Environ. Sci. Technol. **2014,** 48 (15), 8706-8716.
6. Wang, N. C. Y.; Jay Zhao, Q.; Wesselkamper, S. C.; Lambert, J. C.; Petersen, D.; Hess-Wilson, J. K., Application of computational toxicological approaches in human health risk assessment. I. A tiered surrogate approach. Regul. Toxicol. Pharmacol. **2012,** 63 (1), 10-19.
7. Schneider, G.; Böhm, H.-J., Virtual screening and fast automated docking methods. Drug Discov. Today **2002,** 7 (1), 64-70.
8. Kemsley, J., Rethinking breast cancer toxicology. Chem. Eng. News **2010,** 88 (10), 40-41.
9. Lauro, G.; Masullo, M.; Piacente, S.; Riccio, R.; Bifulco, G., Inverse Virtual Screening allows the discovery of the biological activity of natural compounds. Bioorg. Med. Chem. **2012,** 20 (11), 3596-3602.
10. Hondermarck, H.; Tastet, C.; El Yazidi-Belkoura, I.; Toillon, R.-A.; Le Bourhis, X., Proteomics of Breast Cancer: The Quest for Markers and Therapeutic Targets. J. Proteome Res. **2008,** 7 (4), 1403-1411.
11. (a) Yu, K.; Toral-Barza, L.; Discafani, C.; Zhang, W. G.; Skotnicki, J.; Frost, P.; Gibbons, J. J., mTOR, a novel target in breast cancer: the effect of CCI-779, an mTOR inhibitor, in preclinical models of breast cancer. Endocr. Relat. Cancer. **2001,** 8 (3), 249-258; (b) Lee, H.-R.; Hwang, K.-A.; Nam, K.-H.; Kim, H.-C.; Choi, K.-C., Progression of breast cancer cells was enhanced by endocrine-disrupting chemicals, triclosan and octylphenol, via an estrogen receptor-dependent signaling pathway in cellular and mouse xenograft models. Chem. Res. Toxicol. **2014,** 27 (5), 834-842.
12. Trott, O.; Olson, A. J., AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. **2010,** 31 (2), 455-461.
13. Kolšek, K.; Mavri, J.; Sollner Dolenc, M.; Gobec, S.; Turk, S., Endocrine Disruptome—An Open Source Prediction Tool for Assessing Endocrine Disruption Potential through Nuclear Receptor Binding. J. Chem. Inf. Model. **2014,** 54 (4), 1254-1267.
14. Montes-Grajales, D.; Olivero-Verbel, J., Computer-aided identification of novel protein targets of bisphenol A. Toxicol. Lett. **2013,** 222 (3), 312-320.
15. Montes-Grajales, D.; Olivero-Verbel, J., EDCs DataBank: 3D-Structure database of endocrine disrupting chemicals. Toxicology **2015,** 327 (0), 87-94.

16. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. Nucleic Acids Res. **2000,** 28 (1), 235-242.
17. Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J., AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. J. Comput. Chem. **2009,** 30 (16), 2785-2791.
18. Meunier, B.; Dumas, E.; Piec, I.; Bechet, D.; Hebraud, M.; Hocquette, J.-F., Assessment of hierarchical clustering methodologies for proteomic data mining. J. Proteome Res. **2007,** 6 (1), 358-366.
19. Warnes, G. R.; Bolker, B.; Bonebakker, L.; Gentleman, R.; Huber, W.; Liaw, A.; Lumley, T.; Maechler, M.; Magnusson, A.; Moeller, S., gplots: Various R programming tools for plotting data. R package version **2009,** 2 (4).
20. Maldonado-Rojas, W.; Olivero-Verbel, J., Food-related compounds that modulate expression of inducible nitric oxide synthase may act as its inhibitors. Molecules **2012,** 17 (7), 8118-8135.
21. Wolber, G.; Langer, T., LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. J. Chem. Inf. Model. **2005,** 45 (1), 160-169.
22. Lorber, M.; Schecter, A.; Paepke, O.; Shropshire, W.; Christensen, K.; Birnbaum, L., Exposure assessment of adult intake of bisphenol A (BPA) with emphasis on canned food dietary exposures. Environ. Int. **2015,** 77, 55-62.
23. Fang, X.; Cao, S.; Liu, R., Interaction of Bisphenol A with Bovine Hemoglobin Using Spectroscopic and Molecular Modeling Methods. Appl. Spectrosc. **2011,** 65 (11), 1250-1253.
24. Montgomery, C.; Pei, Z.; Watkins, P. A.; Miziorko, H. M., Identification and Characterization of an Extramitochondrial Human 3-Hydroxy-3-methylglutaryl-CoA Lyase. J. Biol. Chem. **2012,** 287 (40), 33227-33236.
25. Tzitzilonis, C.; Eichmann, C.; Maslennikov, I.; Choe, S.; Riek, R., Detergent/Nanodisc Screening for High-Resolution NMR Studies of an Integral Membrane Protein Containing a Cytoplasmic Domain. PLoS ONE **2013,** 8 (1), e54378.
26. Maruyama, K.; Akiyama, M.; Kokame, K.; Sekiya, A.; Morishita, E.; Miyata, T., ELISA-Based Detection System for Protein S K196E Mutation, a Genetic Risk Factor for Venous Thromboembolism. PLoS ONE **2015,** 10 (7), e0133196.
27. Rimmerman, N.; Ben-Hail, D.; Porat, Z.; Juknat, A.; Kozela, E.; Daniels, M. P.; Connelly, P. S.; Leishman, E.; Bradshaw, H. B.; Shoshan-Barmatz, V.; Vogel, Z., Direct modulation of the outer mitochondrial membrane channel, voltage-dependent anion channel 1 (VDAC1) by cannabidiol: a novel mechanism for cannabinoid-induced cell death. Cell. Death Dis. **2013,** 4, e949.
28. Seidel, S. A. I.; Dijkman, P. M.; Lea, W. A.; van den Bogaart, G.; Jerabek-Willemsen, M.; Lazic, A.; Joseph, J. S.; Srinivasan, P.; Baaske, P.; Simeonov, A.; Katritch, I.; Melo, F. A.; Ladbury, J. E.; Schreiber, G.; Watts, A.; Braun, D.; Duhr, S., Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. Methods **2013,** 59 (3), 301-315.

29. Xie, X.; Wang, X.; Xu, X.; Sun, H.; Chen, X., Investigation of the interaction between endocrine disruptor bisphenol A and human serum albumin. Chemosphere **2010,** 80 (9), 1075-1080.
30. Whitmore, L.; Wallace, B. A., DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res. **2004,** 32 (Web Server issue), W668-W673.
31. Szklarczyk, D.; Franceschini, A.; Wyder, S.; Forslund, K.; Heller, D.; Huerta-Cepas, J.; Simonovic, M.; Roth, A.; Santos, A.; Tsafou, K. P.; Kuhn, M.; Bork, P.; Jensen, L. J.; von Mering, C., STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res. **2015,** 43 (D1), D447-D452.
32. Song, S.; Ruan, T.; Wang, T.; Liu, R.; Jiang, G., Distribution and Preliminary Exposure Assessment of Bisphenol AF (BPAF) in Various Environmental Matrices around a Manufacturing Plant in China. Environ. Sci. Technol. **2012,** 46 (24), 13136-13143.
33. Yang, Y.; Yu, J.; Yin, J.; Shao, B.; Zhang, J., Molecularly Imprinted Solid-Phase Extraction for Selective Extraction of Bisphenol Analogues in Beverages and Canned Food. J. Agric. Food. Chem. **2014,** 62 (46), 11130-11137.
34. (a) Lozada, K. W.; Keri, R. A., Bisphenol A increases mammary cancer risk in two distinct mouse models of breast cancer. Biol. Reprod. **2011,** 85 (3), 490-497; (b) Starovoytov, O. N.; Liu, Y.; Tan, L.; Yang, S., Effects of the Hydroxyl Group on Phenyl Based Ligand/ERRγ Protein Binding. Chem. Res. Toxicol. **2014,** 27 (8), 1371-1379.
35. Moore, K. H.; Bertram, K. A.; Gomez, R. R.; Styner, M. J.; Matej, L. A., Sex hormone binding globulin mRNA in human breast cancer: Detection in cell lines and tumor samples. J. Steroid Biochem. Mol. Biol. **1996,** 59 (3–4), 297-304.
36. Naik S L, D.; Hedau, S.; Bahadur, A.; Saha, R.; Kaur, S.; Ray, A., Sex hormone binding globulin in breast cancer. Indian J Clin Biochem **2008,** 23 (3), 250-254.
37. Miguel-Queralt, S.; Hammond, G. L., Sex Hormone-Binding Globulin in Fish Gills Is a Portal for Sex Steroids Breached by Xenobiotics. Endocrinology **2008,** 149 (9), 4269-4275.
38. Cherkasov, A.; Shi, Z.; Fallahi, M.; Hammond, G. L., Successful in Silico Discovery of Novel Nonsteroidal Ligands for Human Sex Hormone Binding Globulin. J. Med. Chem. **2005,** 48 (9), 3203-3213.
39. Hong, E.-J.; Sahu, B.; Jänne, O. A.; Hammond, G. L., Cytoplasmic accumulation of incompletely glycosylated SHBG enhances androgen action in proximal tubule epithelial cells. Mol. Endocrinol. **2010,** 25 (2), 269-281.
40. Oakley, G. G.; Devanaboyina, U.-s.; Robertson, L. W.; Gupta, R. C., Oxidative DNA Damage Induced by Activation of Polychlorinated Biphenyls (PCBs): Implications for PCB-Induced Oxidative Stress in Breast Cancer. Chem. Res. Toxicol. **1996,** 9 (8), 1285-1292.
41. Seewaldt, V. L.; Johnson, B. S.; Parker, M. B.; Collins, S. J.; Swisshelm, K., Expression of retinoic acid receptor beta mediates retinoic acid-induced growth arrest and apoptosis in breast cancer cells. Cell Growth Differ. **1995,** 6 (9), 1077-1088.
42. (a) Bruning, P. F.; Bonfrèr, J. M.; Hart, A. A., Non-protein bound oestradiol, sex hormone binding globulin, breast cancer and breast cancer risk. Br. J. Cancer **1985,** 51 (4), 479-484; (b) Moore, J. W.; Key, T. J.; Bulbrook, R. D.; Clark, G. M.; Allen, D. S.; Wang, D. Y.;

Pike, M. C., Sex hormone binding globulin and risk factors for breast cancer in a population of normal women who had never used exogenous sex hormones. Br. J. Cancer **1987,** 56 (5), 661-666.

43. Bauer, K. R.; Brown, M.; Cress, R. D.; Parise, C. A.; Caggiano, V., Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype. Cancer **2007,** 109 (9), 1721-1728.
44. (a) Brachmann, S. M.; Hofmann, I.; Schnell, C.; Fritsch, C.; Wee, S.; Lane, H.; Wang, S.; Garcia-Echeverria, C.; Maira, S.-M., Specific apoptosis induction by the dual PI3K/mTor inhibitor NVP-BEZ235 in HER2 amplified and PIK3CA mutant breast cancer cells. Proc. Natl. Acad. Sci. U.S.A. **2009,** 106 (52), 22299-22304; (b) Kalinsky, K.; Jacks, L. M.; Heguy, A.; Patil, S.; Drobnjak, M.; Bhanot, U. K.; Hedvat, C. V.; Traina, T. A.; Solit, D.; Gerald, W.; Moynahan, M. E., PIK3CA Mutation Associates with Improved Outcome in Breast Cancer. Clin. Cancer Res. **2009,** 15 (16), 5049-5059.
45. Roche, O.; Schneider, P.; Zuegge, J.; Guba, W.; Kansy, M.; Alanine, A.; Bleicher, K.; Danel, F.; Gutknecht, E.-M.; Rogers-Evans, M.; Neidhart, W.; Stalder, H.; Dillon, M.; Sjögren, E.; Fotouhi, N.; Gillespie, P.; Goodnow, R.; Harris, W.; Jones, P.; Taniguchi, M.; Tsujii, S.; von der Saal, W.; Zimmermann, G.; Schneider, G., Development of a Virtual Screening Method for Identification of "Frequent Hitters" in Compound Libraries. J. Med. Chem. **2002,** 45 (1), 137-142.
46. Ohyama, K.-I.; Satoh, K.; Sakamoto, Y.; Ogata, A.; Nagai, F., Effects of Prenatal Exposure to Styrene Trimers on Genital Organs and Hormones in Male Rats. Exp. Biol. Med. **2007,** 232 (2), 301-308.
47. Staskal, D. F.; Diliberto, J. J.; DeVito, M. J.; Birnbaum, L. S., Inhibition of Human and Rat CYP1A2 by TCDD and Dioxin-like Chemicals. Toxicol. Sci. **2005,** 84 (2), 225-231.
48. Thompson, R. C.; Moore, C. J.; vom Saal, F. S.; Swan, S. H., Plastics, the environment and human health: current consensus and future trends. Philos. Trans. R. Soc. Lond. B. Biol. Sci. **2009,** 364 (1526), 2153-2166.
49. Deceuninck, Y.; Bichon, E.; Marchand, P.; Boquien, C.-Y.; Legrand, A.; Boscher, C.; Antignac, J.; Le Bizec, B., Determination of bisphenol A and related substitutes/analogues in human breast milk using gas chromatography-tandem mass spectrometry. Anal Bioanal Chem **2015,** 407 (9), 2485-2497.
50. Biscardi, J. S.; Ishizawar, R. C.; Silva, C. M.; Parsons, S. J., Tyrosine kinase signalling in breast cancer: Epidermal growth factor receptor and c-Src interactions in breast cancer. Breast Cancer Res. **2000,** 2 (3), 203-210.
51. Madhavan, S.; Gusev, Y.; Singh, S.; Riggins, R. B., ERRγ target genes are poor prognostic factors in Tamoxifen-treated breast cancer. J. Exp. Clin. Cancer Res. **2015,** 34 (1).
52. Su, N.; Qiu, H.; Chen, Y.; Yang, T.; Yan, Q.; Wan, X., miR-205 promotes tumor proliferation and invasion through targeting ESRRG in endometrial carcinoma. Oncol. Rep. **2013,** 29 (6), 2297-2302.
53. Fernandez, M. F.; Arrebola, J. P.; Taoufiki, J.; Navalón, A.; Ballesteros, O.; Pulgar, R.; Vilchez, J. L.; Olea, N., Bisphenol-A and chlorinated derivatives in adipose tissue of women. Reprod. Toxicol. **2007,** 24 (2), 259-264.

54. Eriksen, K. T.; Sørensen, M.; McLaughlin, J. K.; Lipworth, L.; Tjønneland, A.; Overvad, K.; Raaschou-Nielsen, O., Perfluorooctanoate and Perfluorooctanesulfonate Plasma Levels and Risk of Cancer in the General Danish Population. J. Natl. Cancer Inst. **2009,** 101 (8), 605-609.
55. Seinen, W.; Lemmen, J. G.; Pieters, R. H. H.; Verbruggen, E. M. J.; van der Burg, B., AHTN and HHCB show weak estrogenic — but no uterotrophic activity. Toxicol. Lett. **1999,** 111 (1–2), 161-168.
56. Huang, Z.; Chen, G.; Shi, P., Emodin-induced apoptosis in human breast cancer BCap-37 cells through the mitochondrial signaling pathway. Arch. Pharm. Res. **2008,** 31 (6), 742-748.
57. Schlumpf, M.; Jarry, H.; Wuttke, W.; Ma, R.; Lichtensteiger, W., Estrogenic activity and estrogen receptor ß binding of the UV filter 3-benzylidene camphor: Comparison with 4-methylbenzylidene camphor. Toxicology **2004,** 199 (2), 109-120.
58. Tarnow, P.; Tralau, T.; Hunecke, D.; Luch, A., Effects of triclocarban on the transcription of estrogen, androgen and aryl hydrocarbon receptor responsive genes in human breast cancer cells. Toxicol. In Vitro **2013,** 27 (5), 1467-1475.
59. Rundle, A.; Tang, D.; Hibshoosh, H.; Estabrook, A.; Schnabel, F.; Cao, W.; Grumet, S.; Perera, F. P., The relationship between genetic damage from polycyclic aromatic hydrocarbons in breast tissue and breast cancer. Carcinogenesis **2000,** 21 (7), 1281-1289.
60. Lippold, T. G.; Kohli, N.; Endo, P. D.; Hui, H. K.; Lin, S.-M., Indicator dye for use in the detection of presence of oxidizing agents. Google Patents: 2005.
61. Meerts, I. A.; Letcher, R. J.; Hoving, S.; Marsh, G.; Bergman, A.; Lemmen, J. G.; van der Burg, B.; Brouwer, A., *In vitro* estrogenicity of polybrominated diphenyl ethers, hydroxylated PDBEs, and polybrominated bisphenol A compounds. Environ. Health Perspect. **2001,** 109 (4), 399-407.
62. Grishkovskaya, I.; Avvakumov, G. V.; Hammond, G. L.; Catalano, M. G.; Muller, Y. A., Steroid Ligands Bind Human Sex Hormone-binding Globulin in Specific Orientations and Produce Distinct Changes in Protein Conformation. J. Biol. Chem. **2002,** 277 (35), 32086-32093.
63. (a) Déchaud, H.; Ravard, C.; Claustrat, F.; de la Perrière, A. B.; Pugeat, M., Xenoestrogen interaction with human sex hormone-binding globulin (hSHBG)1. Steroids **1999,** 64 (5), 328-334; (b) Gale, W. L.; Patiño, R.; Maule, A. G., Interaction of xenobiotics with estrogen receptors α and ß and a putative plasma sex hormone-binding globulin from channel catfish (Ictalurus punctatus). Gen. Comp. Endocrinol. **2004,** 136 (3), 338-345.
64. Hildebrand, C.; Bocchinfuso, W. P.; Dales, D.; Hammond, G. L., Resolution of the steroid-binding and dimerization domains of human sex hormone-binding globulin by expression in Escherichia coli. Biochemistry **1995,** 34 (10), 3231-3238.

CHAPTER 6

Conclusions and final remarks

This thesis was conceived to explore the plausibility of endocrine disruptors to affect the signaling related to breast cancer through protein-ligand interactions. Interestingly, we found that these pollutants are able to bind a large number of proteins, not only related to this disease, but with other important processes as revealed by the inverse virtual screening with BPA. This fact could be associated to structural properties shared by many EDCs such as the high hydrophobicity, small molecular weight and the presence of routable bonds, which facilitate the non-covalent interaction on the binding site of different proteins. The promiscuity of these compounds is concerning as these pollutants may affect numerous processes in the organism, as suggested by epidemiological data that indicates that the exposure to xenoestrogens induce diverse deleterious effects, ranging from diabetes to neurodevelopmental disorders. Therefore, the data generated in this work could be useful to guide the *in vitro* and *in vivo* screening process, which are imperative to establish the real impact of these chemicals in human and wildlife health, as starting point to take action on their regulation and use.

Other alarming fact could be the inclusion of new pollutants to the market with unknown effects to replace the prohibited or regulated chemicals, as these could be even more aggressive than the original compounds. The vHTS showed that BPA analogs tend to bind stronger on a greater number of breast cancer proteins than BPA, and even when the population has been informed to avoid this compound in food plastic containers, the new substituents may be present in those labeled as BPA-free.

Finally and undoubtedly, one of the greatest contributions to the field of EDCs made in this thesis project was the creation of the first database with 3D structures available worldwide. Currently, this has been used in more than one hundred countries (Figure 6.1), and is expected to be a valuable tool to help research, education and information to the general population regarding the risk of these pollutants.

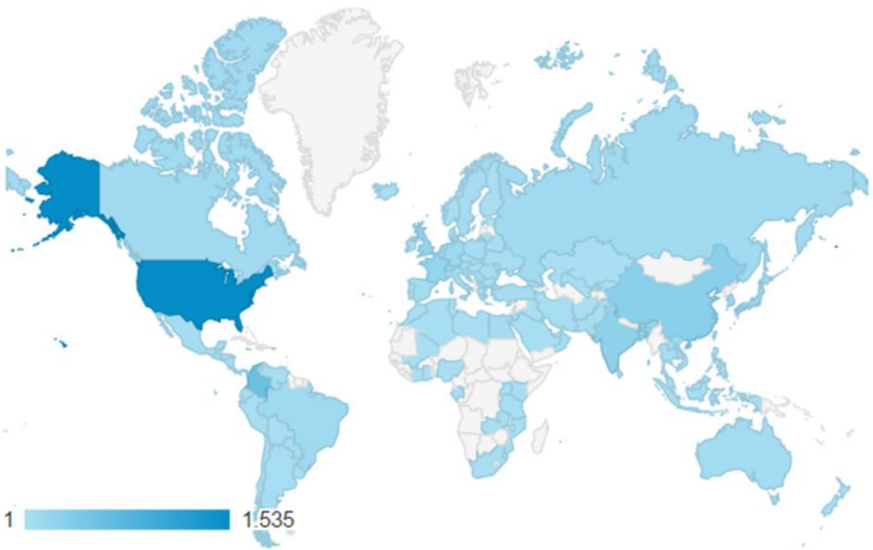


Figure 6.1. Geographic information of the visits to EDCs DataBank, by Google analytics (22-mar-2014 to 1-aug-2015).

Some key points regarding this thesis are:

- EDCs DataBank is the unique database with three-dimensional structures of EDCs for virtual screening.

- EDCs DataBank is suitable for virtual screening and other computational approaches as the molecular structures are available in several formats.

- EDCs DataBank is a valuable repository to study the molecular basis of the interaction between these molecules and macromolecules involved in several pathologies.

- This database provides helpful information for research, academia and general population.

- BPA may interact with other targets different from the ER.

- BPA seemed to interact with proteins involved in several high prevalence diseases.

- Dual specificity protein kinases CLK4, CLK1 and CLK2 showed the best *in silico* affinity by BPA.

- Validation showed that the *in silico* docking protocol reproduces well the BPA pose on the crystallographic complex with ESRRG.

- BPA toxicological effects could be explained through its interactions with key proteins in signaling pathways.

- EDCs have the potential to target breast cancer proteins.

- Many EDCs are frequent hitters and their mechanisms of action may be very complex, affecting several pathways.

- Derivatives and substituents of BPA presented high affinity for breast cancer proteins.

- Experimental validation was performed on testing the interaction between BPA and one of its predicted targets by circular dichroism. The main conformational change occurred after BPA-SHBG binding was an increment of the alpha helix content.

CHAPTER 7

Annexes

In this chapter, the supplementary information available on the web (https://goo.gl/tEG3ZR) is described.

ANNEXES OF THE CHAPTER 3

Annex 3.1

The factor scores resultant of the PCA of the 300 independent molecular descriptors of the EDCs in EDCs DataBank, as well as the eigenvalues for the three principal components are shown in Annex 3.1. (see Microsoft Excel book: "Annex_3_1") In the following tables:

- Table 1. Principal component scores.
- Table 2. Eigenvalues for the three principal components.

The principal components scores show all the components determined by the PCA analysis and the punctuation for each molecule; and the revision of the eigenvalues for the best principal components gives an idea of which of the molecular descriptors contributes in a major proportion to them.

ANNEXES OF THE CHAPTER 4

Annex 4.1

The extended results of the virtual screening between BPA and proteins involved in different pathways are presented and organized by categories in

this Annex (see Microsoft Excel book: "Annex_4_1") as well as the results of the *in silico* studies of the protein-ligand interactions of the best complexes.

- Table 1. Docking affinity scores for BPA binding to nuclear receptors.
- Table 2. Docking affinity scores for BPA binding to circadian clock related proteins.
- Table 3. Docking affinity scores for BPA binding to insulin receptor pathway proteins.
- Table 4. Docking affinity scores for BPA binding to serum proteins.
- Table 5. Docking affinity scores for BPA binding to breast cancer proteins.
- Table 6. Docking affinity scores for BPA binding to other proteins from data mining.
- Table 7. Interacting residues between BPA and proteins with the best affinity score.

ANNEXES OF THE CHAPTER 5

Annex 5.1

The output of text mining regarding breast cancer related proteins and the high-throughput virtual screening results, in terms of docking affinity (kcal/mol) and classified according to the source of exposure, are presented in this Annex (see Microsoft Excel book: "Annex_5_1"), as follows:

- Table 1. List of proteins and genes associated to breast cancer in the literature according to FABLE (Accessed: 12 dic 2012).
- Table 2. Docking affinity scores for dioxins and related molecules with breast cancer associated proteins.
- Table 3. Docking affinity scores for EDCs from plastics and other polymers with breast cancer associated proteins.
- Table 4. Docking affinity scores for EDCs from everyday products with breast cancer associated proteins.

- Table 5. Docking affinity scores for miscellaneous EDCs with breast cancer associated proteins.

Annexes 5.2-5.5

This group of annexes shows an interesting graphical view of the docking results. It includes hierarchical heatmaps of each group of EDCs against the breast cancer proteins, in the color code those highlighted in red obtained good *in silico* docking affinity.

- Annex 5.2. Cluster heat map of the docking affinity scores obtained for EDCs in the group of dioxins and related molecules (columns) and breast cancer proteins (rows). (See PDF file: "Annex_5_2_Heatmap_Dioxins").
- Annex 5.3. Cluster heat map of the docking affinity scores obtained for EDCs in the group of plastics and other types of polymers (columns) and breast cancer proteins (rows). (See PDF file: "Annex_5_3_Heatmap_Plastics").
- Annex 5.4. Cluster heat map of the docking affinity scores obtained for EDCs in the group of Everyday Products (columns) and breast cancer proteins (rows). (See PDF file: "Annex_5_4_Heatmap_Everyday").
- Annex 5.5. Cluster heat map of the docking affinity scores obtained for EDCs in the group of Miscellaneous (columns) and breast cancer proteins (rows). (See PDF file: "Annex_5_5_Heatmap_Miscellaneous").

ACKNOWLEDGEMENTS

I would like to thank science and religion, to provide me the necessary answers to exist without too many philosophical conflicts; but my main acknowledgement is for all the people who inspire me. The first place is for my family, Enrique Montes, Gloria Grajales, Ana María Montes, Andrés Montes and Sebastiana Castro, an amazing group of people with critical and independent thinking, for all their support and inspiration. I would also like to thank one of the most influential school professors I had, Prof. Ariel Acosta, for helping me to spark the scientific curiosity. And to the Prof. Jordi Quintana, Gemma Arsequell, Gregori Valencia, and Daniel Blasi of the Scientific Park of Barcelona, for helping me to discover I wanted to do a Ph.D.

A special recognition to my advisor, Prof. Jesús Olivero, the most tireless person I have ever met, for founding this Ph.D. Program in Environmental Toxicology (DTA), for accepting me in his team since I was first semester student of bachelor degree, and the entire patience and effort that implies this non-easy educational process. All the academic community of the DTA and the Environmental and Computational Chemistry Group merit my respect and gratitude. They enriched my scientific carrier by learning from their very multidisciplinary experiences, knowledge and even emotional intelligence.

I am very grateful to Prof. Gonçalo Bernardes, an extraordinary scientist with cutting-edge ideas, to giving me the opportunity to work in his group at the University of Cambridge and all the great people of GBernardes Lab, specially Prof. Francisco Corazana, Prof. Omar Boitera, Dr. Nuria Martinez, Dr. Victor Somovilla, Allyson, and the amazing person who taught me all of what I needed to know about experimental protein expression in less than one week, Beatrice Amgarten. I would also want to thank people from other groups and technical staff that helped me with the project there, some of them are Ryan, Alex, Marco di Antonio and Yiyan.

Thanks to my enthusiastic friends, Ingrid Sarmiento and Lorena Martinez, to keep me near the things I like the most, art and friendship, and for all the fantastic moments we spent together. I would also like to thank all the good international friends I have, for inspiration and support; and to all those who helped me with revisions and advices for EDCs DataBank, specially to the novation program students, Nely Mejía and Jessica Cerón; to my friends Sebastian González and Carlos Kleber for helping me to revise the code till late at night; and to my brother, Andrés Montes, for designing the logo of the database and the source icons.

Finally, I would like to thank the University of Cartagena and Colciencias for financial support (Grant: 567-2012). As well as the graduate student travel award of the Society of Toxicology of US (SOT, 2013); and the Sao Paulo Research Foundation, FAPESP, for the travel awards for the courses Sao Paulo School of Advanced Sciences: Advances on Molecular Oncology: Translating Molecular Biology into Cancer Treatment (2013) and Advanced topics in computational Biology - Agrochemical and Drug Design (2012).

GLOSSARY

AC_{50}: it is the half-maximal activity concentration, also known as experimental activating concentration 50%.

Blind docking: it is the process by which a standard docking tool is applied to a whole surface of the protein.

Circular dichroism (CD): it is a spectroscopic method used to determine the secondary structural content of proteins.

Contact residues: are those participating in the protein-ligand binding, through polar or van der Waals interactions.

CSS: cascading style sheets is a language that allows the programming of style the websites, in term of colors, font sizes, distribution of the space, among others; making the code compact, and the webpage good looking, consistently displayable, and quick and efficient to change.

Docking: it is an *in silico* technique to identify the correct poses in protein-ligand complexes, simultaneously predicting their binding affinity.

Flame retardants: these are chemicals added to a broad range of consumer products to reduce the fire-related injury and property damage.

HTML: the hypertext markup language is the standard markup language used to create web pages.

Inverse docking: it is a technique that has been used to identify potential receptor targets of small molecules through docking studies.

JavaScript: it is a programming language that allows developers to create client interaction interfaces for web applications.

JQuery: it is a JavaScript library that simplifies how you traverse HTML documents, handle events, perform animations, and so forth.

Microscale thermophoresis (MST): it is a technique based on thermophoresis that allows the analysis of protein interactions in free solution and with low sample consumption.

MySQL: my structured query language (MySQL) is a database management system that allows the communication with the relational database.

Non-redundant: it means that there are no duplicates in the data set.

Pharmacophore features: it includes chemical and structural properties related to the binding. These include: hydrogen bond donors, hydrogen bond acceptors, and hydrophobic and ring aromatic features, among others.

PHP: PHP hypertext preprocessor is scripting language that allows embedding executable code into HTML.

POPs: persistent organic pollutants are a group of chemicals that persist in the environment, bioaccumulate through the food cycle, and exhibit toxic effects that may affect the health of humans and wildlife.

Recombinant DNA technology: it is a series of procedures used to connect DNA segments of different organisms, containing a gene of interest. This construct can be inserted in a host organism, and used for the production of proteins.

Text mining: it is the discovery of new, previously unknown information, by automatically extracting information from different written resources using computer aided tools.

Virtual high-throughput screening (vHTS): This is an approach that allows the *in silico* identification of biologically relevant molecules amongst large libraries of compounds.

Xenobiotic: it is a foreign chemical found within an organism that is not normally naturally produced by or expected to be present within that organism.

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

Spanish

BPA	bisfenol A
EDCs	disruptores endocrinos
ER	receptor de estrógenos
ESR1	receptor de estrógenos 1
OMS	Organización Mundial de la Salud
PDB	Protein Data Bank
PFOS	ácido sulfónico de perfluorooctano
PGR	receptor de progesterone
SHBG	globulina fijadora de hormonas sexuales
vHTS	cribado virtual de alto rendimiento

English

AC_{50}	experimental activating concentration 50%
ADIPOQ	adiponectin
ADORA2A	adenosine receptor A2A
AMPK	AMP-activated protein kinase
APOD	apolipoprotein D
AR	androgen receptor
BPA	bisphenol A
BPAF	bisphenol AF
BRCA1	breast cancer type 1 susceptibility protein
BRCA2	breast cancer type 2 susceptibility protein
C	Control
CAR	constitutive androstane receptor
CAT	Catalase
CCL2	C-C motif chemokine 2
CCRIS	Chemical Carcinogenesis Research Information System
CD	circular dichroism

CD40	tumor necrosis factor receptor superfamily member 5
CD69	early activation antigen CD69
CDK2	cyclin-dependent kinase 2
CID	PubChem compound identifier
CK5	Cytokeratin
CLK	Cdc2-like kinase
CSS	cascading style sheets
CTD	Comparative Toxicogenomics Database
CXCL10	C-X-C motif chemokine 10
CYP1A1	cytochromeP450 1A1
CYP1A2	cytochrome P450 1A2
CYP2C19	cytochrome P450 2C19
CYP3A4	cytochrome P450 3A4
DBP	dibutyl phthalate
DEHP	bis(2-ethylhexyl)phthalate
DFT	density functional theory
EADB	Estrogenic Activity Database
EAR-7	thyroid hormone receptor alpha
EDC	endocrine disrupting chemical
EDCs	endocrine disrupting chemicals
EDID	Endocrine Disrupting Chemicals–Diet Interaction Database
EDKB	Established Knowledge Base for Endocrine Disrupting Chemicals
EDPSD	Endocrine Disruptor Priority Settings Database
EPA	US Environmental Protection Agency
ER	estrogen receptor
ESR1	estrogen receptor 1
ESRRG	estrogen-related receptor gamma
Fable	fast automated biomedical literature extraction
FGFR2	fibroblast growth factor receptor 2
FPLC	fast protein liquid chromatography
GABA	γ-aminobutyric acid
GENE-TOX	Genetic Toxicology Data Bank
HBCD	Hexabromocyclododecane
HER-1	human epidermal growth factor receptor

HER-2	human epidermal growth factor receptor 2
HSDB	Hazardous Substances Data Bank
HTML	hypertext markup language
IGF1R	type 1 insulin-like growth factor receptor
IHop	Information Hyperlinked over Proteins
IPTG	isopropyl 1-thio-ß-d-galactopyranoside
IRIS	Integrated Risk Information System
ITER	International Toxicity Estimates for Risk
MAPK	Ras–mitogen-activated protein kinase
MMP-1	matrix metalloproteinase-1
MMP-13	matrix metalloproteinase-13
MMP-8	matrix metalloproteinase-8
MMP-9	matrix metalloproteinase-9
MST	microscale thermophoresis
mTOR	mammalian target of rapamycin
MW	molecular weight
MWCO	molecular weight cutoff
MySQL	my structured query language
NR3C1	glucocorticoid receptor
PAHs	polycyclic aromatic hydrocarbons
PBBs	polybrominated biphenyls
PBDEs	polybrominated diphenyl ethers
PBS	phosphate-buffered saline
PCA	principal component analysis
PCBs	polychlorinated biphenyls
PCDDs	dibenzo-p-dioxins
PCDFs	Dibenzofurans
PDB	Protein Data Bank
PFOS	perfluorooctane sulfonic acid
PGC-1α	peroxisome proliferator-activated receptor-γ coactivator
PGR	progesterone receptor
PHP	PHP hypertext preprocessor
PI3K	PI3-kinase
PIK3CA	phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic

	subunit alpha isoform
PIK3CG	phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit gamma isoform
PKCθ	protein kinase C theta
POPs	persistent organic pollutants
RAR-alpha	retinoic acid receptor alpha
RARB	retinoic acid receptor beta
RB	retinoblastoma-associated
RBD	Development of the Receptor Database
RMSD	root mean square deviation
RORA	nuclear receptor ROR-alpha
RORC	nuclear receptor ROR-gamma
ROS	reactive oxygen species
R-PTP-epsilon	receptor-type tyrosine-protein phosphatase epsilon
RXR-beta	retinoic acid receptor RXR-beta
SHBG	sex hormone-binding globulin
SMILES	simplified molecular input line entry specification
SR	serine-arginine-rich
SRC	proto-oncogene tyrosine-protein kinase Src
t=O/N	overnight incubation
t=0	time zero
TBBPA	tetrabromobisphenol A
TCDD	2,3,7,8-tetrachlorodibenzo-p-dioxin
TGFB1	transforming growth factor beta-1
TGF-ß	growth factor-beta
TIMP-3	metalloproteinase inhibitor 3
TOP2A	DNA topoisomerase 2-alpha
TOXNET	Toxicology Data Network
TP53	and tumor protein
RTK	tyrosine kinase receptors
UNEP	United Nations Environment Programme
VCAM1	vascular cell adhesion protein 1
VDR	vitamin D3 receptor
vHTS	virtual high-throughput screening
WHO	World Health Organization