UNIVERSITÀ DI PARMA

UNIVERSITÀ DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN

"Scienze degli Alimenti"

CICLO XXXIV

A Food Contact Materials 3D database for food safety using nuclear receptors (ER and AR) as biological cancer target

Coordinatore:

Chiar.ma Prof. **Chiara Dall'Asta**

Tutore:

Chiar.mo Prof. **Pietro Cozzini**

Dottorando: **Francesca Cavaliere**

Anni Accademici 2018/2019 - 2020/2021

Table of Contents

Abstract

Food contact chemicals (FCCs) include a very heterogeneous class of molecules and the term refers to all chemicals that come into contact with food during the entire food chain, from the raw material to the final products. FCCs mainly derived from three different sources: natural, intentional, and unintentional.

Microorganisms, plants, and animals, which directly interact with food, can produce and release molecules as a natural defence mechanism. These molecules may have a toxic effect on other organisms, including humans. Mycotoxins are a typical example of food contact chemicals that naturally occur in food.

Two other important sources should be considered: the molecules that are intentionally added to food and those unintentionally present. To the first class belongs all substances which are intentionally added to enhance some food properties, and to ensure and maintain food safety, like preservatives. Instead, all food contact chemicals that accidentally come into contact with food, as contaminant molecules belong to the second class. They can occur at different levels of the food chain, e.g., processing, storing, packaging, or consumption.

It is often difficult to evaluate the total extent of exposure as well as their safety for humans and the environment because information on food contact chemicals is widespread across multiple online sources, which have also a certain degree of errors. Moreover, these errors propagate very easily across the internet. Therefore, the data are not standardized, and this issue limited the analyses to evaluate the safety of these compounds. To make matters worse, food contact chemicals include thousands of chemicals which make it difficult to test their safety using experimental methods. They would require a long time and high-costs to be performed. In fact, a lot of chemicals are waiting for their safety evaluation.

Based on these considerations, this Doctoral Thesis aims at developing a database of all food contact chemicals (foodchem DB) and at screening their endocrine-disrupting properties using computational methods. This may overcome experimental limits (cost and time) in order to prioritize the molecules that require further investigations.

In the first part of this PhD thesis, some specific case studies have been considered. First, the class of bisphenols has been evaluated for the estrogenic and androgenic pathways.

Second, some pesticides have been evaluated considering an animal model, *Daphnia Magna*, with the aim to show the utility of computational methods to also reduce experimental tests.

Once the benefit of using computational methods to identify endocrine disruptor chemicals has been evaluated, in the second part of this PhD thesis, it has been considered a huge number of molecules. Firstly, it has been developed a food contact chemical database to take into account all molecules that come into contact naturally, intentionally, or unintentionally with food. The foodchem DB lists 11059 substances divided into ten subclasses: additive, bisphenol, dioxin, flavouring, furan, mycotoxin, PCB, pesticide, phthalate, and the food packaging forum (FPF) database lists (FCCdb). To pursue the objective to include all these molecules, regulatory lists, industry inventories, and public data sources have been considered. Moreover, in this section, it is also addressed the evaluation of the endocrine-disrupting properties undertaken by food contact chemicals using computational methods. The importance of the consensus scoring approach to overcome molecular docking limitations has also been illustrated. In this study, several nuclear receptors have been considered to evaluate a general endocrine disruptor effect produced by all food contact chemicals. However, to consider specifically the involvement of these molecules in breast and prostate cancers, in Chapter 8, a little focus has been made on estrogen and androgen receptors since they are two well-known nuclear receptors involved in these cancers.

The combination of molecular docking and consensus scoring techniques used in these studies highlighted how these methods are useful to study the interaction between food contact chemicals and nuclear receptors, allowing the identification of the chemicals which have most the suitable physical-chemical characteristic to interact and disrupt the endocrine pathway. On the other side, the effect of that interaction on the receptor (from a structural point of view) can be analysed using molecular dynamic simulation. Molecules can affect nuclear receptors activity by acting as agonist or antagonist compounds. However, it is not fully understood yet how molecules induce an antagonistic effect on the androgen receptor. Thus, to study the mechanism of action (MoA) of antiandrogens on the homodimer stability of the androgen receptor, in Chapter 9, molecular dynamics simulation has been carried out. This study might be useful to decipher food contact chemical effects on the androgen receptor using computational methods.

Finally, in the last chapter, two future perspectives have been illustrated. In study one, it is highlighted the advantage to use the isothermal titration calorimetry (ITC) to study protein-ligand interaction for hazard identification. In study 2, it is shown the use of two computational methods (machine learning and robust consensus scoring) to identify the substances of very high concern for the endocrine system among food contact chemicals.

CHAPTER 1

Food contact chemicals and *in silico* methods

1.1 Food contact chemicals

The term "Food contact chemicals" (FCCs) may be misleading since different scientists refer to them with a different meaning concerning the food molecules the term should include. Basically, food contact chemicals (FCCs) can be defined as all chemical substances which are not part of food items but that come into contact with it. The Food Packaging Forum (FPF) (https://www.foodpackagingforum.org/) is a foundation established to provide scientific information of high-quality related to food packaging and health. With the term food contact chemical, the FPF refers to all chemical constituents of food contact materials and finished food articles, considering also the substances that are intentionally added or present in food for other reasons (Muncke et al., 2020). This definition highlights two important terms in the food contact chemical context: food contact material (FCM) and food contact article (FCA). Although they are often used to refers to the same thing, it is not totally correct. While food contact material refers to the materials that come into contact with food, such as plastics, papers, glass, adhesive, etc., food contact article is the final product used to store and/or to contain food, such as bottles and wraps. In this latter case, FCA can be a combination of FCMs or can be composed of a single FCM. Food contact materials can be divided into two different subgroups: intentionally added substances (IASs) and non-intentionally added substances (NIASs). The former refers to all chemical components that are deliberately used to manufacture FCMs and FCAs. The latter term, NIASs, was firstly introduced in a European regulation (Commission Regulation (EU) No 10/2011) regarding plastic food contact materials, but they are not limited to plastic sources and can occur from all other food contact materials. NIASs are chemical components present in FCMs but that has not been added intentionally for a technical purpose during the production process and, thus, they do not have any specific function. NIASs can come from different sources: side products formed during the production of starting substances, breakdown products, or contaminants (impurities of raw materials and/or contaminants during the production). Food contact materials or articles are perhaps the most obvious example of food packaging. However, food can come into contact with materials in many other situations such as during its manufacture, transport, storage, preparation, and consumption. These include the materials used for storage vessels, conveyor belts, tubing, food preparation surfaces, and cooking and eating utensils (Barnes et al., 2006). Food contact materials (FCMs) can cause food contamination when one or more of its constituents (chemicals) migrates into

the food source. The migration phenomenon has been studied since the 1950s (Muncke et al., 2020). It might be a contradiction the fact that food packaging, which has the main purpose to protect food from external contamination, such as from physical damage, soiling, and microbial spoilage, can be itself a source of unwanted chemicals. The term migration usually describes a diffusion process that is subjected to both thermodynamic and kinetic control and can be described by Fick's first and second law (Arvanitoyannis and Bosnea, 2004). While the kinetic dimension dictates the velocity of the migration process, the thermodynamic dimension dictates the extension of substances that will be transferred at the end of migration, i.e., at the equilibrium. The migration of chemicals from FCMs has two major impacts on food since it is both a food safety and a food quality issue. Some substances can have adverse health effects in humans or the environment (food safety), but, at the same time, they can also affect the taint or odour of food, reducing the consumer appeal (food quality).

Ensuring food safety is not a simple task. Different aspects should be considered, such as the good manufacturing practice (GMP) that must be followed during the food contact materials manufacturing chain, or the procedure that should be adopted to perform safety assessments of FCM constituents.

In Europe, FCMs and FCAs are regulated by two different types of legislation: the Community legislation adopted by the EU, and the national legislation adopted by Member states. The Regulation (EC) No 1935/2004, commonly denoted as FCM Framework Regulation, sets out the general principles to be complied with by all food contact materials. The principles mainly establish that FCMs: *a*) must be sufficiently inert to avoid the transfer of their constituents into food at levels harmful to human health; *b*) must not change food composition or deteriorate food organoleptic properties. In addition to this general legislation, some specific materials are covered by specific EU measures (ceramic materials, regenerated cellulose film, plastics, active and intelligent materials). Moreover, other EU legislations are also enacted for specific substances. For example, the Commission Regulation (EU) 2018/213 regards the use of bisphenol A in varnishes, coatings, and plastics intended to come into contact with food, and it amends the Regulation (EU) No 10/2011. The use of certain epoxy derivatives is restricted by the Commission Regulation 1895/2005/EC. To the promulgation of the EC No 1935/2004 was followed the Commission Regulation (EC) No 2023/2006 on good manufacturing practice (GMP) for materials and articles intended to come into contact with food. As the

regulation itself states, GMP "*means those aspects of quality assurance which ensure that materials and articles are consistently produced and controlled to ensure conformity with the rules applicable to them and with the quality standards appropriate to their intended use by not endangering human health or causing an unacceptable change in the composition of the food or causing a deterioration in the organoleptic characteristics thereof*". Wherever EU measures are absent, EU Member State countries can define their national provisions on food contact materials under Article 6 of Regulation 1935/2004.

Food contact material is not the only source of unintentional molecules present in food. This class can include all chemicals which accidentally contaminate food product, like environmental pollutants, chemical residues due to human activities such as farming, industry, or car exhausts, or as a result of human cooking and processing. It is a very broader and heterogeneous class consisting of metals, dioxins, polychlorinated biphenyls (PCBs), furans, acrylamide, pesticides, veterinary medicines, and feed additives.

Several metals, such as arsenic, cadmium, lead. and mercury can be accumulated in the environment due to their presence at different levels, e.g., soil, water, and atmosphere. Once in the environment, these compounds might also contaminate food and water. Moreover, metals can also occur as a result of human activities, such as farming, industry or car exhausts, or from contamination during food processing and storage. Several factors affect food contamination levels. The varying exposure is the main factor, but the level of contamination may be also influenced by the different plant uptake mechanisms (Stasinos et al., 2014). The EU Regulation 315/93/EEC sets out the general principles on contaminants in food. Generally, the legislation explains that food contaminants must be as low as can, setting the maximum level to protect public health and that all food containing a contaminant at a toxicological level must not be placed on the market. An additional regulation, the Regulation EC 1881/2006, defines the maximum levels for specific contaminants, including lead, cadmium, mercury, and inorganic tin.

Dioxins and polychlorinated biphenyls (PCBs) belong to the class of unintentionally added substances. These molecules are characterized by pronounced chemical stability and persistence in the environment. Furthermore, they are highly lipophilic and show low volatility. Since they accumulate in the environment and food chain (especially in animal fat), food is the major source of exposure in the general population. The term "dioxins" is commonly used to indicate two different chemical subfamilies: polychlorinated dibenzo-p-dioxins (PCDDs) and polychlorinated dibenzofurans (PCDFs). Dioxin' food

contamination mainly comes from thermal and industrial processes as unwanted and often unavoidable by-products, while PCBs had widely used in numerous industrial applications for several years until the 1980s when they were banned in most countries. Although dioxins and PCBs are found at low levels in many foods, they are characterized by a metabolic half-life of several years in the human body, and longer-term exposure to these substances causes a range of adverse effects in humans. Since the toxic effects of these compounds, in 2001, the European Commission establishes for the first time maximum acceptable levels for dioxins in food, which were extended to dioxin-like PCBs in 2006. After that, the regulation was updated with two other regulations: Regulation EU 1259/2011 and Regulation EU 277/2012, which set the new maximum levels for non-dioxin-like PCBs in food and feed, respectively.

Food contamination can derive at different levels. Industrial food processing as well home-made cooking, where high temperatures are used, can determine the release of furan substances in the food product. Furans are organic molecules formed during heat treatment and they have shown carcinogenic properties in animal experiments. The data collected between 2004 and 2010 and analysed by EFSA (European Food Safety Agency), derived from analytical tests for measuring furan levels in heat-treated commercial food products, show that furan exposure is highest in adults and toddlers with coffee and jarred baby foods as the food major sources (Crews, 2009; Fromberg et al., 2009). Since furans levels detected in heat-treated foodstuffs were close to the doses that cause carcinogenic effects in experimental animals, EFSA considers furans as an urgent issue and claims for further data on both toxicity and exposure to make a reliable risk assessment (Commission Recommendation of 28 March 2007).

High-temperature cooking (> 120 °C) determines the formation of acrylamide, which is naturally produced as a consequence of the Maillard Reaction. Acrylamide forms from sugars and amino acids, and it is mainly found in products like potato crisps, French fries, bread, biscuits, and coffee. It is defined as a contaminant in the Council Regulation (EEC) No 315/93. After a risk assessment, in 2015, EFSA concluded that acrylamide potentially increases the risk of developing cancer for consumers. Following the EFSA opinion, Member Stated started to determine appropriate regulatory measures to reduce its presence in food, and in 2017, the Commission Regulation (EU) 2017/2158 was adopted.

Other important sources of unintentionally food added substances are pesticides and veterinary medicines. Although it is obvious to think that pesticides and veterinary

medicines contaminate plants and animals, respectively, there is not a real demarcation line. Animals may be themselves exposed to residues of pesticides and contaminants due to their presence in food to which animals have been reared. Plants may be contaminated by veterinary medicines due to their releasing into the environment. Considering their intended use, animals may be treated with veterinary medicines to prevent or cure disease. Pesticide, instead, refers to all substances used for plant protection products, including biocides that are intended for non-plant uses. Moreover, the term pesticide also includes herbicides, fungicides, insecticides, acaricides, plant growth regulators, and repellents. According to the Special Report 2/2019 published by the European Court of Auditors, the use of 492 active pesticide' substances and 666 pharmacologically active substances are approved. Two different regulations regulate the two classes. The Commission Regulation (EU) No 37/2010 regulates the maximum residue limits of pharmacologically active substances in foodstuffs of animal origin, and the Commission Regulation (EC) No 1107/2009 concerns the placing of plant protection products on the market. Concerning animals, other important substances that may contaminate the food are feed additives, which are products used with the purpose to improve the quality of feed and food or to promote animal performances and heath, e.g., enhancing ingestion, absorption, assimilation of nutrients, and growth. Two different regulatory frameworks (Regulation 1831/2003 and Regulation 429/2008) set out the authorized additives for use in animal feed and provide the rules for the presentation of the application to authorise new feed additives.

Intentionally added substances refer to chemicals added to food for enhancing some properties, such as for prolonging shelf life, modifying its texture, making the products more appealing and attractive, and/or for improving its flavour. Food additives can be used for different purposes and the European Union legislation defines 26 "technological purposes". Among other things, they are mainly used as colorants, preservatives (for prolonging the shelf-life of foods by protecting them against micro-organisms), antioxidants (for protecting the food against oxidation), and flour treatment agents (for improving baking quality). All food additives are identified by an E number in the European Union. Flavourings are products added to food to enhance or modify odour, taste, and/or aroma. The safety evaluation of flavour and additive ingredients added to the food supply is an important subject and it is paramount to protect public health (Smith et al., 2018). The potential risks associated with food additives and flavourings concern

different aspects, such as the use of unauthorized molecules, the use of molecules that do not comply with purity criteria, and/or the use of excessive quantities. According to Annex I of the Special Report 2/2019, to date 334 food additives and 2549 food flavourings are approved in the European Union. Moreover, the term food flavouring also refers to molecules added as smoke flavourings and, with the Regulation 1321/2013 of 10 December 2013, 10 different substances have been authorized as primary products for the production of derived smoke flavourings. Food additives and food flavourings are regulated by the Regulation EC 1333/2008, which sets a list of approved molecules based on safety assessment and the technological need, and for ensuring that their use will not mislead consumers. Moreover, it also sets the rules to be followed: conditions of use, labelling, and procedures.

Man-made compounds are not the only molecules to be concerned about in the food contamination context. Several molecules are naturally occurring in the food supply due to their release in food products by plants, animals, or microorganisms. These toxins are produced by living organisms as a natural defence mechanism. Although they do not affect the organism itself, these molecules may be toxic for other organisms, including humans, when eaten. These chemicals may be very heterogenous in terms of their structure and may have different biological functions and toxicity. Moreover, some of them are not eliminated by cooking and/or freezing. Algae are a source of toxins that can contaminate shellfish (particularly, mussels, scallops, and oysters) and water. Since these molecules are retained in shellfish and fish when eaten, they can cause different side effects (Van Dolah, 2000) (paralysis, tingling, neurotoxic effect, etc.) in humans, mammals in general, and/or in fish itself. Cyanogenic glycosides, furocoumarins, pyrrolizidine alkaloids (PAs), solanines, and chaconine are phytotoxins produced by plants in response to stress, such as physical damage, UV light, attacks from insect pests, etc. These molecules may occur in different plants that are used as a food source and may determine different toxic effects to exposed humans that range from gastrointestinal problems to more adverse outcomes, such as cardiac arrest ((CONTAM) et al., 2019). Mycotoxins are probably the most known source of contamination. They are produced as secondary metabolites by certain types of fungi. Mycotoxins can enter the food chain as a result of pre- or post-harvest crop infection and they are found in a variety of food products, such as cereals, dried fruits, nuts, and spices. Most mycotoxins are chemically stable and may survive food processing. The presence of these molecules in food and feed

can cause very severe effects in humans and animals with symptoms ranging from gastrointestinal problems to immunodeficiency and cancer. Different Commission Regulations exist to regulate the maximum levels of toxins and mycotoxins in food and feed ((EC) No 1881/2006, (EC) No 401/2006, 2012/154/EU, 2014/662/EU). Moreover, following the conclusion' EFSA opinions, the Standing Committee recommended the monitoring of other plant toxins and mycotoxins.

1.2 Endocrine Disrupting Compounds (EDCs)

The endocrine system is a complex network that involves several actors: glands, hormones, and receptors. Hormones are produced by glands and, by using the bloodstream, they reach their target tissues. Working with the nervous system, hormones control important functions, such as reproduction, immunity, metabolism, and behaviour. The rapid development of industrial technologies has progressively increased the risk level for human health. In fact, all living beings are daily exposed to a mixture of exogenous substances that may act as endocrine disruptors. Endocrine disruptors (EDs) are a great concern in the food safety field since several molecules naturally present, intentionally, or unintentionally added in food may act as endocrine-disrupting compounds (EDCs). The U.S. Environmental Protection Agency (EPA) defined an endocrine-disrupting compound as "*an exogenous agent that interferes with synthesis, secretion, transport, metabolism, binding action, or elimination of natural blood-borne hormones that are present in the body and are responsible for homeostasis, reproduction, and developmental process*". As it is evident by the definition, EDCs are molecules that can interfere with the endocrine system at different levels. There are generally two ways by which these molecules may affect the hormone system. Initially, endocrine disrupting compounds were thought to exert their actions primarily mimicking the effect of natural hormones through their interaction with nuclear receptors (NRs). However, the mechanisms of action (MoA) of EDCs are much broader than originally recognized. EDs may act through an indirect mechanism by interacting with membrane proteins, nonsteroid receptors, enzymatic proteins, or via other mechanisms which are all involved or control some aspects of the endocrine and reproductive systems (Diamanti-Kandarakis et al., 2009). The interference with the endocrine systems leads to the development of several harmful effects for humans, such as malformations, reproductive system problems, an increase in cancer risk, and disturbances in the immune and nervous system function. In the 1970s, it was reported the first evidence about the harmful effect of EDCs

for humans when the oestrogenic drug diethylstilbestrol (DES), administrated during the 1940s, was associated with women and male's reproductive tract problems. As new evidence was obtained about the potential problem of endocrine disruptors, the European Commission, in 1999, adopted a strategy following the precautionary principle (COM(1999)706) to respond quickly and effectively to this concern and to set out the Commission's obligation to protect humans and the environment. Endocrine disruptors are an emerging problem for public health. In fact, in REACH (Registration, Evaluation, Authorisation, and Restriction of Chemicals), a European Union regulation concerning the risks posed by chemical substances (EC 1907/2006), ECDs are considered as substances of very high concern (SVHCs) similarly to the regulatory concern posed to cancerogenic, mutagenic and toxic molecules. Today, several existing legislations are present within the EU, and some legislations exist to govern specific chemicals, such as pesticides and biocides. In general, before entering the marketing system, a substance undergoes three different evaluations:

- *hazard identification*: the potential hazard of chemicals to cause harm is assessed using different tests according to the OECD (Organization for Economic Co-operation and Development);
- *risk assessment*: the probability that a molecule may cause a harmful effect is evaluated based on the results obtained by the hazard identification and the environmental levels;
- *risk management*: according to the previous steps, the use of some chemicals can be restricted or reviewed.

Based on this workflow, after the release of the COM(1999)706 legislation, three other updates were presented as an implementation of the strategy (2001 (COM(2001)262), 2004 (SEC(2004)1372) and 2007 (SEC (2007) 1635)). Following the precaution principle, all these legislations update the document adding new substances as suspected molecules capable of interfering with the hormone systems of humans and wildlife. However, the increasing number of molecules released every year combined with the different steps needed to evaluate a single substance make the endocrine-disrupting evaluation very challenging, at least in a limited time and at low costs. Moreover, in the past decades, regulatory efforts and policies have been proved inefficient to minimize and decrease human exposure to EDCs. In addition, a lot of chemicals are waiting for evaluating their endocrine-disrupting properties and some weaknesses exist in testing

approaches. To this end, the European Union should consider expanded and alternative test methods to conclusively identify EDCs (Kassotis et al., 2020).

1.3 Nuclear Receptors

Nuclear receptors (NRs) are an important family of transcription factors. In multicellular organisms, they play an important role in several biological functions, such as they regulate reproduction, development, and nutrient utilization (Bookout et al., 2006). After the first steroid receptors were cloned in the mid-1980s, a great number of nuclear receptors have been identified through the screening of cDNA libraries mainly due to the extensive amino acid and gene sequence similarity (Chawla et al., 2001). The nuclear receptors superfamily is composed of 48 different members in humans. Most NRs are regulated by endogenous small lipophilic ligands such as steroids, retinoids, and phospholipids, which can cross the plasma membrane and bind their targets. However, the class also includes "orphan" receptors for which a natural ligand has not yet been discovered. Once activated, nuclear receptors directly bind the promoter genes that regulate, and activate the gene transcription.

Nuclear receptors share a common structure (**Figure 1**) comprising a variable amino-terminal domain, which contains the activation function region (AF1, also known as A/B domain), a core DNA-binding domain (DBD), a short hinge region responsible for protein flexibility (D domain), and a ligand-binding domain (LBD). Additionally, other nuclear receptors consist of a short C-terminal region, known as the F domain, of unknown function (Chawla et al., 2001; Sever and Glass, 2013).

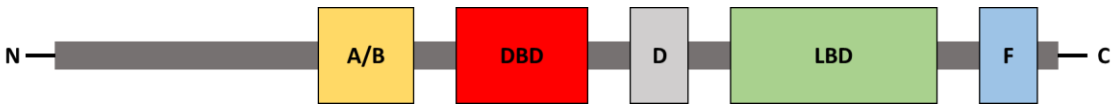


Figure 1. The general gene structure of nuclear receptors.

Acting as monomers, homodimers, or heterodimers, nuclear receptors recognize specific DNA sequences named hormone response elements (HREs), which can also be located far away from the transcriptional start site, such as in the enhancer regions. Some NRs are localized in the cytoplasm, and once bound by the ligand, migrate inside the nucleus and activate gene transcription both as monomers and dimers. Others reside in the nucleus bound to the HREs, and the gene transcription is inhibited by the interaction with corepressors. After the ligand binding, corepressors are replaced by coactivators which

facilitate the activation of target genes (Glass and Rosenfeld, 2000; Sever and Glass, 2013; Weikum et al., 2018).

Ligands bind nuclear receptors within the ligand-binding pocket (LBP) of the LBD and promote conformation changes allowing the binding of coactivators and supporting the active conformation. Other ligands can act as antagonistic molecules inducing the receptor to constitutively be in the inactive conformation.

A peculiar characteristic of nuclear receptors is their ability to regulate different genes in different cell types, controlling a wide variety of processes. However, because they bind to small molecules, they often represent an easy target for several exogenous molecules, which may act as agonists or antagonists. Following this interaction, their dysregulation may promote numerous diseases, including cancer, diabetes, and infertility. Two nuclear receptors, estrogen and androgen receptors (ER and AR, respectively), are often the target of endocrine disruptor compounds (EDCs). The binding of these molecules to ER and AR has been associated with increased cancer risk in males and female.

1.3.1 Estrogen receptor

Estrogen receptor belongs to the family of nuclear receptors. To date, two different isoforms were identified and cloned: ERα (Greene et al., 1986) and ERβ (Mosselman et al., 1996). They are synthesized by two different genes and share a very similar structure and function. ERα and ERβ have a high homology sequence in the DNA-binding domain region (96%), but the ligand-binding domain (LBB) shows only a 53% homology (Journé et al., 2008). Despite the 53% sequence identity of LBD, ERα and ERβ exhibit some differences in ligand binding specificity as for some anti-estrogens and phytoestrogens. On the contrary, the natural ligand of ER, 17β-estradiol, binds the two isoforms with a quite similar binding affinity (Kuiper et al., 1998, 1997). The two isoforms are expressed in different parts of the human body and are mainly found in the breast, brain, cardiovascular system, urogenital tract, and bone. However, ERα and ERβ show their peculiarities in their abundance in the body and the tissues. For example, while ERα is mainly found in the liver, beta isoform is more abundant in the colon. A different localization is also highlighted within the ovary, where ERα is largely present in the thecal and interstitial cells, and ERβ in the granulosa cells (Pearce and Jordan, 2004). Thus, ERα and ERβ have different biological actions due to their tissue-specificity.

In absence of hormone, the estrogen receptor is in inactive conformation bound to the heat shock protein 90 (Hsp90), which prevents the degradation of the receptor. After the

binding of 17β-estradiol or another agonist to the LBD, the receptor forms homodimer or heterodimers, enters into the nucleus, and interacts with DNA regulatory sequences (estrogen response elements, ERE) located in the promoter regions of target genes.

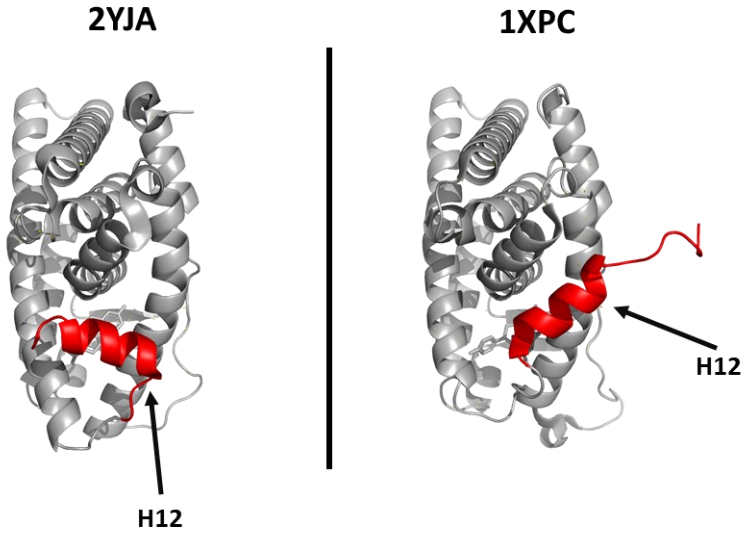


Figure 2. On the left, the structure of LBD in the agonist conformation, with the H12 close to the body of the receptor. On the right, the three-dimensional structure of the receptor in the unbound or antagonist conformation.

Ligand binding influences the conformation of a short helix, also known as AF2 (activation function 2) or helix 12 (H12) located at the carboxy-terminal end of the LBD (**Figure 2**). Upon the agonist binding, the helix 12 is located close to the structure of the LBD resulting in the presentation of a consensus sequence LxxLL to coactivator proteins. On the other side, in the unbound state or when an NR antagonist binds the LBP, the helix 12 is completely dissociated from the body of the LBD and this prevents coactivator interaction.

The estrogen receptors and their ligands have several roles in the human body and regulate a lot of functions, such as the growth, development, and physiology of the human reproductive system, but they also influence neuroendocrine, skeletal, adipose, and cardiovascular systems. Since the estrogen receptors are also involved in cell proliferation, a lot of evidence has reported that the expression of ERα and ERβ is closely associated with breast cancer and with the development of tumors (Jameera Begam et al., 2017). Moreover, it is found that almost 70% of breast cancers are estrogen receptor α (ERα) positive (Xiong et al., 2016). Therefore, dysregulation in the function of these two isoforms may have adverse effects. For this reason, endocrine disrupting compounds (EDCs) are a very concern. In fact, after the binding with the LBD of ERα and ERβ, they

can impact the estrogen signaling by inappropriate activation of the receptors. Pharmaceuticals, phytoestrogens, pesticides, and industrial chemicals are some examples of such direct-acting of EDCs (Shanle and Xu, 2011).

1.3.2 Androgen receptor

As estrogen receptors, the androgen receptor (AR) belongs to the steroid nuclear receptors. It was firstly cloned in 1988 (Lubahn et al., 1988). The AR gene was localized on the human Xq 11-12 chromosome and consists of eight exons. Androgen receptor shares a common three-dimensional structure to the other steroid hormone nuclear receptors. However, the C-terminal of the protein contains a β-sheet that is not found in the other NRs. The presence of this region may restrict the mobility of helix 12 differently from what can be observed in the estrogen receptor (Bohl et al., 2007). In fact, to date, there is not a crystal structure of a wild-type androgen receptor in the open conformation. AR is localized to the cytoplasm of the cell in the absence of a ligand. After ligand binding, it translocates to the nucleus and plays its physiological function by activating genes transcription. Recently, an X-ray crystal structure of AR LBD homodimer has been determined (Nadal et al., 2017), suggesting that upon the ligand interaction, the receptor undergoes homodimerization.

Testosterone (TES) and dihydrotestosterone (DHT) are the two natural ligands of the receptor and bind to the canonical ligand-binding pocket. However, the transcriptional activity of the receptor may be also influenced by protein-protein interaction. Androgen receptor consists of a binding function 3 (BF3) site on the surface of the AR. By recruiting some coregulators, such as FKBP52 (De Leon et al., 2011) and Bag-1L (Jehle et al., 2014), BF3 may induce conformational changes in the adjacent AF2 domain (Grosdidier et al., 2012) influencing genes transcription.

Androgen receptor is expressed in a wide variety of tissues and plays an important role in bone, muscle, prostate, adipose tissue, cardiovascular, immune, neural, hemopoietic systems, and both male and female reproductive development and function (Hu et al., 2020). Prostate cancer (PCa) affects approximately 1 in 8 men. The AR signalling plays a significant role in many early and late-stage PCa tumors (Jillson et al., 2021). Malignant cells express high doses of prostate-specific antigen (PSA) (Kim and Coetzee, 2004). For that reason, the PSA test is the most validated test for the diagnosis of prostate cancer (PCa) (Logozzi et al., 2017). Endocrine disruptor compounds (EDCs) can influence prostate cancers by binding to the androgen receptor, acting both as agonists and

antagonists. It has been reported that some EDCs may stimulate PSA expression by increasing the progression of prostate cancers; others may contribution to PCa growth. Moreover, some endocrine-disrupting compounds may also affect therapeutic response in a subset of tumors that harbour AR ligand-binding domain mutations (Hess-Wilson and Knudsen, 2006).

1.4 Molecular Docking & Consensus Scoring

Toxicology risk assessments are essential to ensure the safety of consumers and the environment. In section 1.2, it has been emphasised the need for new methodologies to perform hazard identification, e.g., the probability that a compound may determine a harmful effect on humans and the environment. A large amount of chemicals remains unexplored. *In vitro* and *in vivo* analyses are often time-consuming and require a terrific amount of money at least in light of the increasing number of molecules that require a risk evaluation. *In silico* methodologies have proven to be a very useful approach in the context of medicinal chemistry and, in the last ten years, a lot of papers have shown their utility in risk evaluation. With these methods, it is possible to filter a battery of compounds to predict their potential activity towards a protein target. The adverse outcome pathway (AOP) is defined as the molecular and cellular events required to produce a toxic effect after the exposure of an individual to a chemical. In most cases, at the beginning of AOP, it is possible to identify a molecular initiating event (MIE) which often involves the interaction between a compound and a macromolecule (proteins, nucleic acids, complex carbohydrates, lipids, etc.) (Allen et al., 2014).

1.4.1 Molecular Docking

Molecular docking is a computational strategy that attempts to predict the interaction between two constituent molecules: protein-protein, protein-ligand, protein-nucleic acid, ligand-nucleic acid, and so on. Since endocrine-disrupting compounds often interact with nuclear receptors establishing a protein-ligand complex, this latter case has been discussed in more detail.

The “lock and key” concept was first introduced by Emil Fischer in 1894 to suggest protein-ligand interaction, where the “lock” describes the protein/enzyme, and the “key” describes the ligand. As for a lock and key mechanism, the ligand (key) should fit appropriately into the hole (binding pocket) of the protein (lock). Ligands that are too small or large, or with chemical features incorrectly positioned will not properly fit into

the protein (Tripathi and Bankaitis, 2017). By analogy to the lock and key concept, molecular docking attempts to find the best complementary between the two rigid bodies. However, this model is too simplistic because both the protein and the ligand are not a rigid body and protein/ligand flexibility should be considered, in accordance with the induced-fit concept. Molecular docking is mainly composed of two different steps. In the first step, it predicts the most favourable protein-ligand binding mode using molecular docking algorithms. In the second step, it ranks a family of ligands using the values obtained from the scoring function implemented in the docking software. The scoring functions are mathematical functions used to provide a value that predicts how tightly the two molecules interact (i.e., a kind of binding affinity) (Elokely and Doerksen, 2013). Each molecular docking software has its own algorithms and scoring functions. However, considering both the protein and the ligand as two flexible bodies increases computational requirements, since the software should ideally search for the most populated alternatives of protein-ligand solution. In fact, while considering the flexibility of small molecules is quite simple from a computational point of view, addressing the full protein flexibility is a challenging task. For that reason, most of the docking software uses a compromise between cost and benefit. Several docking algorithms exist, and they may consider the protein as a rigid body or as a semi-flexible entity. In this latter case, the docking software may allow the flexibility of some side-chain or certain protein domain. Alternatively, multiple three-dimensional protein structures may be used to account for protein flexibility that represents an ensemble of rigid protein structures (Elokely and Doerksen, 2013).

1.4.2 Consensus Scoring

Molecular docking mainly relies on an algorithm and a scoring function, and each docking software is a combination of the two elements. At the beginning of every computational protocol, one faces the decision of which software package should use. There is not the best docking software since each of them has its performances that are strictly correlated to the complexes used during the training phase. Thus, a best practice may be to choose a docking package that is trained with protein-ligand complexes that are close to the complex under investigation. However, when it is needed to test very different ligands and/or different proteins, the question about which docking software choose still remains. With the increase of computational performance, molecular docking calculations are computationally efficient and very fast even though large datasets are

tested. It has been found that combining the results obtained from different molecular docking to obtain a final rank or score leads to an increase in the result reliability also decreasing the number of false positives (Bissantz et al., 2000; Ericksen et al., 2017; Oda et al., 2006; Wang and Wang, 2001). For that reason, when multiple protein-ligand complexes need to be investigated, a consensus docking approach can be a good practice.

1.5 Molecular Dynamic

Molecular dynamics (MD) simulation is a useful method to investigate the time-dependent behaviour of a molecular system. It also allows studying the fluctuation and the conformational changes of a molecule. The molecular dynamic method was first introduced in the late 1950s when Alder and Wainwright (Alder and Wainwright, 1957) used the integration of Newton's second law of motion to study the interaction of two hard spheres for a total simulation of 9.2ps (picoseconds). After that, MD simulations were used for different systems such as small solutes, peptides or to study the folding of small proteins. However, it was only in 1977 that MD was used to study the behaviour of a folded protein, the bovine pancreatic trypsin inhibitor, that was simulated for 8.8ps (McCammon et al., 1977). Classical molecular dynamics is based on Newton' second law, i.e., the equation of motion:

$$F_i = m_i * a_i$$

where F_i is the force exerted on the particle i, m_i is the mass of particle i, and a_i is the acceleration of particle i. By solving Newton's equation, it is possible to generate a trajectory that describes the positions, the velocities, and the acceleration of each atom during the time. In fact, by deriving the equation of motion it is also possible to determine the acceleration:

$$F_i = -\frac{\partial E_{pot}}{\partial x_i}$$

where E_{pot} is the potential energy and x_i is the *x, y, z* coordinates of the particle i. The potential energy is separated into bonded (E_{bonded}) and non-bonded ($E_{nonbonded}$) terms:

$$E_{pot} = E_{bonded} + E_{nonbonded}$$

Moreover, the bonded energy is given by the stretching, bending, and torsional energy terms:

$$E_{bonded} = E_{str} + E_{bend} + E_{tor}$$

and the non-bonded energy is given by the electric and Van der Wall contributions:

$$E_{nonbonded} = E_{elec} + E_{vdw}$$

Force field is the collection of equations (Guvench and MacKerell, 2008) and the associated parameters that are used to describe the potential energy of the tested system. The parameters of the force field are often derived from data obtained by higher-level calculations and/or experiments. In this way, the parametrization of most biomolecules, such as proteins, DNA, lipids, and sugars can be performed only once (Mortier et al., 2015). Several force fields should be chosen based on the system under investigation. For example, specific force fields are developed for proteins, nucleic acid, carbohydrates, etc. However, the detailed information about force fields is beyond the scope of the present thesis.

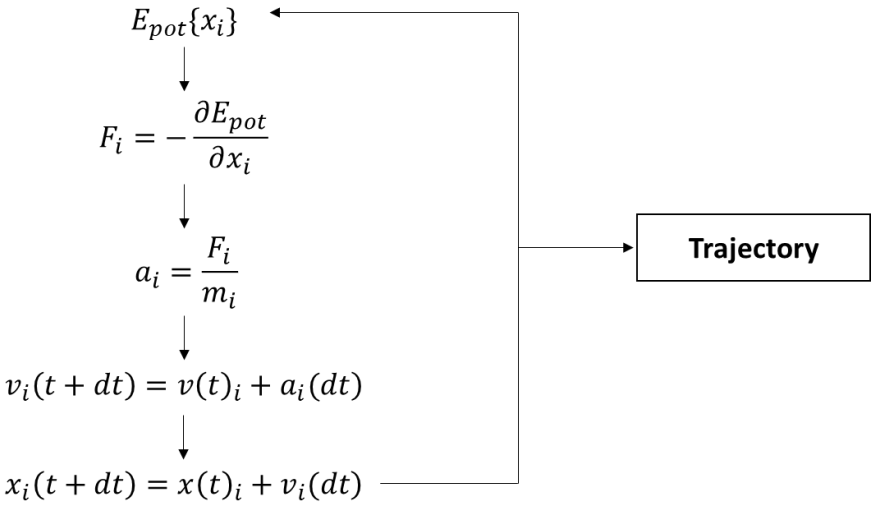


Figure 3. The formulas used to calculate the entire trajectory over the time simulation.

Once the potential energy has been calculated using the proper force field, it is possible to calculate the acceleration since at the beginning of the simulation the coordinates of the system are also known. Using the acceleration, both the velocity and the position at the time t can be derived. By solving the equation of motion over time, it is possible to obtain the full trajectory (**Figure 3**). The resulting MD trajectory can be analysed to extract important information about the system. However, a lot of biological processes occur in milliseconds/seconds time scale. While the atom vibrations occur in femtoseconds/picoseconds (Fichou et al., 2015), some processes, such as substrate binding, hydrolysis, ion transport, etc., occur in milliseconds/seconds (Szöllősi et al.,

2018). Simulate a system for this long time is computationally challenging, especially when it is simulated a system composed of millions of atoms. Nowadays, computational methods are undergoing to rapid development with the increase in computer performances, especially with the advent of high-performance computers (HPCs). Combining with the improvement in MD methodology, it is now possible to simulate biologically relevant processes (Wolf et al., 2020). The use of MD simulations is now an essential methodology to study relevant events at the microscopic level which helps to rationalize some experimental data. The important achievements reached by computational methods have also a profound impact on hazard identification. For example, the application of MD simulation may shed light on the effect of endocrine-disrupting compounds over the simulation time.

1.6 High-Performance Computer

The utilization of high-performance computing (HPC) tooks the computational methods to the next level. High-Performance Computers (HPC) process data and perform complex calculations at high speed. For example, while a laptop PC can perform around 3 billion calculations per second with a 3 GHz processor, HPC solutions can perform quadrillions of calculations per second. High-performance computer architecture is composed of computer servers (hundreds or thousands) that are networked together into a computer cluster, where each computer server is called “node”. Software programs and algorithms, such as molecular dynamic calculation, are run in parallel on the nodes of the clusters, significantly increasing the processing speed. Another important component of a high-performance computer center is data storage, which is a physical device used to capture the output during the calculation.

Two different types of calculation can significantly benefit from the use of the high-performance computer:

1) Calculations that are composed of many similar jobs with different parameters or data sets have good performances on HPC clusters. The jobs are submitted together and the cluster itself manages the entire workflow.
2) Complex calculations (as molecular dynamic) are split into smaller sub-jobs and each of them is run on different nodes. This type of program is called parallel because it uses many cores that collaborate to solve the same problem. Breaking a job in sub-jobs could increase speed. A parallel job could be like building a house: the more workers are, the faster is completed the job. However, some parts

of the calculation are highly parallel instead other steps may depend on the others. For instance, some calculations cannot be finished until one step is done. Thus, there are some limitations in the speed-up of parallel jobs because there are some “slow steps”. This effect is called Amdahl’s law and it gives an estimation about the speed-up is possible to obtain from a given calculation. In fact, in some cases, although more resources are requested, the calculation does not run faster.

Calculations may also benefit from graphical processing units (GPUs) on HPC. GPU hardware contains thousands of cores which allows parallel jobs, increasing the speed of the calculation 19 times compared to the single CPU. Thus, the advent of high-performance computer centers has drastically increased the velocity of computational calculation, significantly boosting their use.

References

(CONTAM), E.P. on C. in the F.C., Schrenk, D., Bignami, M., Bodin, L., Chipman, J.K., del Mazo, J., Grasl-Kraupp, B., Hogstrand, C., Hoogenboom, L. (Ron), Leblanc, J.-C., Nebbia, C.S., Nielsen, E., Ntzani, E., Petersen, A., Sand, S., Vleminckx, C., Wallace, H., Benford, D., Brimer, L., Mancini, F.R., Metzler, M., Viviani, B., Altieri, A., Arcella, D., Steinkellner, H., Schwerdtle, T., 2019. Evaluation of the health risks related to the presence of cyanogenic glycosides in foods other than raw apricot kernels. EFSA J. 17, e05662. https://doi.org/https://doi.org/10.2903/j.efsa.2019.5662

Alder, B.J., Wainwright, T.E., 1957. Phase Transition for a Hard Sphere System. J. Chem. Phys. 27, 1208–1209. https://doi.org/10.1063/1.1743957

Allen, T.E.H., Goodman, J.M., Gutsell, S., Russell, P.J., 2014. Defining Molecular Initiating Events in the Adverse Outcome Pathway Framework for Risk Assessment. Chem. Res. Toxicol. 27, 2100–2112. https://doi.org/10.1021/tx500345j

Arvanitoyannis, I.S., Bosnea, L., 2004. Migration of Substances from Food Packaging Materials to Foods. Crit. Rev. Food Sci. Nutr. 44, 63–76. https://doi.org/10.1080/10408690490424621

Barnes, K., Sinclair, R., Watson, D., 2006. Chemical Migration and Food Contact Materials, 1st ed. Elsevier.

Bissantz, C., Folkers, G., Rognan, D., 2000. Protein-Based Virtual Screening of Chemical Databases . 1 . Evaluation of Different Docking / Scoring Combinations. J. Med. Chem. 43, 4759–4767. https://doi.org/10.1021/jm001044l

Bohl, C.E., Wu, Z., Miller, D.D., Bell, C.E., Dalton, J.T., 2007. Crystal structure of the T877A human androgen receptor ligand-binding domain complexed to cyproterone acetate provides insight for ligand-induced conformational changes and structure-based drug design. J. Biol. Chem. 282, 13648–13655. https://doi.org/10.1074/jbc.M611711200

Bookout, A.L., Jeong, Y., Downes, M., Yu, R.T., Evans, R.M., Mangelsdorf, D.J., 2006. Anatomical profiling of nuclear receptor expression reveals a hierarchical transcriptional network. Cell 126, 789–799. https://doi.org/10.1016/j.cell.2006.06.049

Chawla, A., Joyce, R.J., Ronald, E.M., Mangelsdorf, D.J., 2001. Nuclear Receptors and Lipid Physiology: Opening the X-Files. Science (80-.). 294, 1866–1870. https://doi.org/10.1126/science.294.5548.1866

Crews, C., 2009. Consumer exposure to furan from heat – processed food and kitchen air. EFSA Support. Publ. 6, 30E. https://doi.org/https://doi.org/10.2903/sp.efsa.2009.EN-30

De Leon, J.T., Iwai, A., Feau, C., Garcia, Y., Balsiger, H.A., Storer, C.L., Suro, R.M., Garza, K.M., Lee, S., Sang Kim, Y., Chen, Y., Ning, Y.-M., Riggs, D.L., Fletterick, R.J., Guy, R.K., Trepel, J.B., Neckers, L.M., Cox, M.B., 2011. Targeting the regulation of androgen receptor signaling by the heat shock protein 90 cochaperone FKBP52 in prostate cancer cells. Proc. Natl. Acad. Sci. 108, 11878 LP – 11883. https://doi.org/10.1073/pnas.1105160108

Diamanti-Kandarakis, E., Bourguignon, J.-P., Giudice, L.C., Hauser, R., Prins, G.S., Soto, A.M., Zoeller, R.T., Gore, A.C., 2009. Endocrine-disrupting chemicals: an Endocrine Society scientific statement. Endocr. Rev. 30, 293–342. https://doi.org/10.1210/er.2009-0002

Elokely, K.M., Doerksen, R.J., 2013. Docking Challenge: Protein Sampling and Molecular Docking Performance. J. Chem. Inf. Model. 53, 1934–1945. https://doi.org/10.1021/ci400040d

Ericksen, S.S., Wu, H., Zhang, H., Michael, L.A., Newton, M.A., Hoffmann, F.M., Wildman, S.A., 2017. Machine Learning Consensus Scoring Improves Performance Across Targets in Structure-Based Virtual Screening. J. Chem. Inf. Model. 57, 1579–1590. https://doi.org/10.1021/acs.jcim.7b00153

Fichou, Y., Schirò, G., Gallat, F.-X., Laguri, C., Moulin, M., Combet, J., Zamponi, M., Härtlein, M., Picart, C., Mossou, E., Lortat-Jacob, H., Colletier, J.-P., Tobias, D.J., Weik, M., 2015. Hydration water mobility is enhanced around tau amyloid fibers. Proc. Natl. Acad. Sci. 112, 6365 LP – 6370. https://doi.org/10.1073/pnas.1422824112

Fromberg, A., Fagt, S., Granby, K., 2009. Furan in heat processed food products including home cooked food products and ready-to-eat products. EFSA Support. Publ. 6, 1E. https://doi.org/https://doi.org/10.2903/sp.efsa.2009.EN-1

Glass, C.K., Rosenfeld, M.G., 2000. The coregulator exchange in transcriptional functions of nuclear receptors. Genes Dev. 14, 121–141.

Greene, G.L., Gilna, P., Waterfield, M., Baker, A., Hort, Y., Shine, J., 1986. Sequence and expression of human estrogen receptor complementary DNA. Science 231, 1150–1154. https://doi.org/10.1126/science.3753802

Grosdidier, S., Carbó, L.R., Buzón, V., Brooke, G., Nguyen, P., Baxter, J.D., Bevan, C., Webb, P., Estébanez-Perpiñá, E., Fernández-Recio, J., 2012. Allosteric Conversation in the Androgen Receptor Ligand-Binding Domain Surfaces. Mol. Endocrinol. 26, 1078–1090. https://doi.org/10.1210/me.2011-1281

Guvench, O., MacKerell, A.D.J., 2008. Comparison of protein force fields for molecular dynamics simulations. Methods Mol. Biol. 443, 63–88. https://doi.org/10.1007/978-1-59745-177-2_4

Hess-Wilson, J.K., Knudsen, K.E., 2006. Endocrine disrupting compounds and prostate cancer. Cancer Lett. 241, 1–12. https://doi.org/https://doi.org/10.1016/j.canlet.2005.10.006

Hu, C., Fang, D., Xu, H., Wang, Q., Xia, H., 2020. The androgen receptor expression and association with patient's survival in different cancers. Genomics 112, 1926–1940. https://doi.org/https://doi.org/10.1016/j.ygeno.2019.11.005

Jameera Begam, A., Jubie, S., Nanjan, M.J., 2017. Estrogen receptor agonists/antagonists in breast cancer therapy: A critical review. Bioorg. Chem. 71, 257–274. https://doi.org/https://doi.org/10.1016/j.bioorg.2017.02.011

Jehle, K., Cato, L., Neeb, A., Muhle-Goll, C., Jung, N., Smith, E.W., Buzon, V., Carbó, L.R., Estébanez-Perpiñá, E., Schmitz, K., Fruk, L., Luy, B., Chen, Y., Cox, M.B., Bräse, S., Brown, M., Cato, A.C.B., 2014. Coregulator Control of Androgen Receptor Action by a Novel Nuclear Receptor-binding Motif * . J. Biol. Chem. 289, 8839–8851. https://doi.org/10.1074/jbc.M113.534859

Jillson, L.K., Yette, G.A., Laajala, T.D., Tilley, W.D., Costello, J.C., Cramer, S.D., 2021. Androgen Receptor Signaling in Prostate Cancer Genomic Subtypes. Cancers . https://doi.org/10.3390/cancers13133272

Journé, F., Body, J.-J., Leclercq, G., Laurent, G., 2008. Hormone therapy for breast

cancer, with an emphasis on the pure antiestrogen fulvestrant: mode of action, antitumor efficacy and effects on bone health. Expert Opin. Drug Saf. 7, 241–258. https://doi.org/10.1517/14740338.7.3.241

Kassotis, C.D., Vandenberg, L.N., Demeneix, B.A., Porta, M., Slama, R., Trasande, L., 2020. Endocrine-disrupting chemicals: economic, regulatory, and policy implications. lancet. Diabetes Endocrinol. 8, 719–730. https://doi.org/10.1016/S2213-8587(20)30128-5

Kim, J., Coetzee, G.A., 2004. Prostate specific antigen gene regulation by androgen receptor. J. Cell. Biochem. 93, 233–241. https://doi.org/10.1002/jcb.20228

Kuiper, G.G.J.M., Carlsson, B., Grandien, K., Enmark, E., Häggblad, J., Nilsson, S., Gustafsson, J.-A., 1997. Comparison of the Ligand Binding Specificity and Transcript Tissue Distribution of Estrogen Receptors α and β. Endocrinology 138, 863–870. https://doi.org/10.1210/endo.138.3.4979

Kuiper, G.G.J.M., Lemmen, J.G., Carlsson, B., Corton, J.C., Safe, S.H., van der Saag, P.T., van der Burg, B., Gustafsson, J.-A., 1998. Interaction of Estrogenic Chemicals and Phytoestrogens with Estrogen Receptor β. Endocrinology 139, 4252–4263. https://doi.org/10.1210/endo.139.10.6216

Logozzi, M., Angelini, D.F., Iessi, E., Mizzoni, D., Di Raimo, R., Federici, C., Lugini, L., Borsellino, G., Gentilucci, A., Pierella, F., Marzio, V., Sciarra, A., Battistini, L., Fais, S., 2017. Increased PSA expression on prostate cancer exosomes in in vitro condition and in cancer patients. Cancer Lett. 403, 318–329. https://doi.org/10.1016/j.canlet.2017.06.036

Lubahn, D.B., Joseph, D.R., Sullivan, P.M., Willard, H.F., French, F.S., Wilson, E.M., 1988. Cloning of Human Androgen Receptor Complementary DNA and Localization to the X Chromosome. Science (80-.). 240, 327–330. https://doi.org/10.1126/science.3353727

McCammon, J.A., Gelin, B.R., Karplus, M., 1977. Dynamics of folded proteins. Nature 267, 585–590. https://doi.org/10.1038/267585a0

Mortier, J., Rakers, C., Bermudez, M., Murgueitio, M.S., Riniker, S., Wolber, G., 2015. The impact of molecular dynamics on drug design: applications for the

characterization of ligand–macromolecule complexes. Drug Discov. Today 20, 686–702. https://doi.org/https://doi.org/10.1016/j.drudis.2015.01.003

Mosselman, S., Polman, J., Dijkema, R., 1996. ER beta: identification and characterization of a novel human estrogen receptor. FEBS Lett. 392, 49–53. https://doi.org/10.1016/0014-5793(96)00782-x

Muncke, J., Andersson, A.-M., Backhaus, T., Boucher, J.M., Carney Almroth, B., Castillo Castillo, A., Chevrier, J., Demeneix, B.A., Emmanuel, J.A., Fini, J.-B., Gee, D., Geueke, B., Groh, K., Heindel, J.J., Houlihan, J., Kassotis, C.D., Kwiatkowski, C.F., Lefferts, L.Y., Maffini, M. V, Martin, O. V, Myers, J.P., Nadal, A., Nerin, C., Pelch, K.E., Fernández, S.R., Sargis, R.M., Soto, A.M., Trasande, L., Vandenberg, L.N., Wagner, M., Wu, C., Zoeller, R.T., Scheringer, M., 2020. Impacts of food contact chemicals on human health: a consensus statement. Environ. Heal. 19, 25. https://doi.org/10.1186/s12940-020-0572-5

Nadal, M., Prekovic, S., Gallastegui, N., Helsen, C., Abella, M., Zielinska, K., Gay, M., Vilaseca, M., Taulès, M., Houtsmuller, A.B., van Royen, M.E., Claessens, F., Fuentes-Prior, P., Estébanez-Perpiñá, E., 2017. Structure of the homodimeric androgen receptor ligand-binding domain. Nat. Commun. 8, 14388. https://doi.org/10.1038/ncomms14388

Oda, A., Tsuchida, K., Takakura, T., Yamaotsu, N., Hirono, S., 2006. Comparison of consensus scoring strategies for evaluating computational models of protein-ligand complexes. J. Chem. Inf. Model. 46, 380–391. https://doi.org/10.1021/ci050283k

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

Sever, R., Glass, C.K., 2013. Signaling by Nuclear Receptors. Cold Spring Harb Perspect Biol 1–5. https://doi.org/10.1101/cshperspect.a016709

Shanle, E.K., Xu, W., 2011. Endocrine disrupting chemicals targeting estrogen receptor signaling: identification and mechanisms of action. Chem. Res. Toxicol. 24, 6–19. https://doi.org/10.1021/tx100231n

Smith, R.L., Cohen, S.M., Fukushima, S., Gooderham, N.J., Hecht, S.S., Guengerich,

F.P., Rietjens, I.M.C.M., Bastaki, M., Harman, C.L., McGowen, M.M., Taylor, S. V, 2018. The safety evaluation of food flavouring substances: the role of metabolic studies. Toxicol. Res. (Camb). 7, 618–646. https://doi.org/10.1039/c7tx00254h

Stasinos, S., Nasopoulou, C., Tsikrika, C., Zabetakis, I., 2014. The Bioaccumulation and Physiological Effects of Heavy Metals in Carrots, Onions, and Potatoes and Dietary Implications for Cr and Ni: A Review. J. Food Sci. 79, R765–R780. https://doi.org/https://doi.org/10.1111/1750-3841.12433

Szöllősi, D., Rose-Sperling, D., Hellmich, U.A., Stockner, T., 2018. Comparison of mechanistic transport cycle models of ABC exporters. Biochim. Biophys. Acta - Biomembr. 1860, 818–832. https://doi.org/https://doi.org/10.1016/j.bbamem.2017.10.028

Tripathi, A., Bankaitis, V.A., 2017. Molecular Docking: From Lock and Key to Combination Lock. J. Mol. Med. Clin. Appl. 2, 10.16966/2575-0305.106. https://doi.org/10.16966/2575-0305.106

Van Dolah, F.M., 2000. Marine algal toxins: origins, health effects, and their increased occurrence. Environ. Health Perspect. 108 Suppl, 133–141. https://doi.org/10.1289/ehp.00108s1133

Wang, R., Wang, S., 2001. How does consensus scoring work for virtual library screening? An idealized computer experiment. J. Chem. Inf. Comput. Sci. 41, 1422–1426. https://doi.org/10.1021/ci010025x

Weikum, E.R., Liu, X., Ortlund, E.A., 2018. The nuclear receptor superfamily: A structural perspective. Protein Sci. 27, 1876–1892. https://doi.org/10.1002/pro.3496

Wolf, S., Lickert, B., Bray, S., Stock, G., 2020. Multisecond ligand dissociation dynamics from atomistic simulations. Nat. Commun. 11, 2918. https://doi.org/10.1038/s41467-020-16655-1

Xiong, R., Patel, H.K., Gutgesell, L.M., Zhao, J., Delgado-Rivera, L., Pham, T.N.D., Zhao, H., Carlson, K., Martin, T., Katzenellenbogen, J.A., Moore, T.W., Tonetti, D.A., Thatcher, G.R.J., 2016. Selective Human Estrogen Receptor Partial Agonists (ShERPAs) for Tamoxifen-Resistant Breast Cancer. J. Med. Chem. 59, 219–237. https://doi.org/10.1021/acs.jmedchem.5b01276

CHAPTER 2

Database System

2.1 Introduction

The increasing amounts of data coming from food risk assessment tests prompt the need for efficient databases where collect multiple data sources. By using databases, food safety authorities can have rapid access to essential information for deciding about food contact chemicals safety to prevent and manage disease outbreaks. Several food safety databases exist that mainly collect information about the experimental results of molecules that have been already tested (JECFA Database), the contaminant levels of molecules in food (GEMS Food contamination database) or the results about food consumption levels. These databases are very useful for risk assessment. However, there is the necessity to collect different types of data to accelerate the process of hazard identification since a lot of molecules have not yet been tested. Considering this scenario, the use of efficient databases has become vital for maintaining and monitoring food safety.

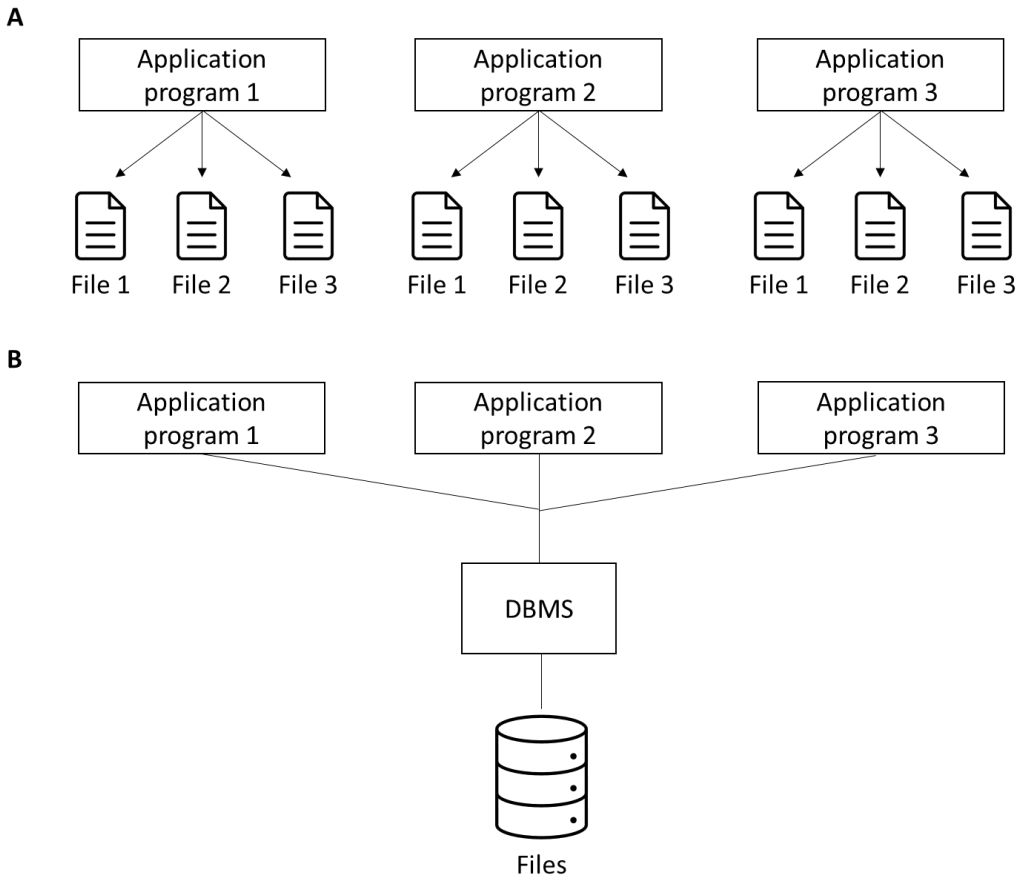


Figure 1. Schematical representation of how to access data (A) before and (B) after Database System development.

Before the development of the database system concept, all information was stored in simple linear files, where each file is independent of other files (**Figure 1**). Each file can be seen as a collection of records in which the elementary data (attributes and fields) are memorized. Without the database system application, sharing data among different application programs can be made using shared files. This approach has different disadvantages:

- Data file formats can be incompatible among different programs, and that lead to (a) programs should adapt to different convention over the time, and (b) data sharing may be very challenging.
- Any changes in the data structure of the file will require the modification of all the programs that use these data.
- Information may be redundant if the data are not shared.

For these reasons, there is the necessity to find tools that simplify the tasks of managing the data and extracting useful information.

2.2 Database system

The term "database system" is often used as a synonym of the database management system (DBMS). But actually, the term means the combination of the DBMS software and the Database (**Figure 2**).

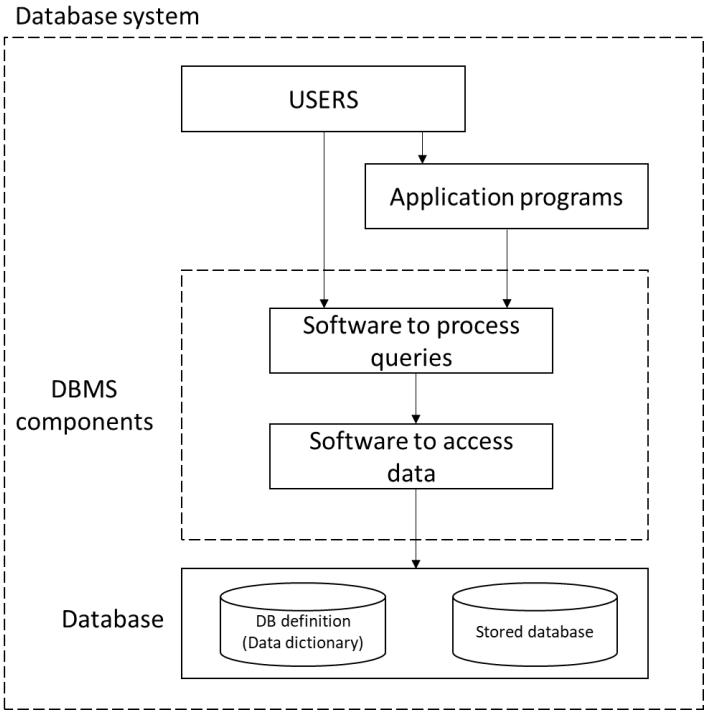


Figure 2. Schematical representation of the components of a database system.

Starting from the 70s, the technological process has made available new informatics tools, such as the database management system (DBMS) which allows integrated management of data. In other words, data although represented once can be used by different application programs (**Figure 1B**). Thus, the data do not rely on a specific application, but each application program can use data throughout DBMS.

Using a DBMS to manage data has many advantages:

1. *Data integration*. Instead of having for each application program a couple of data/programs, there is a unique collection of data that can be reached by several application programs which operate with the data of interest. That, in turn, entails:
   - *Data sharing*. The same data can be used for different users.
   - *Data redundancy*. The existence of a unique database accessible with different modalities removes the need for duplication, allowing a lower memory occupancy and avoiding data inconstancy.
2. *Data independence*. In a database system, the number of programs that use a DB, the type of data, and the relation between data changes over time. In a DBMS, the application programs should not, ideally, be exposed to details of data representation and storage. In fact, the DBMS provides an abstract view of the data needed for the application programs, hiding the other detailed information. Thus, both applications and data can be updated without affects the database.

The components of the database a database system will be addressed in more details in the following paragraphs.

2.2.1 DBMS Users

A variety of people can interact with a database system, and they can be categorized depending on the degree of expertise and on the type of information they want to access. *Naïve Users* or *End Users* use the database invoking one of the application programs which are specifically written for interacting with database. In this class, can be also added online users which indirectly interact with the database using a web user interface. Obviously, this class of users is not aware of the presence of a database and is not specialized in database structures. They mainly interact with the DB to store and use data. *Sophisticated users* should be distinguished from naïve users since they make more extensive use of a DBMS and are able to interact with it without an application program writing their own queries.

Application programmers are computer professionals who write and develop application programs or user interfaces to facilitate data access for naïve and online users. These programmers require a good knowledge of one or many programming languages and have access to the logical schema of the database.

A *database administrator* (DBA) is a person who has the task of designing and completely controlling the database. The DBA has different responsibilities, and the main tasks are:

- Deciding database schema or database contents based on (a) the data type the users need to store and (b) on its use;
- Giving user permission to use the database and deciding which portion of the DB it can view (granting authorities);
- Adopting valid security protocols for ensuring unauthorized data access;
- Ensuring data availability and recovery from failures taking a regular backup in case the system fails;
- Monitoring the performance to maintain adequate performance as requirements change.

2.2.2 Database

In a database system, a database is composed of two different data categories:

1. *Data dictionary (Metadata)*. It stores the definition of data characteristics, restrictions on pertinent data values (data integrity), and relationships. That data schema must be defined before data creation, and it is independent of application programs. It may view as "data about data". For example: when a field storing a molecule Name is created, the DBMS will store (a) when that data was stored in the database, (b) what is the size of the field, and (c) if it is stored as related data to some other data, or if it is independent.
2. *Database*. It is a collection of data stored together; it is the resource for which DBMS was designed.

2.2.3 Database management system (DBMS)

A database management system (DBMS) is basically a tool that is standing between the user and the data. In this way, the users do not directly interact with the memorized data (physical representation), but with a logical representation (Section 2.3). A DBMS is a program, or an ensemble of programs designed to create, organize, memorize, control, and retrieve data in a database. Alternatively, it can be defined as a computerized record-

keeping system whose overall purpose is to store information and to allow users to add, delete, modify, retrieve, and update that information in a way that is both convenient and efficient.

There are two important concepts in the DBMS: data definition and data manipulation. Data definition involves the definition of database structures: what tables, attributes, indexes will be in the database, and so forth, and it is set up with a data definition language (DDL). On the other side, data manipulation refers to the three main operations a user can perform on data stored in a database system: data retrieval, data update, data modification (insertion and deletion of records). Those operations can be communicated to the DBMS using a particular language, the data manipulation language (DML)(Gillenson, 2012). Thus, two are the main languages provided by a DBMS:

- *Data Definition Language* (DDL). It is used to define the content and the structure of the database, i.e., the schema components. The output of DDL is placed in the data dictionary, which contains metadata. General speaking, it is possible to define two different descriptors:

  - Schema. It is the logical description of the entire DB and comprise all the data names, their attributes, and their relations;

  - Subschema. It is the description of the data seen by a specific application; it can be a subset of the SCHEMA, but it may have some changes required from the application itself (additional relations among the data, secure codes, measure units, etc.).

  Each DBMS has its own DDL with specific syntax and semantic.
- *Data Manipulation Language* (DML). It is used to access and manipulate data, such as for adding (inserting), deleting, and modifying (updating) data in a database. The portion of a DML that involves information retrieval is called a *query language* and it is used to figure out data in a database.

Some examples of DBMS software are Microsoft's SQL Server, Oracle Corporation's Oracle, Oracle's MySQL, and IBM's DB2.

2.3 Database definition levels

A database system is a combination of data and application programs that allow the user to access and use a piece of data regardless of how these data are stored and maintained in the system. Moreover, a database system is designed to efficiently retrieve data, and

this efficiency is obtained using complex data structures to represent data. That complexity is normally hidden from users to simplify its interaction with the system. For that reason, there are three different levels of data descriptions (**Figure 3**).

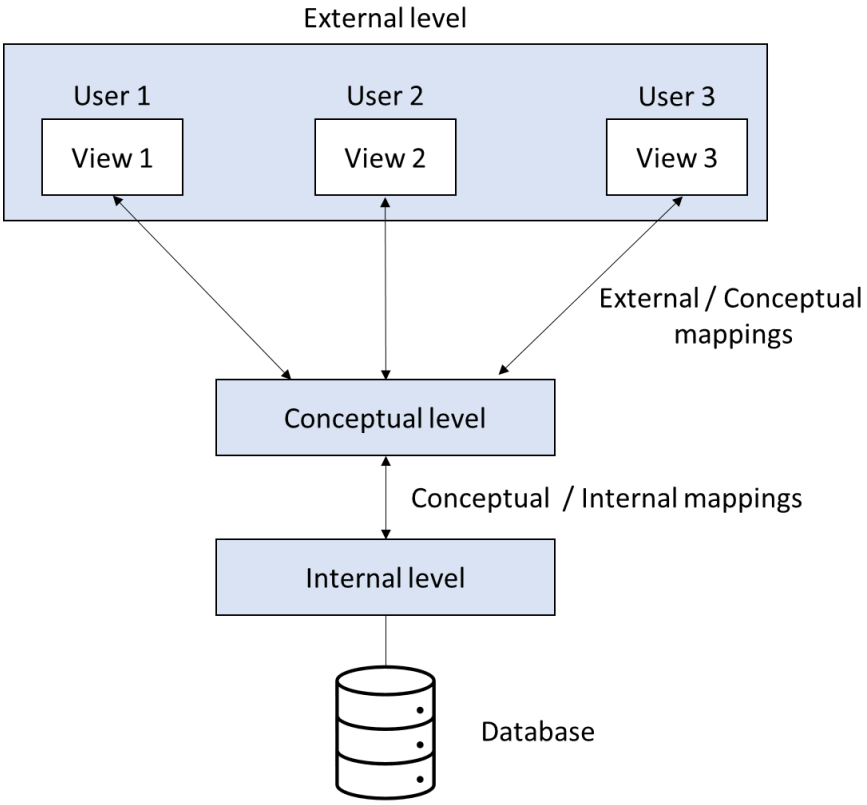


Figure 3. The three-level architecture of a database system.

Physical level (or internal view/schema) is the lowest level of database system architecture. It describes the way in which data should be physically stored and organized in the memory to facilitate their usage.

The *logical level (or conceptual view/schema)* is the next higher level of database system architecture. It describes the overall data structure and what relationships exist among them, without considering their physical organization on the memory (physical data independence). The description of the database is achieved through the Data Description Language (DDL).

External level (or user view / external schema) is the highest level of database system architecture. It describes the actual view of data seen by an individual user or by an application program. Thus, while the logical schema is unique, several external views are defined by the DBA for each user and/or application program.

The communication between the three different levels is guaranteed using two types of mappings: the conceptual/internal mapping and the external/conceptual mappings. This allows transforming the requests and the results among the three levels.
The three levels approach has the advantage to maintain the independence of application programs from the physical and logical organization of the data. Therefore, if it is needed to modify data structures to improve data retrieval performances or if it is needed to change the network node where data are stored to reduce data transfers, it will not be necessary to change application programs since they interact with the external level of data and not with the physical level. Moreover, logical data independence guarantee that a DBA can change the conceptual schema of the database without affecting the external view and, thus, any application programs and users.

2.4 Data Model

A data model is a collection of conceptual/logical tools used to organized data of interest and to describe their schema in a way that is comprehensible to an elaborator. This data schema provides to the DBMS a data representation that allows the management of data organization.
Three main data model categories were developed over the years: hierarchical, network, and relational.
The **hierarchical data model** is one of the oldest DBMS and it was developed at the end of the ’60s when IBM developed and introduced IMS (Information Managing System) on the market. By far, it was also the first DBMS. In the hierarchical model, there is a collection of archives, which are composed of records called *segments*, which are organized in a tree structure. These segments are in hierarchical relations through a parent-child relationship. In a tree structure model (**Figure 4**), it can be identified a 1:N relationship composed of the main segment (parent) and n child segments (N), which can be the parent of other *n* child segments. Thus, each child record can be linked to only one parent.

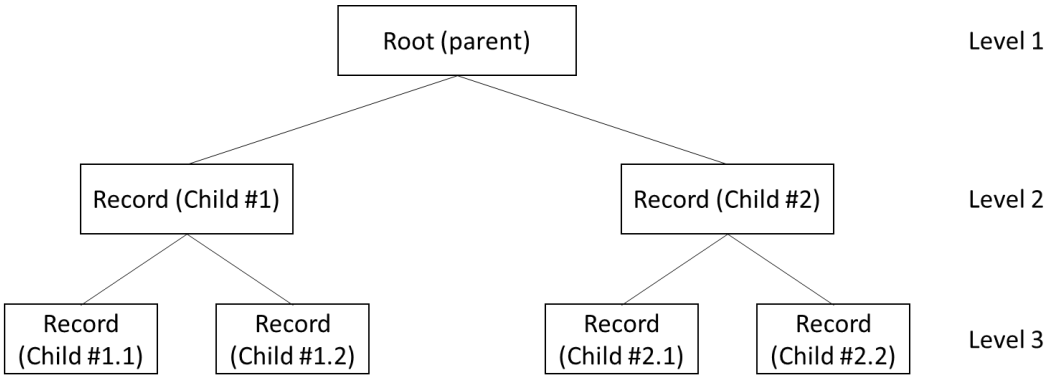


Figure 4. Schematic representation of the three-level architecture of the hierarchical data model

The **network data model** can be seen as a development of the hierarchical model since it is more flexible and can handle more complicated cases. In fact, in a hierarchical model, a child segment can be linked to a single parent segment. In a network data model, each record (child segment) can have *n* child segments (as in the hierarchical model), but also *n* previous segments (parents), allowing a many-to-many relationship in the tree-like structure model. Thus, data are organized more like a graph, where each node is a record, and the arches are the links among the nodes. These links can represent 1:1, 1:N, or N: M relationships.

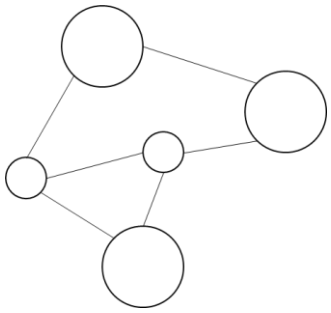


Figure 5. Schematic representation of the network data model.

In the network model, dependencies can be established. For example, the top record in **Figure 5** does not have a direct link with the one on the bottom, but it can reach it using the record in the middle left. On the other side, it can directly reach the record in the middle right. Thus, in the network model, we can often speak of the "navigation" concept referring to the pathway needed to reach a record through these links. It follows that records can be inserted and removed without interfering with the overall structure.

The **relational data model** was firstly introduced in a scientific publication by E.F.Codd in 1970 with the aim to overcome the problem of how to treat data independence. The

relational model is based on two main concepts: tables and relations. In fact, a relational database is perceived by the users as a collection of tables (relations) where data are organized as a set of rows and a set of column fields. In its terminology, each row is named *tuple* and it represents the record that will be memorized in the database. Each record type is defined by a fixed number of columns, which are named *attributes*. Entities with the same columns/attributes will be part of the same table/relation. The model accounts also for the relationships among tables: one-to-one, one-to-many, many-to-many. It is the most data model used by the current database systems.

2.5 Structured Query Language (SQL)

SQL is the abbreviation Structured Query Languages (usually pronounced "sequel"). It is a relational database management system (RDBMS) running under the IBM operating system and the first commercial DBMS that supported SQL was Oracle in 1979. It was standardized and accepted as an American and international standard by ANSI (American National Standards Institute) and ISO (International Organization for Standardization). SQL is a language used to manage a relational database where the information is presented in a tabular form with a set of tuples. Thus, SQL fulfils Data Description Language (DDL), Data Manipulation Language (DML), and Query Language (QL) functions, allowing query and manipulation options based on a set of tuples.

DDL is used to define data structures and to create the database schema. CREATE table, ALTER table, DROP table are typical examples of DDL instructions in SQL.

DML is used to manipulate data itself, using instructions like INSERT, UPDATE, DELETE. Thus, it is used to add, retrieve, and update data.

As for every DBMS, SQL consists of a storage and a management component, where the storage component stores both data and their relationships to which a user can access and query; the management component contains the SQL language to manipulate the database.

One of the main purposes of SQL is to allow access to a relational database by one or more users.

2.6 Big Data and NoSQL

The simplest way to collect and display data is in a table with rows and columns. For most data sets, that is good enough to be read and understand. However, the problem is how to deal with data that are unstructured and that derived from different data sources.

With the increasing volume of data, the term Big Data enters the scene, which normally is referred to the use of a large amount of data, but this is a simplistic definition of the term. However, there is not a single definition for Big Data, but most data specialist supports the concept that Big Data should, at least, have three main characteristics: high-*volume* (extensive amounts of data, ranging from tera to zettabytes), high-*variety* (multiple data formats: structured, semi-structured, and unstructured data), and high-*velocity* (high-speed and real-time processing). In fact, one of the first definitions of the term was proposed by Garner in 2012, who stated: "*Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation*".

NoSQL is the abbreviation for "Not Only SQL", and it refers to database management systems used to manage databases which may not be based on the traditional relational data model and therefore may not have SQL as a query language. In fact, they can store data in both traditional and non-traditional structural languages.

There are at least three main characteristics that distinguish NoSQL from SQL databases:

1. Schema-less: NoSQL databases are independent of schemas, which means that it can be run without the description of all data and data structures in advance as in a relational database.
2. Memorize data in different data formats, storing data in an unstructured or a semi-structured form, such as document, graph, etc. Thus, they do not require rearranging data in a tabular form, with rows and columns.
3. Highly distributable: NoSQL can store and process the information on more than one device while keeping high performances and high scalability.

NoSQL databases can mainly be divided into three different types: document-oriented database, key-value database, and graph database.

As their name suggests, **document-oriented** databases store data in the form of a document, such as XML, JSON (Java Option Notation), or BSON (Binary JSON). Each document has a variable number of properties/attributes that are the information. The most famous NoSQL databases are Mongo DB, CouchDB, and Elasticsearch.

Key-value databases store items as alpha-numeric identifiers (keys) to which are associated values that store the actual data. The values may be simple text strings or more complex data. Since the key allows to quickly retrieve database values, it is normally used

in the context where it is needed a highly scalable retrieval of data, like managing user profiles, retrieving product names, etc.

Graph databases memorized data as graphs with nodes connected with edges. The information may be stored both on nodes and edges. Managing graphs is not a simple task, but they offer a decisive contribution in cases where data are highly interconnected. The strength of graph databases is in the information that can be retrieved following the graph pathway.

2.7 SQL or NoSQL - Factors to consider for selecting a database

Although SQL and NoSQL are often put in opposition, no one is better. In fact, the choice relies on how the data looks like. Important factors to be considered are:

- how data are organized in the data sources where the information has been retrieved;
- how this information should be stored in the database.

In fact, if data are mainly structured, an SQL database may be the right choice. On the other hand, if data are semi-structured or unstructured, NoSQL may be the best bet. Another important aspect that should be considered is the query language. SQL is a milestone of DBMS, and it provides a very intuitive query language, that can also be used by non-experienced users. On the other side, the NoSQL database can store different types of data and they usually have different query languages. Thus, it is often necessary for a further application that is able to convert user queries into NoSQL queries.

SQL and NoSQL have also some differences in their scalability. SQL databases scale vertically, and this means that, for increasing the amount of data in a single server, it is necessary to add new machine components (CPU, RAM, or SSD) to enhance the performances. By contrast, NoSQL databases scale horizontally. This means that it is possible to add more servers to the resources, and data can be distributed across these resources.

Another important difference between SQL and NoSQL databases is the presence or absence of a schema. As highlighted before, SQL databases store data as a set of tables with a fixed predefined schema, and all the data must follow this schema. Consequently, based on the data source where data are retrieved and extracted, it may be necessary a lot of preparation before data uploading. In turn, NoSQL databases allow the use of unstructured data and follow a dynamic schema. Thus, they do not require a predefined structure.

In conclusion, SQL and NoSQL have their specific peculiarities. Making the final decision to use one or the other mainly depends on the type of data in question, the amount of data, and who will be managed the database.

2.8 Data Quality

With the increasing number of data produced and available to the community, the concept of data quality (DQ) enters the scene at the beginning of the 1990s, when data scientists began to consider the problem of the correctness of electronic data stored in databases. In fact, without high data quality, all post-processing analyses will be incorrect, and they will generate erroneous predictions. Thus, to ensure data quality, for making them eligible for the final use, it is necessary to carry out different "preliminary" steps on the data before adding them to a database. Data quality is a fundamental concept in the scientific context. *In vitro* and *in vivo* data are highly heterogeneous both in terms of data types and data quality. Experimental information is usually collected through different sources and disconnected repositories, which makes data quality assessment very challenging. Moreover, it is entirely in the hands of the scientists, which often analyze manually the data sources and try to solve data conflict. Thus, to overcome such shortcomings, scientists need a robust data quality database.

Data quality covers different aspects, although it is normally correlated to the **data accuracy** concept. In fact, when people think about data quality, they often refer to typos' errors (syntactic errors) or to erroneous association values, such as a wrong birth date associated with a person (a semantic error). While syntactic errors are relatively easy to be found and checked, major efforts are needed to dig semantic errors. In fact, identifying a semantic error requires a priori knowledge of the correct value to replace the incorrect one. In most cases, this is not possible. A possible technique to solve this problem consists of looking for the same data in different data sources and finding a consensus match of the information among them (Batini and Scannapieca, 2006). However, data quality is more than data accuracy since it is a multi-dimensional concept. Other important aspects that should be considered are data completeness (missing value) and data consistency. **Data completeness** is one of the most challenges in the big data era and it refers to the presence of missing values in a database. This concept is particularly important when data stored in a database are used to make analyses and predictions. In fact, in some cases, the lack of information makes the data completely unusable, while in other contexts can lead to mistakes or to false conclusions. Thus, an important step in database definition is to

choose which data are critical and necessary and must be complete (not null value). **Data consistency** is another important concept when talking about data quality. It refers to the consistency in the meaning of a particular variable throughout the datasets. This is particularly relevant when data stored in a database have been extracted and aggregated from multiple sources. Discrepancies in the data meanings can create inaccurate and unreliable datasets. A typical example of data discrepancy may be the activity value of *in vitro* experiments, used to calculate protein-ligand interaction. Some laboratories measure this interaction using the IC_{50} value, others the EC_{50} or the relative binding affinity (RBA). Considering a relational database, if this information is aggregated in the same column reporting only the values without providing any other specification, it can generate an inaccurate conclusion about molecule activity. Ensuring data quality can be performed upstream by searching errors in the extracted data, i.e., before uploading them in a warehouse database, or downstream. To ensure data quality, specific competencies are needed since the people involved in these steps should have a clear view of the context in which they are operating.

2.9 Blockchain

Although SQL and NoSQL are undoubtedly important tools to store and retrieve data information, Blockchain technology may be a very useful tool to guarantee the correctness of data. The data in blockchain technology are distributed, decentralized, and, thus, unmodifiable until the modification is not allowed from the other nodes of the block. A blockchain is a digital shared and decentralized archive that can be consulted by everyone who is part of the network. It is a chain of blocks where each block mainly consists of three elements: the data, the hash, and the hash of the previous block. Hash is a unique string of numbers and letters which identify the specific block and its content. Every time a new block is created, a new unique and specific hash is calculated. If some data in the block changes, also the hash will change. Another characteristic that ensures data security is decentralization. The blockchain uses a peer-to-peer network to which everyone can access and participate. Everyone in the network becomes a node and gets a complete copy of the blockchain. When someone creates a new block, this block is sent to every node of the network, which checks the block and if it is correct, the new block is added to the chain. However, if someone adds an incorrect block, this block will be rejected from every node.

In the agri-food context, blockchain technology can be seen as a new “ally”. The important characteristic that has made blockchain appealing for the agri-food field is its capability to ensure the traceability, transparency, and reliability of the food product, i.e., it allows to “tell its story". In more detail, the benefit of blockchain can be more relevant for all the activities that lead to food quality certification. In fact, the blockchain allows creating of an entire supply chain where all the actors can act, such as raw material products, logistics and transports enterprises, packaging industries, etc. However, blockchain technology should not be seen as a warranty of quality that can provide alone the quality of a food product. In fact, if the original data is incorrect, it will not be checked from the blockchain. On the contrary, it guarantees that this incorrect data will be retained as it is along the overall supply chain. It follows that all people involved in the supply chain can see that data, and everyone can ensure its value.

References

Batini, C., Scannapieca, M., 2006. Data Quality - Concepts, Methodologies and Techniques, 1st ed. Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/3-540-33173-5

Gillenson, M.L., 2012. Fundamentals of Database Management Systems, 2nd ed.

CHAPTER 3

Molecular Docking: A Contemporary Story About Food Safety

Francesca Cavaliere, Giulia Spaggiari, Pietro Cozzini

In book: Molecular Docking for Computer-Aided Drug Design.
Publisher: ELSEVIER
Editors: S. Mohane Coumar
https://doi.org/10.1016/B978-0-12-822312-3.00025-4

3.1 Introduction

What is the link between medicinal chemistry and food safety, and could we apply to food problems the same "in silico" approach used in medicinal chemistry in the last 30 years? "Questa o quella per me pari sono ." ("Neither is any different.") sang the Duke of Mantova from Rigoletto by Giuseppe Verdi. In the opera, the meaning is regarding womendall women are equal for the duke of Mantova, no difference among them. In this manuscript, it has no negative facet, but it is just referred to molecules. From a chemistry point of view, all the molecules are "molecules," independently from the research field. Then we can apply the same computational methods to different molecules considered - drugs or lead compound or food contact chemicals. The main difference between medicinal chemistry and food science is shown in Fig. 1.

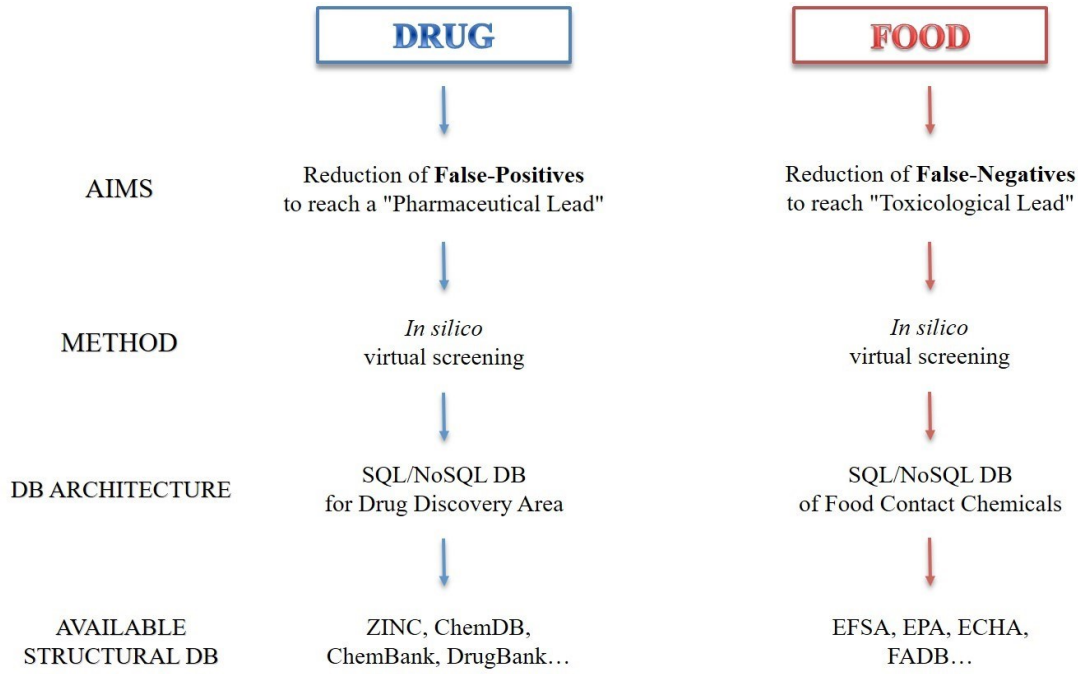


FIG. 1 The different approaches to screen compounds in drug discovery and food safety' areas. The same in silico methods, widely used in drug discovery, can be applied in the food field. The unique difference between them is the aim of the screening process: in drug discovery, it is important to retrieve compounds that strongly bind target protein, avoiding false positives; in food safety, the aim is to retrieve all possible food contaminant molecules that have the capacity to bind the target protein, also with low binding affinity, avoiding excluding true negatives.

Molecular docking is a well-known approach in medicinal chemistry widely used to study the interaction between a receptor and a possible lead compounds, after a screening of a huge number of compounds. While docking in medicinal chemistry is a technique applied for several decades, in food science, it was born 15 years ago, more or less. It could be

considered as a new promising application in food science for the discovery of new possible food contaminants, acting as endocrine disruptors, or to understand a mechanism of binding to activate a flavor (umami, sweet, salty, etc.) or to decipher the activity of a dimer against a monomer. Food safety refers to handling, cooking, and storing food in order to reduce the risk and protect people from foodborne illnesses caused by microbes, chemicals, and other food contact chemicals. A very high number of substances can contaminate food causing a possible risk to the people. An important milestone for screening/docking approaches is the availability of a three-dimensional (3D) database to collect the huge amount of food contact chemicals in order to make possible testing these compounds otherwise unfeasible with traditional in vitro tests. (To give an idea of the huge chemicals that can interact with food, the most collection of substance information is CAS REGISTRY. It contains more than 163 million unique organic and inorganic chemical substances and more than 68 million biosequences.) The application of computational methods, such as repository or database design, screening, and molecular docking, in food safety, could be applied to predict the interaction between food contact chemicals and different receptors/targets involved in human diseases and/or to decipher their mechanism of binding.

3.2 Food Safety

How often do we ask ourselves if the food we are eating is safe? Do we know if it is free from bacteria, viruses, chemicals, and other contaminants? Over the years, food safety is becoming one of the major issues of public concern, food policy, industry, and research. There is no uniform/standard definition of food safety, but anyway in 1993, OECD, the Organisation for Economic Co-operation and Development, gave it a working definition, namely “a reasonable certainty that no harm will result from intended uses under the anticipated conditions of consumption”. Food safety can be defined as the probability of not contracting a disease as a consequence of consuming food. In a broad sense, food safety refers to the scientific process to deal with, manufacture, and store food in order to prevent foodborne diseases. The concept of food safety is closely related to the concept of food security: it is not enough to ensure that the food is safe from a health point of view, but it is necessary to delete the obstacles to food such as the supply, the poverty, and the climate changes. In 1970, the World Food Conference defined food security in terms of food supply; it was the Word Food Summit to provide the final definition: “Food security exists when all people, at all times, have physical and economic access to

sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life". These two elements do not take into account a third important factor: food quality. Let us say that food safety is obtained when everyone has access to food guaranteed as healthy from a hygienic and a nutritional point of view. Therefore, in order to fully understand what food safety means, it is required to define the other two terms: hazard and risk. These two words are often used interchangeably or confused with each other, but they have a different meaning. A hazard is the capacity of a thing to cause harm and in particular referred to food safety. It is any agent (biological, chemical, or physical) or substance in food with the potential to cause adverse consumer health effects. A risk is the probability of an adverse effect in an organism caused by exposure to an agent. For example, salmonella, a biological agent that can contaminate different food such as raw eggs, is considered a biological hazard for the consumer. The risk of getting salmonella food poisoning is minimal when the egg is cooked, but, otherwise, if the eggs are eaten raw, the health risk from salmonella will be higher as a result of the higher likelihood that the hazard will be present and consumed. Ensuring food safety is a significant challenge to protect public health in both developing and developed countries. For this reason, the food safety risk analysis was introduced: it is a fundamental food safety aspect that wants to reduce foodborne illness. This approach aimed at producing high-quality goods and products to ensure safety and protect consumers' health and comply with international and national standards and market regulations; this consists of three components: risk assessment, risk management, and risk communication. In a typical instance, a food safety problem is identified, and risk managers initiate a risk management process, which they then see through to completion. Risk management is defined as "the process of weighing policy alternatives to accept, minimize, or reduce assessed risks and to select and implement appropriate options". The risk assessment process consists of hazard identification and hazard characterization. Fundamental for all these processes and in general for food safety is the Hazard Analysis Critical Control Points (HACCP), an internationally recognized system, composed of seven points, used to identify, evaluate, and control hazards to food safety. These principles are included in the international standard ISO 22000, a complete food safety management system. Apart from this, the presence of regulations established by national and international organizations (such as the European Food Safety Authority [EFSA] in the European Union, which provides scientific advice and information on existing and emerging risks

related to the food chain, and the Environmental Protection Agency [EPA] in the United States, which is in charge of environmental protection and that of human health) ensures that consumers are more protected from health risks. The World Health Organization (WHO) estimated that 600million (almost 1 in 10 people in the world) fall ill after eating and/or drinking contaminated food or water resulting in around 420,000 death every year. In recent years with the movement of the people, the increase of globalization, the modernization of industries, and the international trade, people and/or consumers are exposed on a daily basis to chemical substances, and consequently, the risk of foodborne diseases has increased. Therefore, the control of contaminates and the prevention of foodborne diseases have become one of the most public and private health problems in the contemporary world involving the cooperation of all stages of the food chain: from the field to the table. In the last years, with the increase of diversity and complexity of contaminants and foodborne diseases, not only researchers but also industries and consumers are urged to discover new rapid, sensitive, and selective methods to quantify and qualify damaging substances in food products. Therefore, in vitro and in vivo techniques, such as colorimetric detection, fluorescence sensing (using high quantum yields, narrow and symmetric size-tunable emission, and pronounced photostability, quantum dots, and high signal-to-background ratio and sensitivity as a result of large anti-Stokes shifts, UCNPs), electrochemical sensing, chromatographic separation (high-performance liquid chromatography), immunoassays (enzyme-linked immunosorbent assay), and real-time and in situ analytical methods, have been joined by in silico methods (Liu et al., 2018). These methods, as well as being quick and inexpensive, make up the alternative to animal testing, described by the principles of three Rs (3Rs): replacement, reduction, and refinement (https://www.nc3rs.org.uk/).

3.3 Databases and big data in food safety

Evaluating the effects of food contaminant chemicals is a challenging task. Human exposure can derive from different sources, such as molecules that are naturally present in food products (mycotoxins produced by fungi, flavonoids, etc.), intentionally added molecules (additives, flavorings, etc.), or unintentionally added to food. Some examples are pesticides, biocides that are in contact with the food product, or molecules derived from its packaging and storage such as bisphenols, polycarbonates, etc. The exposure to one chemical can occur via different sources, but rarely humans come in contact with just one single chemical. Instead, we are exposed to a mixture of contaminants. The scenario

becomes more complicated if one also considers environmental chemicals and food contaminant metabolites. Thus, the number of molecules that require risk assessment analysis is very high. Moreover, considering the new molecules that are produced every year and that could accidently be released in food, environment, etc., the number of chemicals that should be investigated increases rapidly. Risk assessment of these huge amounts of chemicals using standard toxicological in vitro methods is unthinkable, although, with the advent of high-throughput screening (HTS), toxicological data can be retrieved quickly. However, considering chemical mixtures exposure, it is physically impossible to test all combinations. Thus, in silico methods can be applied to screen this amount of chemicals in a very fast and economic way. To speed up these analyses, it is fundamental to have access to databases that store all food contact chemicals containing information regarding their physical/chemical properties, the 3D structures, their bioactivity, etc. The huge amount of data produced has raised the need for efficient methods that allow the collection, storage, and processing of data. In this scenario, the big data methods are emerging and becoming an increasingly popular term. Big data is a relatively recent word that has become a ubiquitous termin different sectors of society: business, health care, government, etc. The term is seldom used in the food safety field. The principal reason is that toxicological data were produced very slowly due to laboratory experiment time limitations. However, after the advent of techniques that allow laboratory automation and HTS, toxicological data are produced very rapidly and at a low cost for many molecules. Moreover, with the advances in data mining and deep learning, more chemical information can be also retrieved from various online sources, including scientific articles and patent documents. Thus, from a lack of data, it has been passed to “data overload” (Richarz, 2020). Many definitions of big data exist and the majority of them refer to the characteristics that a database should have, named versus attributes. Currently, there are more or less 10 different attributes for big data, but the 3 common versus are volume, velocity, and variety. Volume refers to the amount of data generated, velocity refers to the speed at which these data are produced, and variety refers to the types of data. Based on the context and the use, big data can include other attributes, such as variability, veracity, value, etc. The European Commission (EC) has defined big data as “the large amounts of different types of data produced with high velocity from a number of various types of sources” (European Commission, 2014). Because of the complexity of data generation and curation, big data requires a high-performing computer

(HPC) infrastructure. HPCs are very helpful not only to store and manage this high-velocity flow of data but also to make possible the collection of new insight, solutions, and decisions based on this information. The EC definition has also stated: “Handling today's highly variable and real-time data sets requires new tools and methods, such as powerful processors, software and algorithms” (European Commission, 2014). We thought that this definition could be the best one in the context of food safety. Data and information are scattered across food, health, and agriculture sectors for food assessment. As the information is derived from different assays and techniques, many different types of data are produced and should be stored and processed. Moreover, considering in silico assessment, it is also mandatory to store chemical information and 3D structures. Thus, different types of sources and data are used (variety). Although data are not yet generated in real time as in other big data fields, the speed by which they are produced is increased in the last years with the advent of HTS, omics technologies, and (bio) monitoring (velocity). The first requirement of big data in food science is the collection of information from different sources considering different aspects of the food toxicology and food safety fields. Thus, a database should be storing and making accessible information regarding the physical/ chemical properties, the 3D structures of molecules along with toxicological data, derived from different assays, and regulatory information. With the free access, online databases, chemical structures, and data are available for their use in cheminformatics, bioinformatics, systems biology, drug discovery, and food science. From the computational point of view, different public databases store important information, which is currently used in drug discovery and design. Just to cite some of them, PubChem is a large public repository containing information on chemical substances, their biological activities, and their chemical structures. Another chemical database is ChemSpider, a free chemical structure database providing fast text and structure search access to 85 million chemical structures from 275 different data sources. ZINC is a free database that contains the 3D formats of over 230 million purchasable compounds in a ready-to-dock format and over 750 million purchasable compounds allowing the possibility to search for analogs in a very fast way. Moreover, 3D databases are also present in literature that are specific for in silico screening in food toxicology. For example, Ginex et al. have released a 3D version of the EAFUS (Everything Added to Food in the United States) list, a sum of WHO, FAO food additive databases (Ginex, Spyrakis, & Cozzini, 2014). Data stored in these databases contain important

toxicological information and comprise a variety of different types of data: in vitro and in vivo assay results, in silico predictions, gene arrays and omics read-outs, regulatory data, 3D and 2D chemical structures, physical/chemical information, etc. All these information represent a big data set (volume) containing several different types of data (variety and variability) and data can be collected in a single repository or otherwise connected. Retrieving information from different sources highlights the importance of uniform data to avoid incongruence among them. In fact, some efforts should be made in the direction of database data quality to enforce the utility of big data in drug design and food safety fields. For example, an important point in chemical toxicity data is the identity name of the chemical used. Each molecule must be having an unambiguous name linked to a unique 3D structure. This issue should be guaranteed by the use of CAS numbers, but it is not uncommon to find some errors in public databases. Moreover, errors in chemical structures are not so rare. Williams and Ekins (Williams & Ekins, 2011) estimated that around 5%-10% of molecule structures have errors in their stereochemistry, valency, and charge. Thus, an important issue is the data curation to improve data quality. There is also a great data variability in terms of differences in data measurement and types of assay across different laboratories. Therefore, data could not be comparable. Data standardization should be desirable. The use of nonrelational databases is becoming more common, as they are open source and horizontally scalable and they are referred to as NoSQL databases. Why big data is becoming so popular? How could it be useful in food safety? Correlated with the concept of the term big data, there are techniques such as text mining and machine learning methods. These methodologies, in some cases, allow us to use the big amount of data to find new knowledge from already available information in a perspective manner. Using information from human cell lines, HTS assays, in vivo animal models could allow the building of predictive models for different applications, such as computer-aided drug design (for the development of new drugs), food toxicology, and/or predictive toxicology (for safety assessment and decision-making). Moreover, the use of big data databases also allows to reduce unnecessary in vivo studies. Hartung et al. (Hartung, 2019) have reported that, on average, every assay was carried out three times and sometimes more than this value. For example, they have reported that two chemicals have been tested more than 90 times in the Draize rabbit eye. Moreover, having a database that stores all chemical information of food contact chemicals, such as the 3D structures, can increase the velocity of in silico methods results.

Virtual screening, molecular docking, and molecular dynamics can take a great advantage by the usage of these data.

3.4 *In silico* methods

In silico methods are computer methods (computing hardware, algorithms, programming, databases, and other domain-specific knowledge) used to study molecular systems in the fields of computational chemistry, computational biology, and material sciences. Computational methods developed since the 1950s with the increase of computers used for predicting and studying the physical-chemical proprieties, the interactions, and the structures of molecules. Molecular modeling includes all those theoretical methods and computational techniques, such as homology modeling, molecular docking, and molecular dynamics, which are used to represent and/or simulate the behavior of molecules. It, therefore, allows the use of innovative in silico methods, based on the use of computers and information technology, to predict the behavior of biological molecules. Molecular modeling, by studying the energy state of molecules and exploiting calculation algorithms and force fields, or a set of parameters that expresses the potential energy of a particle system, is able to predict and determine quickly and at a low cost the final structure of a molecule. The sources of starting data for the molecular modeling come from experimental determinations (X-ray, nuclear magnetic resonance, and cryogenic electron microscopy) or computational structure prediction, based on homology modeling, in the event that the 3D structure is not present. 3D databases, such as the Protein Data Bank (PDB) for protein structures (https://www.rcsb.org/) and the Cambridge Structural Database (CSD) (https://www.ccdc.cam.ac. uk/solutions/csd-system/components/csd/) for small organic molecules, contain experimental data. The three parameters we have to consider to understand the quality limits of structural data are (1) resolution (Å), which is a statement of the accuracy in data collection and not a measure of the accuracy in refinement, (2) R-factor, which is a measure of how well the refined structure explains the observed data, and (3) temperature factor, which models the effects of static and dynamic disorder in the crystal. All these parameters are fundamental to choose the best “starting point” for the following computational prediction. A schema of in silico approaches in food safety is shown in Fig. 2.

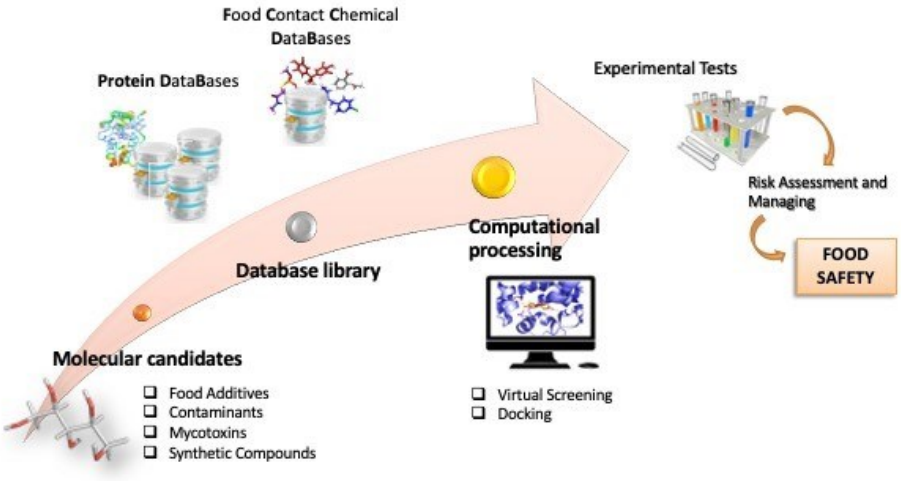


FIG. 2 The in silico approach in food safety schema.

3.4.1 Molecular Docking

Molecular docking is a complex and simple multistep computational technique used to predict and evaluate the structural chemical-physical interaction between two molecules. The method aims to identify the correct positions of the ligands in the binding pocket of a protein and to predict the affinity between the ligand and the protein. Ligand-based and structure-based are the two approaches for virtual screening. Structure-based virtual screening is based on the protein cavity shape, while ligand-based virtual screening refers to the shape of the natural ligand. At the basis of the docking, there is the molecular recognition between the two molecules that interact according to the "lock and key" model developed by Emil Fischer in 1890. In this model the protein has a conformation where the ligand "fits" perfectly, just as it happens for a key inside a lock. The highly specific molecular complementarity between key (ligand) and lock (receptor) plays a fundamental role in biological processes. The receptor's ability to bind to its ligand with high specificity and affinity is due to the formation of a series of weak bonds and favorable interactions. Usually, the interaction between the ligand and its receptor involves the formation of weaker and reversible forces such as (1) hydrogen bonds (10e40 kJ/mol); (2) hydrophobic interactions that constitute the "driving force" capable of promoting bond formation; (3) van der Waals forces (0.03e0.1 kcal/mol); (4) electrostatic interactions (0.3e4 kcal/mol); (5) pep interactions; and (6) coordination with metals. Electrostatic interactions and hydrogen bonds provide specificity to the protein-ligand

interaction and determine its complementarity. During the formation of the complex, a series of enthalpic and entropic interactions are established between protein and ligand, which are mutually concerted. There is, therefore, a variation of enthalpy (due to the formation of intra- and intermolecular noncovalent bonds) and entropy (due to desolvation) in the system, with consequent variation of free energy. The binding affinity between the molecules, a ligand (L) and a protein (P), is characterized by the dissociation constant (Kd):

$$K_d = [L]\,[P] / [LP]$$

corresponding to the process $LP \leftrightarrow L + P$.

The fundamental equation that governs everything is:

$$\Delta G = \Delta H - T\Delta S$$

where ΔG is the change in free energy of a reaction, ΔH and ΔS are the corresponding changes in enthalpy and entropy, and T is the temperature of the system. The binding affinity can be expressed either in terms of the equilibrium constant (K) for the formation of the complex between two molecules:

$$\Delta G^0 = RT\ \mathrm{In}\ K_d$$

where R is the universal gas constant and T is the absolute temperature.

Even while the interactions between protein and ligand are important for generating a positive enthalpy of binding, we must also consider the presence of the water. In fact, molecular recognition takes place in an aqueous medium. Both the protein and the ligand are solvated before complexation; the formation of the intermolecular bond requires the desolvation of the ligand and the macromolecule with simultaneous breakage of the hydrogen waterereceptor and watere ligand bonds (Murcko & Murcko, 1995). The water molecules are organized in such a way as to form as many hydrogen bonds as possible and thus decrease the entropic contribution of the interaction. The clear difference in free energy is often close to zero as many of these breaking bonds are reformed between the ligand and the receptor and the water molecules reorganize around the newly formed complex (Fersht, 1987; Salari & Chong, 2010). For this reason, in order to obtain a reliable energetic estimation of the overall binding process (the total free energy of binding, $\Delta G°_{bind}$), we must use an equation like this:

$$\Delta G°_{bind} = \Delta G^{\circ compl}{}_{solv} - \Delta G^{\circ prot}{}_{solv} - \Delta G^{\circ lig}{}_{solv} + \Delta G°_{int} - T\Delta S° + \Delta\lambda$$

where $\Delta G°_{int}$ is the interaction free energy of the complex, the solvation energy of the ligand ($\Delta G^{\circ lig}{}_{solv}$), the protein ($\Delta G^{\circ prot}{}_{solv}$), and the complex ($\Delta G^{\circ compl}{}_{solv}$), and the entropic

(TΔS°) and conformational (Δλ) changes (Spyrakis et al., 2010). However, as Dill said, "Biological interactions are concerted events, not neat sum of terms where each represents an ingredient of the overall process." Thus, even the best practices in treating each of these disparate interaction types individually will not necessarily yield an accurate and reliable $\Delta G°_{bind}$ in the end (Dill, 1997). Docking software can be differentiated based on their two main components: the sampling algorithm, which searches the possible molecule position, and the scoring function, which evaluates the interaction energy of each position. The first is an "easy" geometrical aspect that "mixes" two or more bodies (molecules) in the same Cartesian space without superpositions and obeying to some elementary chemical rules. In fact, anybody (or molecule) is composed by several or many solid spheres and springs. One of the most difficult part of the molecular docking is due to the fact that it involves many degrees of freedom, the high dimensionality of the energy surface where the search for the global minimum is performed by a docking program. Each algorithm generates poses (where and how a ligand binds a protein); a series of conformations result from rotation about single bonds (Fig. 3). For a molecule with n rotatable bonds, if each torsion angle is rotated in increments of x degrees, the number of conformations is $(360°/x)^n$.

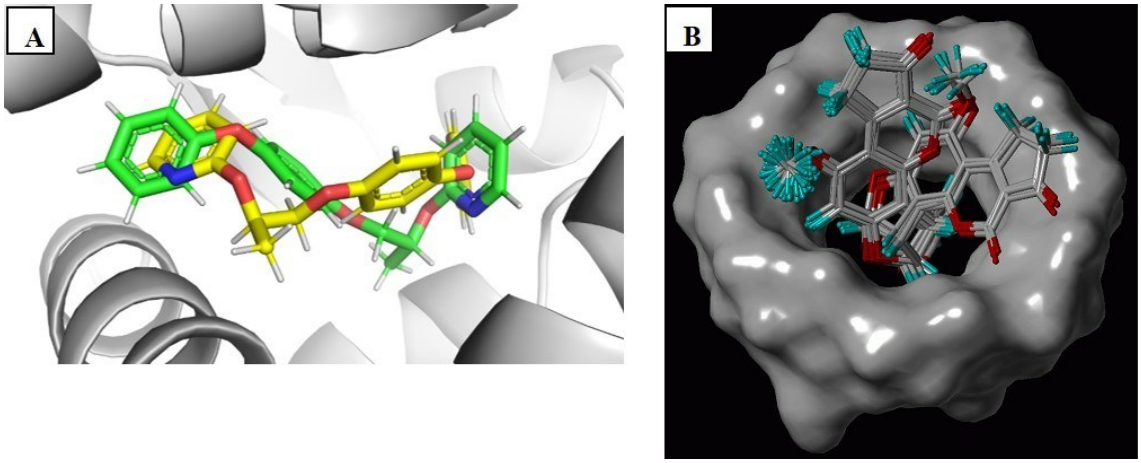


FIG 3 Different poses generated by a generic docking software (GOLD). (A) A protein-ligand complex. In grey the protein (PDB ID: 1FM6) and in yellow and green two different poses of a pesticide (pyriproxyfen) are shown. As we can see the two ligand poses are opposite to each other. (B) The most important poses of a mycotoxins (Aflatoxin) within the cavity of a beta-cyclodextrin. All the poses identify the same position.

We can classify the different search algorithms and consequently the different docking software according to the degrees of freedom that they consider: 1. Rigid docking: This

type of molecular docking ignores the flexibility of the molecules, both for the ligand and for the protein, and treats them like rigid objects. In this case, the side chains and the backbone of the two molecules are kept fixed with no torsion angles or distance between two atoms allowed to change upon the docking simulation. 2. Semiflexible docking: During this docking, the receptor remains unchanged, while the conformation of the ligand changes. It focuses on the changes in the ligand structure and it is usually used for the docking between small ligands and macromolecules. 3. Flexible docking: This docking, the most common today, considers every conformational change of both the protein and the ligand. We can summarize scoring functions in three classes: (1) force fieldebased, where the binding affinity is estimated by the sum of the strength of intermolecular (van der Waals and electrostatic interactions) interactions between all atoms of the two molecules; (2) empirical, where the binding affinity between the two molecules is estimated by the number of various types of interactions; (3) knowledge-based, based on a statistical analysis of observed pairwise distributions. More than 60 different docking software have been reported in the literature, such as AutoDock (Morris et al., 2009), DOCK (Allen et al., 2015), GOLD (Verdonk et al., 2003), FlexX (Schellhammer & Rarey, 2004), Glide (Friesner et al., 2004), Surflex (Jain, 2003), distinguished by the algorithms, the evaluation methods, the docking types (rigid, semiflexible, or flexible docking), and more. One or more scoring functions can be associated with each scoring program. There is no docking-scoring combination valid for each type of analysis, but these combinations must be evaluated based on the characteristics of the target. Most docking scoring functions use very simplified models for hydrophobic interactions; then simulating the binding (or docking) process with explicit terms for entropy has proven to be an elusive goal. To get around this, in 1991, Abraham and Kellogg developed HINT (hydropathic interactions), a scoring function that simulates and quantifies all of the subtle effects contributing to entropy in the docking process (Shoichet & Kuntz, 1993). HINT uses a force field that allows it to evaluate both the entropic aspect (due to desolvation) and the enthalpic aspect (due to interactions). The fact that you also evaluate the entropic aspect is what differentiates HINT from other scoring functions. The function also includes the computational titration method for predicting and optimizing the protonation state of ionizable residues at a complex interface and the Rank algorithm for rationalizing the role of structural water molecules in protein binding pockets (Amadasi et al., 2008; Cozzini et al., 2004). Deciding which

program is the best one is a challenging task. Docking software is normally validated using a training set of proteineligand complexes with the known crystal structure and known binding affinity. The 3D complex is used to validate the internal algorithm of the docking package to predict the correct binding pose based on the crystallographic one. The correlation is usually assessed using the root-main-square deviation between the docked and the crystal ligand pose: the lower the value, the better is the reliability of the docking algorithm. Binding affinity value is used to test the ability of scoring functions to discriminate between compounds having strong-, medium-, and lower binding affinity, or in alternative to test their ability to discriminate among a library of true, false, and decoy compounds. There is no general rule for choosing the best docking program, but it is advisable to utilize a software that was validated against the same class of protein under investigation or with proteins sharing common physical-chemical and shape characteristics. However, the goal of any docking program is to be used for every proteineligand system. It has been estimated that the averaged success rate in predicting the correct poses and top scores is in the range of 54.0%e67.8% for commercial programs and in the range of 47.4%e68.4% for the academic ones (Pagadala et al., 2017). Even if the performances are quite high, a certain grade of uncertainty and error can occur. Thus, the best practice is to apply a consensus score prediction. The concept was introduced to enhance the performances of docking protocols. Multiple scoring functions are simultaneously used rather than a single one. Compounds are then ranked based on the consensus existing among them and only the top scored compounds common to the scoring functions will be used for further in vitro/in vivo assays (Wang & Wang, 2001). The concept of consensus scoring could be seen as weather forecasts: if many of them agree that, during the weekend, there will be the sun, and just one predicts a thunderstorm, then it is more probably there will be a sunny weekend. Compared to a single scoring procedure, it has been shown that the combination of different scoring functions reduces false positives and hence improves the hit rates (Wang & Wang, 2001). It has also been reported that the use of three scoring methods is enough to enhance the capability hit rates of ~50% (Bissantz et al., 2000).

3.5 Case studies

In food science, molecular docking approach is applied for different needs: to study the interaction of a food chemical with a protein receptor understanding the mechanism of binding or the competition between a natural ligand of a protein and a food molecule or

to design chemosensors able to include a toxin in a cavity to take away the dangerous molecule from water or food. Hereinafter, we illustrate a few real cases of docking applications. As the majority of our key studies are focused on nuclear receptors (NRs) and how food contact chemicals may act as endocrine disruptors, Fig. 4 shows a schematic view of the perturbation induced by endocrine disruptor compounds (EDCs) activity.

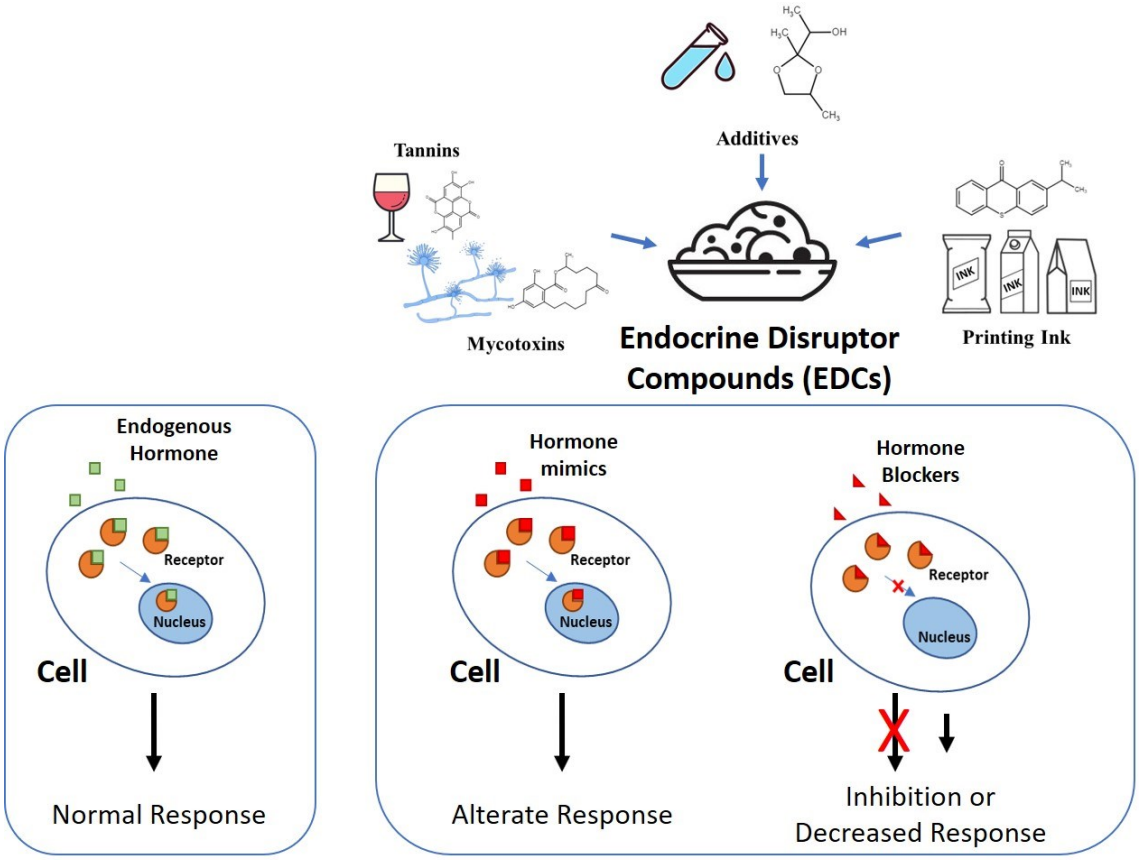


FIG. 4 How food contact chemicals can affect the nuclear receptors pathway. Endocrine disruptors are chemicals that can interfere with endocrine or hormonal systems. These disruptions can cause adverse effects such as tumors, birth deffects, and several other disorders. In fact, these molecules can decrease or increase normal hormone levels altering their normal production.

3.5.1 Mycotoxins Detection

Mycotoxins are important because of possible danger to humans; depending on the intake dose, they can act as endocrine disruptors binding mostly to NRs. Two well-known NRs are recognized as responsible for breast cancer in women and prostate cancer for men: estrogen receptor (ER) and androgen receptor (AR), respectively.

3.5.1.1 Aflatoxins and ochratoxins

Cyclodextrins are cheap and relatively easy to manage, and they show a lipophilic cavity; mycotoxins are small molecules with a hydrophobic and a hydrophilic side, present in plants but dangerous for humans. They could be included in a cyclodextrins cavity, the lipophilic side, and the complex could be detected using spectroscopy fluorescence (Cozzini et al., 2008). Moreover, the mechanism of action (MOA) understanding allows designing specific cyclodextrins customized for different toxin structures (Fig. 5) (Amadasi et al., 2007).

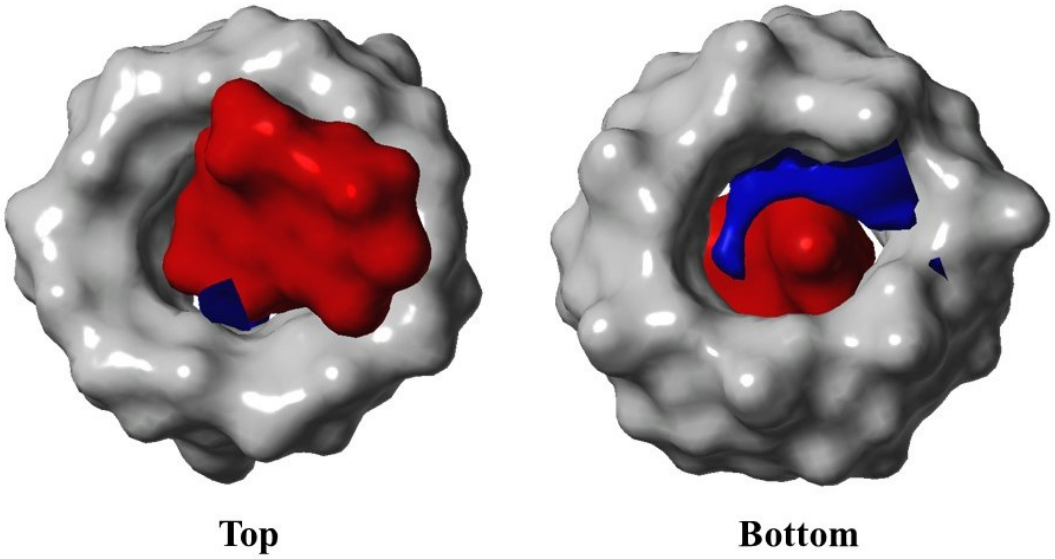


FIG. 5 Results of a molecular docking simulation between beta-cyclodextrin and a mycotoxin. Molecules are depicted using occupancy volume. In red is the volume occupied by the mycotoxin, in gray the cyclodextrin volume, and in blue the empty volume that could be filled by water molecules (GRID analysis).

Aflatoxins and ochratoxins (Fig. 7) affect a large number of mays and grains production in Italy (more or less 5%). They can be detected using cyclodextrins and fluorescence spectroscopy. In particular, in this case, the modeling allowed us to understand why the same beta-cyclodextrin can include aflatoxin and not ochratoxin. The latter requires a specifically designed cyclodextrin.

3.5.1.2 Zearalenone

Zearalenone (ZEN) (Fig. 7) and its metabolites that are known to act through activation of the estrogen receptor alpha (ER alpha) has been studied (Cozzini & Dellafiora, 2012) against ER to understand if they can bind competitively with the endogenous ligand, estradiol. A molecular dockingebased study demonstrates that it is possible to

discriminate between cisand trans-isomers for ZEN using the same docking approach: for the cis isomer, a stronger interaction has been predicted (Dellafiora, Galaverna, et al., 2015). Moreover, ZEN and its reduced metabolites have been used within the framework of reduction, refinement, and replacement of animal experiments (Ehrlich et al., 2015). Mixed methods, docking/scoring, and toxicological methods for identification and characterization of chemical hazards have been developed. The results suggest that activation of ER alpha may play a role in the molecular initiating event and be predictive of adverse effects. The investigation of receptoreligand interactions through docking simulation showed the suitability of the model to address estrogenic potency for this group of compounds. Therefore, the model was further applied to biologically uncharacterized, commercially unavailable, oxidized ZEN metabolites (6 alpha-, 6 beta-, 8 alpha-, 8 beta-, and 13- and 15- OH-ZEN). The main conclusion is that, except for 15-OH-ZEN, the data indicate that in general, the oxidized metabolites would be considered of a lower estrogenic concern than ZEN and reduced metabolites. 5.1.3 Alternariol Another mycotoxin, a widespread microfungi secondary metabolites that may accumulate in crops and enter in contact with some foods, is from Alternaria species (Dellafiora, Dall'Asta, et al., 2015). The whole corn production in Italy is affected by mycotoxin alternariol every year, depending on temperature. Thus, the comprehension of the MOA of alternariol and its derivatives against some proteins is crucial to understand if toxic potencymay drastically be reduced by metabolic modifications. Alternariol (Fig. 7) and alternariol methyl ether show evidence of toxicity binding to topoisomerases but it is not enough. Too many compounds and its derivatives are candidates to be endocrine disruptors because of binders of several proteins. Because wet-lab tests are expensive and require long times, it is really challenging to have a fast and cheap method to discriminate among possible poisons and no poisons as in silico methods. In this work, the methods have been applied for the topoisomerase case. 5.2 Ellagitannin Metabolites Dellafiora et al. have applied the same in silico approach to ellagitannins and their metabolites (glucuronidation, sulfation, and methylation, occurring in vivo) (Fig. 7) (Dellafiora et al., 2013). Urolithin metabolites could act as phytoestrogens able to interact with the ER binding cavity. These hydroxylation patterns are presented in our models coming from berries, walnuts, pomegranate, and oak-aged red wines. They are well known as “natural drugs” that can contribute to decreasing the risk of some ERdependent diseases. Ones

again, the in silico approach to study the MOA suggested that hydroxylation can play an important role in the agonistic behavior of these derivates.

3.5.3 Printing Inks

As stated, another tumor marker is the AR, involved in prostate cancer, able to interact with many food contact chemicals. Thioxanthones are analogs of xanthone and are largely used as photoinitiators (TX) by printing industry to promote ink polymerization. However, a certain level of contamination by isopropyl thioxanthone (ITX) and 2-ethylhexyl-4-dimethylaminobenzoate has been found in food products, especially in infant formulas (as reported by the European Food Safety Authority in 2005). Ginex et al. (Ginex, Dall'Asta, & Cozzini, 2014) have reported an in silico approach to predict the binding affinity of thioxanthone derivates and thioxanthone metabolites against AR. In fact, it is well known by in vitro analyses that this class of compounds is able to bind to AR. Using the in vitro affinity values of some TX compounds as validation test of in silico procedure, different metabolites have been computationally analyzed to predict their binding affinity for the ligand binding cavity of AR. The authors have found that different metabolites have the same or higher binding affinity of 2-ITX, 4-ITX, and 2-Chloro-TX, which are the three well-known AR-mediated endocrine disrupting compounds.

3.5.4 Food Additives

More than 3000 substances could be added to the food depending on the different countries' laws. In the search for xenoestrogens within food additives, the Joint FAO-WHO expert committee database, containing 1500 compounds, was checked using an integrated in silico and in vitro approaches (Amadasi et al., 2009). The main question was, are we confident about the safety of food additives allowed? Docking and screening could assume the same meaning but, usually, screening is reserved to “screen” a huge number of molecules against one or more receptors based on ligand structure or receptor cavity structure. Both techniques can be applied in a pipeline to extract a smaller set of data from a big database (screening) to be docked within a receptor cavity. Wet lab tests applied to predicted molecules identified propyl gallate as an antagonist and 4-hexylresorcinol as a potent transactivator (nanomolar concentration) based to in silico prediction. The final meaning is to consider these two compounds as probable ER interactors but not certified as poison.

3.5.5 Bisphenols in Food

The bisphenol case is another example used to demonstrate that docking methods could be a valid approach to screen estrogenic and androgenic activity of food contact materials (Cavaliere et al., 2020). One of the most common bisphenols is bisphenol A (BPA) or 4,40-isopropylidenediphenol (Fig. 7). This plastic, used to make many food containers, has been classified by the European Chemical Agency (ECHA) as a substance of very high concern for its toxicological effect on reproduction and its endocrine disrupting properties. EDCs can exert their adverse effects binding directly with the ligand binding domain of NRs interfering with the normal hormone response. Thus, a lot of efforts is made to find alternative molecules that can exert the same plasticizing effects in polycarbonate materials (Fig. 6) with no or lower adverse effects for human health. The estrogenic and androgenic effects of 26 different bisphenols (including 7 BPA metabolites) have been evaluated using a mix of molecular docking and consensus scoring methods to evaluate the activity of some BPA alternatives and BPA metabolites. Six different NRs have been included in the analysis: three NRs for the estrogenic pathway and three NRs for the androgenic one. The ligand binding pockets of these NRs have different physicochemical properties. Thus, two different molecular docking software and four different scoring functions have been applied to overcome the possible limitations derived by molecular docking package and to reduce the number of false positive across different targets. The results have shown that (1) some BPA metabolites could lower the harmful effects of BPA exposure; (2) bisphenol S, a BPA' substitute, turned out a lower interactor for all NRs, except for AR, for which its binding activity is found similar to a pharmacological antiandrogen; (3) only 2,2-bis(4-hydroxyphenyl) propanol (BPAol), a BPA metabolite, was predicted as a lower interactor for all NRs considered.

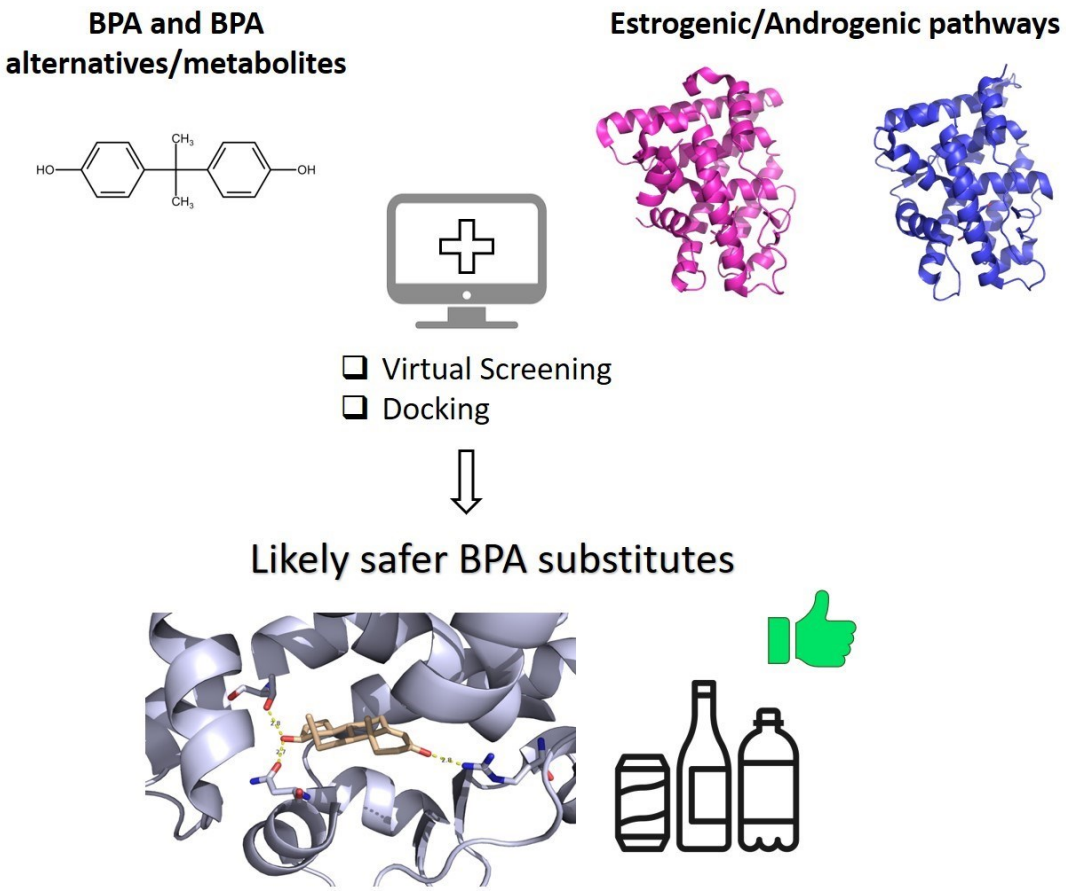


FIG. 6 The case of docking/scoring application on food contact materials: Bisphenols.

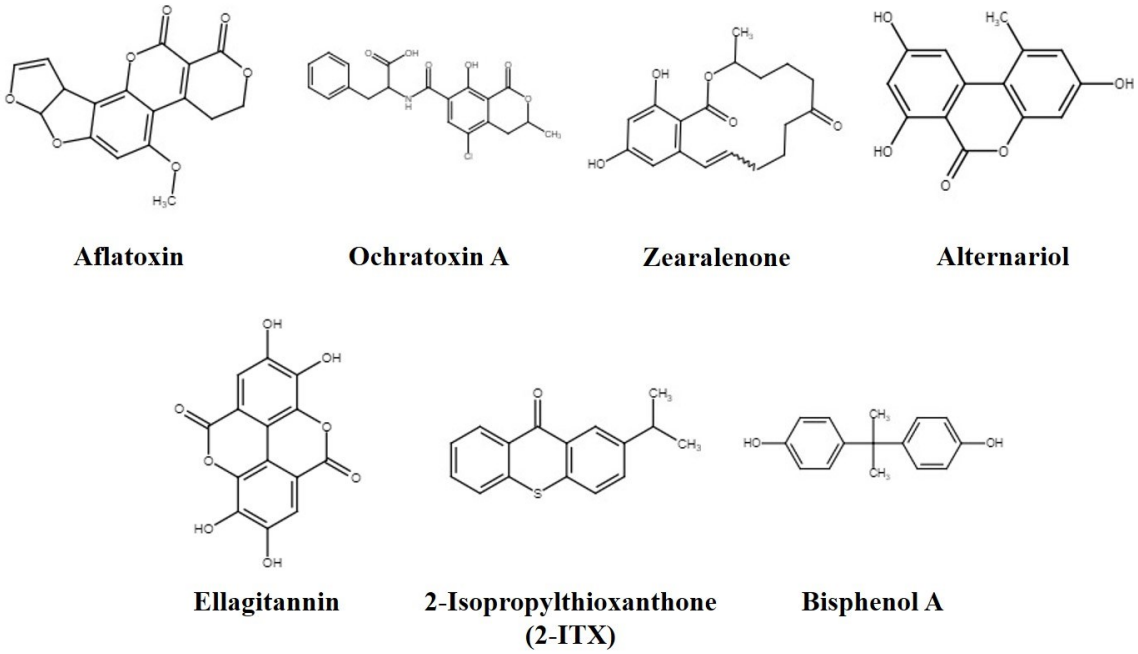


FIG. 7 The chemical structures of case studies compounds.

3.6 Conclusions

The lesson learning from medicinal chemistry suggests we can use computational simulations in food safety, in particular molecular modeling and molecular dynamics. The possibility to screen a huge number of chemicals to find endocrine disruptors in a reasonable time is, to date, a real low-cost opportunity, allowing to apply wet lab test only to the chemicals predicted as most probable interactors. From this chapter we got few take-home messages: (1) be careful with starting structural data (check the structural parameters); (2) be careful in choosing the software, there is not a general package able to solve all modeling problems; (3) a complete analysis should include a lot of factors: waters, protons, metals, cofactors, etc.; and (4) do not trust docking results blindly without a discussion; the software is not a wizard able to predict exactly the future

References

Allen, W. J., Balius, T. E., Mukherjee, S., Brozell, S. R., Moustakas, D. T., Lang, P. T., Case, D. A., Kuntz, I. D., & Rizzo, R. C. (2015). DOCK 6: Impact of new features and current docking performance. Journal of Computational Chemistry, 36(15), 1132e1156. https://doi.org/10.1002/jcc.23905

Amadasi, A., Dall'Asta, C., Ingletto, G., Pela, R., Marchelli, R., & Cozzini, P. (2007). Explaining cyclodextrin-mycotoxin interactions using a "natural" force field. Bioorganic & Medicinal Chemistry, 15(13), 4585e4594. https://doi.org/ 10.1016/j.bmc.2007.04.006

Amadasi, A., Mozzarelli, A., Meda, C., Maggi, A., & Cozzini, P. (2009). Identification of xenoestrogens in food additives by an integrated in silico and in vitro approach. Chemical Research in Toxicology, 12, 52e63.

Amadasi, A., Surface, J. A., Spyrakis, F., Cozzini, P., Mozzarelli, A., & Kellogg, G. E. (2008). Robust classification of "relevant" water molecules in putative protein binding sites. Journal of Medicinal Chemistry, 51(4), 1063e1067. https://doi.org/10.1021/jm701023h

Bissantz, C., Folkers, G., & Rognan, D. (2000). Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. Journal of Medical Chemistry, 43, 4759e4767. https://doi.org/10.1021/jm001044l

Cavaliere, F., Lorenzetti, S., & Cozzini, P. (2020). Molecular modelling methods in food safety: Bisphenols as case study. Food and Chemical Toxicology, 137. https://doi.org/ 10.1016/j.fct.2020.111116

Cozzini, P., & Dellafiora, L. (2012). In silico approach to evaluate molecular interaction between mycotoxins and the estrogen receptors ligand binding domain: A case study on zearalenone and its metabolites. Toxicology Letters. https://doi.org/10.1016/j.toxlet.2012.07.023

Cozzini, P., Fornabaio, M., Marabotti, A., Abraham, D. J., Kellogg, G. E., & Mozzarelli, A. (2004). Free energy of ligand binding to protein: Evaluation of the contribution of water molecules by computational methods. Current Medicinal Chemistry, 11(23), 3093e3118. https://doi.org/10.2174/0929867043363929

Cozzini, P., Ingletto, G., Singh, R., & Dall'Asta, C. (2008). Mycotoxin detection plays “cops and robbers”: Cyclodextrin chemosensors as specialized police? International Journal of Molecular Sciences, 9(12), 2474e2494. https://doi.org/10.3390/ijms9122474

Dellafiora, L., Dall'Asta, C., Cruciani, G., Galaverna, G., & Cozzini, P. (2015). Molecular modelling approach to evaluate poisoning of topoisomerase I by alternariol derivatives. Food Chemistry, 189, 93e101. https://doi.org/10.1016/j.foodchem.2015.02.083

Dellafiora, L., Galaverna, G., Dall'Asta, C., & Cozzini, P. (2015). Hazard identification of cis/trans-zearalenone through the looking-glass. Food and Chemical Toxicology, 86, 65e71. https://doi.org/10.1016/j.fct.2015.09.009

Dellafiora, L., Mena, P., Cozzini, P., Brighenti, F., & Del Rio, D. (2013). Modelling the possible bioactivity of ellagitanninderived metabolites. In silico tools to evaluate their potential xenoestrogenic behavior. Food & Function, 4(10), 1442e1451. https://doi.org/10.1039/c3fo60117j

Dill, K. A. (1997). Additivity principles in biochemistry. The Journal of Biological Chemistry, 272(2), 701e704. https://doi.org/10.1074/jbc.272.2.701

Ehrlich, V. A., Dellafiora, L., Mollergues, J., Asta, C. D., & Serrant, P. (2015). Hazard assessment through hybrid in vitro/in silico approach : The case of zearalenone. ALTEX: Alternativen Zu Tierexperimenten, 32(4), 275e286.

European Commission. (2014). Towards a thriving data-driven economy. Brussels: Communication from the Commission to the European Parliament, The Council, The European Economic and Social Committee and The Committee of the Regions.

Fersht, A. R. (1987). The hydrogen bond in molecular recognition. Trends in Biochemical Sciences, 12(C), 301e304. https://doi.org/10.1016/0968-0004(87)90146-0

Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw David, E., Francis, P., & Shenkin, P. S. (2004). Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. Journal of Medicinal Chemistry, 47(7), 1739e1749. https://doi.org/10.1021/jm0306430

Ginex, T., Dall'Asta, C., & Cozzini, P. (2014). Preliminary hazard evaluation of androgen receptor-mediated endocrinedisrupting effects of thioxanthone metabolites through structure-based molecular docking. Chemical Research in Toxicology, 27(2), 279e289. https://doi.org/10.1021/tx400383p

Ginex, T., Spyrakis, F., & Cozzini, P. (2014). FADB: A food additive molecular database for in silico screening in food toxicology. Food Additives & Contaminants Part A, Chemistry, Analysis, Control, Exposure & Risk Assessment, 31(5), 792e798. https://doi.org/10.1080/19440049.2014.888784

Hartung, T. (2019). Predicting toxicity of chemicals: Software beats animal testing. EFSA Journal, 17(S1), e170710. https://doi.org/10.2903/j.efsa.2019.e170710

Jain, A. N. (2003). Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. Journal of Medicinal Chemistry, 46(4), 499e511. https://doi.org/10.1021/jm020406h

Liu, J.-M., Wang, Z.-H., Ma, H., & Wang, S. (2018). Probing and quantifying the food-borne pathogens and toxins: From in vitro to in vivo. Journal of Agricultural and Food Chemistry, 66(5), 1061e1066. https://doi.org/10.1021/acs.jafc.7b05225

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). Auto- Dock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of Computational Chemistry, 30(16), 2785e2791. https://doi.org/10.1002/jcc.21256

Murcko, A., & Murcko, M. A. (1995). Computational methods to predict binding free energy in ligand-receptor complexes. Journal of Medicinal Chemistry, 38(26), 4953e4967. https://doi.org/10.1021/jm00026a001

Pagadala, N. S., Syed, K., & Tuszynski, J. (2017). Software for molecular docking: A review. Biophysical Reviews, 9(2), 91e102. https://doi.org/10.1007/s12551-016-0247-1

Richarz, A.-N. (2020). Big data in predictive toxicology: Challenges, opportunities and perspectives. In Big data in predictive toxicology (pp. 1e37). The Royal Society of Chemistry. https://doi.org/10.1039/9781782623656-00001.

Salari, R., & Chong, L. T. (2010). Desolvation costs of salt bridges across protein binding interfaces: Similarities and differences between implicit and explicit solvent models. The Journal of Physical Chemistry Letters, 1(19),2844e2848. https://doi.org/10.1021/jz1010863

Schellhammer, I., & Rarey, M. (2004). FlexX-scan: Fast, structure-based virtual screening. Proteins, 57(3), 504e517. https://doi.org/10.1002/prot.20217

Shoichet, B. K., & Kuntz, I. D. (1993). Matching chemistry and shape in molecular docking. Protein Engineering Design and Selection, 6(7), 723e732. https://doi.org/10.1093/protein/6.7.723

Spyrakis, F., Cozzini, P., & Kellogg, G. E. (2010). Docking and scoring in drug discovery. In Burger's medicinal chemistry and drug discovery (pp. 601e684). https://doi.org/10.1002/0471266949.bmc140

Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., & Taylor, R. D. (2003). Improved protein-ligand docking using GOLD. Proteins, 52(4), 609e623. https://doi.org/10.1002/prot.10465

Wang, R., & Wang, S. (2001). How does consensus scoring work for virtual library screening? An idealized computer experiment. Journal of Chemical Information and Computer Sciences, 41(5), 1422e1426. https://doi.org/10.1021/ci010025x

Williams, A. J., & Ekins, S. (2011). A quality alert and call for improved curation of public chemistry databases. Drug Discovery Today, 16(17), 747e750. https://doi.org/10.1016/j.drudis.2011.07.007

CHAPTER 4
Computational Applications on Food Contact Chemicals as Nuclear Receptor Binders

Francesca Cavaliere, Giulia Spaggiari, Pietro Cozzini

In book: Nuclear Receptors
Publisher: Springer
Editors: Mostafa Z. Badr
https://doi.org/10.1007/978-3-030-78315-0

Abstract

Humans, but also animals, are exposed to chemicals in everyday life. Many of these compounds are present in our food as food contact chemicals (FCCs), naturally occurring (toxins produced by plants), intentionally added (food additives, flavourings) or unintentionally added (pesticides, bisphenols, polychlorinated biphenyls). It is well-known that some of them can act as endocrine-disrupting chemicals (EDCs), which can interfere with the endocrine systems mainly by acting through their interaction with nuclear receptors (NRs). NRs are a superfamily constitute of 48 ligand-regulated transcription factors that are expressed in the animal kingdom and are essential for cell signalling, survival and proliferation. Thus, the alteration of nuclear receptor pathways is correlated to a large number of pathologies. Given the high number of EDCs we are exposed to, it is fundamental to test the endocrine disruptor properties of FCCs with alternative methods to animal testing. In this chapter, we focus our attention on the most common in vitro bioassays and in silico analysis as methods that can consider different endpoints of the NR pathway.

Keywords Endocrine disruptors • Food contact chemicals • Food safety prediction • *In silico* methodology • *In vitro* bioassays • Nuclear receptor-associated diseases

4.1 Introduction

Nuclear receptors (NRs) are a superfamily constitute of 48 ligand-regulated transcription factors that are expressed in the animal kingdom. NRs because of the activation of small molecules play diverse roles in cell differentiation/development, proliferation and metabolism. Nuclear receptors share a common structural organization (Fig. 1). The N-terminal region, called the A/B domain, is highly variable and contains the transcriptional activation function (AF-1) and other transactivation domains. The most conserved region is the DNA-binding domain (DBD), or C domain, which contains a P-box and two zinc fingers. The former is responsible for DNA-binding specificity, and it is involved in the dimerization of NRs, while the latter is essential for protein-protein interactions. A D-domain, localized between the DNA-binding and the ligand-binding domains, contains the nuclear localization signal. The ligand-binding domain (LBD) is the largest, and it is contained in the E-F domain, close to the carboxy terminus. The LBD, contained a conserved core of 12 α-helices (H1-H12) and two short f-sheets, is responsible for ligand recognition but also the coactivator and corepressor binding. These ligands can be classified into two different categories: (i) agonists, which promote the nuclear receptors activity, and (ii) antagonists, which block the effect of agonist through competitive interactions to the same binding site. The activation of LBD is determined by the equilibrium of different a-helix 12 (H12) conformations induced by the ligand. In fact, it rather changes the equilibrium towards more active conformations, characterized by a close H12, in the case of agonists and inactive conformations, characterized by an open H12, in the case of antagonists [8].

Depending on the structure and the ligands, the nuclear receptors could be divided into seven *subfamilies* (Table 1). The first group (*Subfamily* 0) is composed of only two proteins characterized by only a ligand-binding domain. *Subfamilies* 1 and 3 are composed of a large variety of receptors (peroxisome proliferation-activated receptors, liver X receptor, progesterone receptor, and many others) that can interact with a vastness of ligands.

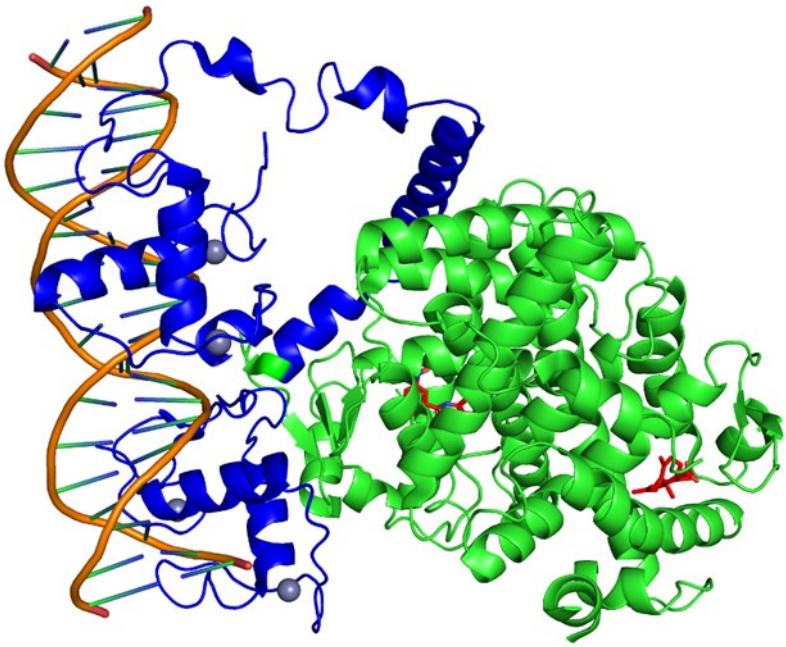


Fig. 1 Representative nuclear receptors' structure (PDB ID: 3E00). The ligand-binding site is in green (in this case it is composed by two nuclear receptors, RXRα and PPARγ, and by their respectively ligands, 9-cis-retinoic acid and 2-chloro-5-nitro-N-phenylbenzamide, in red), while the DNA-binding domain is in blue.

The following receptors' group (*Subfamily* 2) contains orphan receptors, so called because the putative ligand remains to be identified, and the retinoid X receptor (RXR), important receptor because the capability of form heterodimeric complexes with other NRs. Finally, *Subfamilies* 4, 5 and 6 contain orphan receptors important for the development and the metabolism [25].

Along with the endogenous ligands, very different types of molecules are able to bind nuclear receptors. These compounds can be divided into two groups: the first one are molecules that are synthesized to treat NR diseases (drugs) and the other one consists of unintentional binders. In this latter case, they are able to change the important biochemical pathway in which NRs are involved and they are named endocrine disruptor compounds. They include a variety of molecules such as bisphenols (BPs), mycotoxins, food additives, cosmetics, printing ink, plasticizers, etc. From a purely chemical point of view, there are no differences between these different molecules. The common result is that many of them are possible endocrine disruptors. These molecules can be present intentionally or unintentionally in our food, and in this chapter, we define this huge set of molecules simply as food contact, chemicals (FCCS).

4.2 Methods to Evaluate Food Contaminants as Nuclear Receptor Modulators

In this chapter, we want to talk about alternative animal tests that can be used to screen food contact chemicals against nuclear receptors to evaluate their endocrine disruptor proprieties, Firstly, we have considered in vitro bioassays that are currently accepted by different agencies involved in this field. Secondly, the in silico methods have been discussed to evaluate the interactions and the mechanism of action of FCCs as possible EDC molecules.

4.3 *In Vitro* Bioassays to Study the Mechanism of Action (MoA) of Endocrine Disruptor Compounds

Since the 1930s endocrine disruptor compounds have been studied using a range of in vivo models. However, since the use of animals has ethical, economic and scientific limitations, European legislation has prompted the reduction, refinement and replacement (3R) of animal experiments. Considering the higher number of compounds that are synthesized every year, it is not feasible to screen this huge amount using only in vivo study because of their high costs and low throughput. Moreover, it has been shown that a mixture of chemicals can act additively inducing a more potent endocrine-disrupting outcome [15].

Table 4.1 List of the 48 nuclear receptors with the respective diseases

No.	Subfamily	Approved Name	Gene name	Crystallography PDB structure (*)	NR diseases
1	NR0B1	Dosage-sensitive sex reversal, critical region on the X chromosome	DAX1-AHC	Yes (1)	Adrenal failure, salt-losing crises in the neonatal period, nausea, weight loss, hypotension, hyper-pigmentation, Ewing tumours typical of children, adolescents and young adults, adrenocortical tumours, ovarian, endometrial, prostate, lung and breast cancer, X-linked adrenal hypoplasia congenita (AHC), hypogonadotropic hypogonadism (HHG)
2	NR0B2	Small heterodimer partner	SHP	Yes (6)	Obesity
3	NR1A1	Thyroid hormone receptor alpha	THRA	Yes (8)	Alzheimer diseases, thyroid neoplasms, osteoporosis, cardiovascular and coronary disorders, osteoarthritis, hypertension
4	NR1A2	Thyroid hormone receptor beta	THRB	Yes (18)	Hyperthyroidism, diabetes, female infertility, end-organ unresponsiveness to thyroid hormone, abnormal growth and bone maturation, and deafness, asthma, abortion, narcolepsy
5	NR1B1	Retinoic acid receptor alpha	RARA	Yes (6)	Cleft lip and palate, diabetes, autistic and bipolar diseases, obesity, myopia, neural tube defects, neoplasms, diabetes, mental disorders, schizophrenia
6	NR1B2	Retinoic acid receptor beta	RARB	Yes (6)	Bipolar and autistic disorders, Creutzfeldt-Jakob syndrome, mental disorders, gout, diabetes, myopia, meningomyelocele, cleft lip and palate
7	NR1B3	Retinoic acid receptor gamma	RARG	Yes (11)	Diabetes, liver cirrhosis, bipolar and autistic disorders, edema, Alzheimer disease, neoplasms
8	NR1C1	Peroxisome proliferator activated receptor alpha	PPARA	Yes (18)	Dementia, coronary restenosis and stenosis, carcinoma, brain ischemia, diabetes, esophageal, lung and liver neoplasms, rhinitis, kidney failure, IGA, ventricular dysfunction, obesity, thrombosis, premature birth, ovarian and

					prostatic neoplasms, hepatitis C, hyperlipidaemia, hypercholesterolemia, chorioamnionitis
9	NR1C2	Peroxisome proliferator activated receptor delta	PPARD	Yes (41)	Bipolar disorder, adenoma, diabetes, adenocarcinoma, colonic and esophageal neoplasms, edema, hypertrophy, growth disorders, gout, weight gain and loss, multiple myeloma, personality inventory, obesity, schizophrenia, metabolic syndrome X, peripheral nervous system diseases, coronary and cardiovascular diseases
10	NR1C3	Peroxisome proliferator activated receptor gamma	PPARG	Yes (178)	Diabetes mellitus, metabolic syndrome X, cardiovascular and coronary artery diseases, colorectal and lung neoplasms, myocardial infarction, weight gain and loss, obesity, rectal and prostatic neoplasms, diseases progression, stroke, leiomyoma, atherosclerosis and arteriosclerosis, edema, pulmonary and metabolic diseases, peptic ulcer, chronic and Hodgkin diseases, dementia, sleep apnea, neoplasms, memory and mental disorders, hip fractures, lipid metabolism and growth disorders
11	NR1D1	Rev-ErbA-alpha	THRAL/ErbA	Yes (1)	Hypothyroidism congenital nongoitrous, rem sleep behaviour disorder, major depressive disorder, enhanced s-cone syndrome, delayed sleep phase disorder
12	NR1D2	Rev-Erb beta	Rev-ErbB	Yes (4)	Atrioventricular septal defect (AVSD), metabolic disorders, plasmin system abnormalities, cardiovascular diseases
13	NR1F1	RAR-related orphan receptor A	RORA	Yes (3)	Wet molecular degeneration, mental and macular disorders, edema, choroidal neovascularization, bipolar and depressive disorders, mood and sleep disorders, vasculitis
14	NR1F2	RAR-related orphan receptor B	RORB	No	Epilepsy, enhanced s-cone syndrome, refractive error
15	NR1F3	RAR-related orphan receptor C	RORG	Yes (81)	Diabetes, celiac diseases, breast neoplasms, carcinoma, lymphedema
16	NR1H2	Liver X receptor B	LXRB	Yes (17)	Calcinosis, dementia, atherosclerosis, encephalitis, edema, coronary and crohn diseases, obesity, diabetes mellitus type 2, metabolic syndrome X, neoplasms, colitis
17	NR1H3	Liver X receptor A	LXRA	Yes (7)	Cardiovascular diseases, diabetes mellitus type 2, metabolic syndrome X, polycystic ovary syndrome, coronary and cerebrovascular diseases, edema, myocardial ischemia, dyslipidaemias, hypertension, lymphoma, dementia
18	NR1H4	Farnesoid X receptor	FXR	Yes (73)	Lung neoplasms, pregnancy complications, inflammatory bowel diseases, insulin resistance, liver cirrhosis, cholestasis, colitis, coronary and crohn diseases, calcinosis, diarrhoea, Hepatitis C, dyslipidaemias, liver and cardiovascular diseases, Metabolic syndrome X, neoplasms, osteoporosis, overweight, irritable bowel syndrome, urinary bladder neoplasms, Alzheimer disease
19	NR1I1	Vitamin D receptor	VDR	Yes (45)	Periodontitis, vitamin D deficiency, diseases progression, obesity, diabetes, tuberculosis, rickets, melanoma, adenoma, prostatic hyperplasia, psoriasis, lead poisoning, carcinoma, kidney calculi
20	NR1I2	Pregnane X receptor	PXR	Yes (23)	Liver cirrhosis and neoplasms, asthma, diabetes, edema, lung and liver neoplasms, leukaemia, dementia, head and neck neoplasms, viremia, anaemia, crohn and cardiovascular diseases, carcinoma, acquired immunodeficient syndrome
21	NR1I3	Constitutive androstane receptor	CAR	Yes (2)	Renal carcinoma, neutropenia, prostatic neoplasms, memory and mental disorders, leukopenia, dementia, hypertriglyceridemia
22	NR2A1	Hepatocyte nuclear factor 4-alpha	HNF4A	Yes (5)	Hyperinsulinism, tubulointerstitial kidney disease, diabetes, Fanconi renotubular syndrome 4 with maturity-onset diabetes of the young
23	NR2A2	Hepatocyte nuclear factor 4-gamma	HNF4G	Yes (1)	Maturity-onset diabetes of the young, hyperuricemia, chromosome 8q21.11 deletion syndrome, ulcer, diabetes mellitus, colitis, crohn diseases, pancreatic neoplasm, carcinoma, inflammatory bowel diseases, dengue fever
24	NR2B1	Retinoic acid receptor alpha	RXRA	Yes (85)	Carcinoma, colonic and colorectal neoplasms, coronary stenosis and diseases, diabetes, autistic and bipolar disorders, keratoconus, microsatellite instability, neoplasms, hypercholesterolemia and hypertriglyceridemia, schizophrenia, pulmonary diseases

25	NR2B2	Retinoic acid receptor beta	RXRB	Yes (6)	Diabetes, gallstones, gallbladder neoplasms, bile duct neoplasms, arthritis, neoplasms, cryptorchidism, pulmonary diseases, psoriasis, hypospadias, lung and prostatic neoplasms, tonsillitis
26	NR2B3	Retinoic acid receptor gamma	RXRG	Yes (1)	Obesity, hypospadias, esophageal neoplasms, metabolic syndrome X, adenocarcinoma, neoplasms, autistic and bipolar disorders, diabetes, acquired immunodeficiency syndrome, carcinoma, Alzheimer diseases
27	NR2C1	Testicular receptor 2	TR2	No	Urothelial cancer, infertility
28	NR2C2	Testicular receptor 4	TR4	Yes (1)	Premature aging, lateral myocardial infarction, epilepsy, anterior cerebral artery infarction, teratocarcinoma, cancer
29	NR2E1	Tailless homolog	TLX	Yes (1)	Enhanced s-cone syndrome, retinitis pigmentosa, chromosome 17q21.31 duplication syndrome, microphthalmia, autism spectrum disorder, bipolar disorder, neurological diseases
30	NR2E3	Photoreceptor-specific nuclear receptor	PNR	Yes (1)	Enhanced s-cone syndrome, retinitis pigmentosa, colour vision deficiency, cone-rod dystrophy
31	NR2F1	Chicken ovalbumin upstream promoter transcription factor I	COUP-TFI	No	Exotropia, bosch-boonstra-schaaf optic atrophy syndrome, unilateral polymicrogyria, adrenal cortical adenoma, cerebral visual impairment
32	NR2F2	Chicken ovalbumin upstream promoter Transcription factor II	COUP-TFII	Yes (1)	Congenital heart defects multiple types 4 (CHTD4), complete atrioventricular canal-tetralogy of Fallot syndrome, complete atrioventricular canal-left heart obstruction syndrome, complete atrioventricular canal-ventricle hypoplasia syndrome, partial atrioventricular canal
33	NR2F6	V-erbA-related protein 2	EAR-2	No	Patulous eustachian tube, eustachian tube disease
34	NR3A1	Estrogen receptor alpha	ERA	Yes (266)	Osteoporosis, breast and prostatic diseases, cardiovascular diseases, female infertility, hypertension, scoliosis, uterine and colorectal neoplasms, cryptorchidism, polycystic ovary syndrome, inflammation, primary ovarian syndrome, stroke, osteoarthritis, hip fractures, leiomyoma, metabolic syndrome X
35	NR3A2	Estrogen receptor beta	ERB	Yes (32)	Cardiovascular diseases, beast, colorectal, and prostatic neoplasms, osteoporosis, endometriosis, male and female infertility, hypertension, oligospermia, obesity, Parkinson and Alzheimer diseases, azoospermia, adenocarcinoma, ovarian and testicular neoplasms, colonic and endometrial neoplasms, gallstones, hip fractures, anorexia nervosa, abortion, inflammation
36	NR3B1	Estrogen related receptor alpha	ESRRA	Yes (4)	Diabetes Mellitus type 2, cardiovascular diseases, edema, glandular and epithelial neoplasms, obesity, ovarian neoplasm
37	NR3B2	Estrogen related receptor beta	ESRRB	No	Deafness autosomal recessive, adrenal hypoplasia, hereditary hearing loss and deafness, autosomal recessive non-syndromic sensorineural deafness type DFNB
38	NR3B3	Estrogen related receptor gamma	ESRRG	Yes (17)	Breast and colorectal neoplasms, Diabetes Mellitus type 2, hearing, neoplasms, osteoporosis, overweight, stomach neoplasms
39	NR3C1	Glucocorticoid receptor	GR	Yes (43)	Obesity, bipolar disorder, bronchiolitis, cardiovascular and coronary diseases, diabetes mellitus, hypertension, inflammation, fatigue syndrome, metabolic syndrome X, adenoma, premature birth, schizophrenia, mental and psychotic disorders, multiple sclerosis
40	NR3C2	Mineralocorticoid receptor	MRL	Yes (25)	Bipolar and attention deficit disorders, myocardial infarction, reward, stress and mental disorders, metabolic syndrome X, hypotension, hyperkalaemia, child behaviour disorders, pseudo hypoaldosteronism, pregnancy complications, edema
41	NR3C3	Progesterone receptor	PGR	Yes (20)	Breast and ovarian neoplasms, premature birth, uterine and prostatic neoplasms, male and female infertility, carcinoma, gallstones, abortion, musculoskeletal diseases, neoplasms, vertigo, thrombophilia, skin and pulmonary diseases, obesity
42	NR3C4	Androgen receptor	AR	Yes (82)	Prostatic and breast neoplasms, infertility, male, polycystic ovary syndrome, alopecia, ovarian neoplasms, oligospermia, prostatic hyperplasia, testicular neoplasms, disease progression,

					endometrial, neoplasms, carcinoma, cryptorchidism, insulin resistance, neoplasms, hypospadias, hypogonadism, Klinefelter syndrome, diabetes mellitus type 2, adenocarcinoma, acne, androgen-insensitivity syndrome, obesity, azoospermia, cardiovascular diseases, Alzheimer disease, leiomyoma hyperandrogenism, osteoporosis, ovarian failure, gender identity, metabolic Syndrome X, abortion, autistic disorder, chromosome aberrations, depressive disorder, endometriosis
43	NR4A1	Nerve growth factor IB-like receptor	NGF IB	Yes (15)	Pseudohypoaldosteronism, salivary gland carcinoma, pyomyositis, night blindness congenital stationary type 1h, salivary gland disease, metabolic disease, colorectal and pancreatic cancer, lung and breast cancer, inflammatory disease
44	NR4A2	NGFI-B/nur77 beta type transcription factor homolog	NURR1	Yes (2)	Parkinson disease, arthritis, rheumatoid arthritis, attention deficit-hyperactivity disorder, alcohol dependence, colorectal, lung, adrenocortical and cervical cancer
45	NR4A3	Neuron-derived orphan receptor 1	NOR1	No	Chondrosarcoma, epithelial-myoepithelial carcinoma, myxoid and extraosseous chondrosarcoma, Ewing sarcoma (ES)
46	NR5A1	Steroidogenic factor 1	STF1	Yes (4)	46,XY sex reversal 3 (SRXY3), 46,XX sex reversal 4 (SRXX4), premature ovarian failure 7 (POF7), spermatogenic failure 8 (SPGF8), adrenal insufficiency NR5A1-related (AINR), prostate cancer
47	NR5A2	Liver receptor homolog-1	LRH1	Yes (17)	Edema, diarrhoea, obesity, osteoporosis, irritable bowel syndrome, adenocarcinoma, cardiovascular diseases, diabetes mellitus type 2
48	NR6A1	Germ cell nuclear facto	GCNF	No	Embryonal carcinoma, teratocarcinoma, ureter cancer, retinitis pigmentosa

Thus, human exposure to different contaminants could also have an additive effect. Although in vivo studies have their advantages, the complexity of the biological system often generates controversial results: for example, where the same experiment was possible between humans and animals, a correlation of 60% has been found [14]. Thus, the results of animal tests cannot be always related to the human outcomes. Keeping this in mind, in silico and in vitro analysis should be preferred over in vivo studies. This section does not want to be an exhaustive list of in vitro studies to screen endocrine disruptor compounds, but we want to make a brief discussion about the most accepted and used in vitro tests considering the guidelines of agencies that are most active in this field: US Environmental Protection Agency (EPA) (https://www.epa.gov/), European Chemicals Agency (ECHA) (https://echa.europa.eu/it/home), Organisation for Economic Co-operation and Development (OECD) (https://www.oecd.org/) and European Food Safety Authority (EFSA) (https://www.efsa.europa.eu/it). Chemicals intentionally or unintentionally coming in contact with food can adverse human health, in most cases functioning as endocrine disruptor compounds. To evaluate their potential ED proprieties, different in vitro tests are currently used. To understand how they function, a short discussion about how NRs work should be made. Nuclear receptors are composed of two principal domains: a DNA-binding domain and a ligand-binding domain. When an

agonist ligand binds to the ligand-binding pocket of LBD, the receptor (as a monomer, homodimer or heterodimer) can migrate inside the nucleus, where the DBD recognized specific DNA sequence named DNA-responsive element located upstream to the gene regulated by the receptor. Once the NR is bound to the DNA, it recruits additional proteins of the transcriptional machinery and activates the transcription and transduction of the gene. Thus, an endocrine disruptor is a compound able to bind the nuclear receptor inducing its activation or deactivation. As a consequence, it determines an upregulation or downregulation of the genes that the receptor modulates. The endocrine-disrupting issue is not a recent discussion. In 1998, EPA convened a committee for developing a tiered approach to evaluate the oestrogen, androgen and thyroid-related effects of a great number of chemical contaminants for a rapid prioritization following by in vivo tests on only relevant compounds. After that, in 2012, the OECD has released a revised guidance document in which test guidelines are exposed for evaluating chemicals for endocrine disruption, and that has been updated in 2018. In vitro assays are part of the Level 2 Framework of OECD, and most of them refer to the oestrogenic and androgenic pathway as well as steroidogenesis (Fig. 2). However, compounds could also interfere with other nuclear receptors/pathways, and thus additional in vitro assays are required to detect all endocrine activity. Operating with the same principle, most of the in vitro tests cited in the OECD document are also available for other NRs. Since the scope of this section should be to make a brief discussion about in vitro tests that could be useful to validate in silico methodologies and that are used for studying EDCs, we have categorized them according to the biological endpoint under investigation. Accordingly, since an EDC could act at different levels of the biological systems and induce different responses, we discuss bioassays considering the effect resulting from EDC exposure: the chemical interaction with hormone receptors, the induced gene expression by the ligand binding to the receptor, and the cellular responses to EDCs (Table 2).

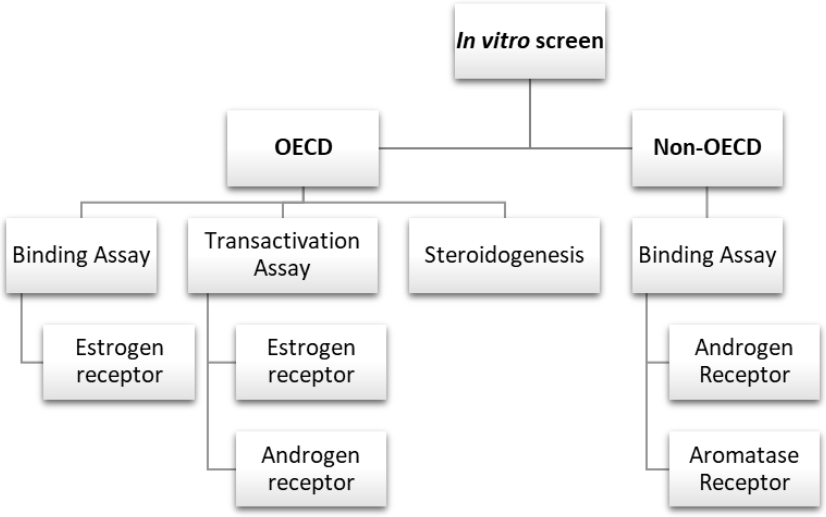


Fig. 2 Flowchart of in vitro approaches

Table 2 List of the in vitro tests based on the biological endpoint under investigation

Level of response	Type of bioassay	Mechanism	Endpoint
Receptor	Ligand binding assays	Detect the direct ligand binding to nuclear receptors	Receptor binding
	Isothermal Titration Calorimetry		
	Differential Scanning Fluorimetry		
Transcription	Reporter gene assays	Detect the agonistic/antagonistic effects	Receptor transactivation
Steroidogenesis	H295R assay	Detect metabolic activation induced by EDCs	Hormone Production
	Aromatase assay		

4.3.1 Ligand-Binding Assays

OECD guidelines and EPA documents refer to only the oestrogenic/androgenic binding assay. The first in vitro test to screen the capability of a compound to bind a nuclear receptor is the ligand-binding assay. The older version of the test, applied to the oestrogen receptor, used rat uterine cytosol as a source of the oestrogen receptor protein. Since there are two different isoforms (alpha and beta) of the receptor, the test does not make any distinction between them (although the alpha isoform is the most abundant protein in rat uterine cytosol). Thus, a homogenates tissue extract can be used that must be specific for the receptor under investigation: that is, it is advisable to use a cell line derived from a specific tissue that expresses the nuclear receptor at higher levels. However, a modern

binding assay method uses a recombinant protein of the human receptor produced in and isolated from baculovirus-infected insect cells expressing a full-length human recombinant protein or the human recombinant ligand-binding domain only. Homogenates of cells or tissues with a radiolabelled or fluorescent (fluorescent polarization binding assay) compound are incubated together with different concentrations of tested compounds. Plotting the bound reference ligand against the log concentration of the tested compound gives the possibility to generate the competitive binding curve. The binding activity is quantified as the concentration of the competitor needed to displace half of the reference compound (IC50) or as relative binding affinity calculated from the ratio between the IC50 of the reference compound and the test chemical. Thus, the binding assay cannot determine whether a compound is an agonist or an antagonist since it does not consider the transcriptional activity of the receptor, but it only divides compounds into binders (strong to weak) and non- binders. As an advantage, since the ligand-binding domain of some nuclear receptors is highly conserved across different vertebrate species, the assay results could be referred to many taxa. Among in vitro tests used, this represents the best method for correlating in silico results since it refers to the direct binding of a compound to the receptor. However, the great advantage of in silico methods could be evident: although the rationale of binding assay could be extended to other nuclear receptors, the document of 2018 guidelines is only provided for oestrogen-, androgen- and aromatase-receptor. The great versatility of in silico methods lies in the possibility to consider different NRs all at once to predict the direct binding of chemical contaminants to receptors. This can be done in a very fast and inexpensive way compared to the binding assay. They do not require cell lines; no solutions and no compounds are needed for the experimentation. Moreover, there are fewer interfering factors compared to cell lines, where it has been reported that a certain grade of variability exists in the assay results influenced by protein concentration and/or plate temperature. The phenomenon of partial degradation and/or denaturation of the protein could influence the ligand-receptor interaction inducing a reported decrease of the binding that it is, in reality, a false-positive result. Another important issue can be encountered that when being tested compounds that themselves fluoresce or interfere with light emission report an erroneous interaction. However, although LBA is the only direct binding assay in the OECD guidelines and the only one accepted by EPA, which refers to OECD documents, in literature some other in vitro methodologies have been used to evaluate the direct

endocrine disruptor binding with NRs. Isothermal titration calorimetry (ITC) assay, for example, is often used to study the binding of a small molecule to large macromolecules, such as proteins. It directly measures the heat realized or absorbed along with a bimolecular reaction depending on the type of binding, i.e. whether exothermic or endothermic. The instrument is composed of two different cells that are kept at steady temperature and pressure: (i) the main cell where the NR ligand- binding domain is placed in its buffer solution and (ii) the reference cell which is generally filled with water or with the solvent used for the analysis. During the experiment, the tested compound is titrated into the receptor solution (main cell). Since the reaction leads to a heat release or consumption, the binding induces a variation in the temperature of the main cell. For maintaining it at the same temperature as the reference cell, the instrument spends energy. The heat change is, calculated by integrating the power spent over the time (seconds) that corresponds to the enthalpy of the reaction and, thus, to the fraction of bound ligand. For instance, Zhang and colleagues have used an in silico approach to screen indoor dust contaminants against thyroid hormone receptor BI (THRβ1) [27]. Of the 31 compounds predicted as potential (THRβ) binders, 5 have been tested using ITC, and the binding affinity has been calculated. The results showed that four of five molecules were THRβ1 binders. ITC is often useful when the synergic effect of compounds would be studied. Balaguer and colleagues used ITC to study the cocktail effects of two molecules alone and in combination against the peroxisome X receptor (PXR) reporting that the two compounds can interact contemporarily with the nuclear receptor [2]. Thouennon and co-workers used the ITC to characterize the ability of some environmental chemical contaminants to bind oestrogen-related receptor y (ERRy) finding that bisphenol E was a more potent binder compared to bisphenol A [22]. An additional in vitro technique that exploits protein thermodynamic characteristic to study ligand-protein binding is the differential scanning fluorimetry (DSF), also known as thermal shift assay (TSA) or thermal denaturation assay (TDA). The methodology is based on the principle that bounded nuclear receptors are more stable than the apo-form and thus are much less prone to denaturation process induced by the heating temperature. DSF uses a real-time PCR instrument to monitor thermally induced denaturation of protein at different ligand concentrations by measuring the fluorescence of a dye that binds preferentially unfolded proteins. Compounds that significantly increase the protein Tm as compared to the vehicle controls are good binders of the nuclear receptor. Since the magnitude of ΔTm is

negatively correlated to K_d of the interaction, DSF allows obtaining the binding affinity of different compounds. DeSantis and colleagues have reported the capability of this technique to identify known interactors of ERα contained in a commercially available compound library, showing that the two agonists β-estradiol and estrone and the antagonist tamoxifen citrate can increase significantly the Tm of the receptor compared to the control sample [11].

4.3.2 Gene Reporter Assays

If the experiment's purpose is to distinguish against agonist and antagonist com- pounds, reporter gene or gene transactivation assay can be applied. Monitoring the transcriptional levels of downstream genes is an efficient in vitro test to screen endocrine-disrupting properties of food contact chemicals. Cell cultures are co-transfected with two plasmids: the first one containing the genomic sequence of a nuclear receptor and the second one reporting the specific DNA-responsive element fused with the genomic sequence of a product that can easily be quantified (e.g. luciferase, a fluorescent protein or B-galactosidase). Cells are treated with tested compounds, and the agonistic activity could be detected by monitoring the NR-mediated transactivation of the reported gene compared to control cells (normally treated with the vehicle alone). The antagonistic activity of a compound can be instead detected co-treating cells with a chemical and a potent agonist to establish whether it determines a reduction in response and data are compared to cells treated with the potent agonist alone. Finally, if the compound is not able to bind the nuclear receptor and/or induce an agonistic or antagonistic activity, no differences will be reported in the transcription of the reported gene in both experiments. One of the first versions of this assay utilized yeast cell lines carrying the human nuclear receptor together with a vector containing the reported gene, and it is widely used for screening environmental samples. Actually, more specific human mammalian cell lines could be used. They could be properly selected for the type of nuclear receptor under investigation, i.e. cell lines that are well-known to express at high dose the NR. In this latter case, cells are only transfected with report gene construct using selected mammalian cells that naturally express the receptor of interest. Alternative, dual receptor-reporter transfections are also common for mammalian endocrine-screening assays. However, reducing performances could be encountered due to the transcriptional activation of the reporter gene construct induced by non-ER or non-AR-mediated process. To solve this issue, a chimeric construct is utilized in some cases that involve the use of the human

ligand binding fused with the DBD of a yeast-specific protein. Importantly, this in vitro test has significant interlaboratory variability, in part influenced by assay parameters such as pH and solvent effect. In silico methods can be, in some circumstances, compared to gene transactivation assay result. Generally speaking, molecular docking allows predicting if a compound is a good, a weak or a bad binder of the receptor since it predicts the binding strength of a protein-ligand interaction without considering the effect of this interaction in terms of agonistic and/or antagonistic activity of a compound. However, for some NRs, such as the oestrogen receptor, two different protein conformations are well-known differing for the helix 12 positions: a close (agonist) conformation where the H12 is located towards the receptor and an open (antagonist) conformation, where H12 is displaced from the receptor. Taking in consideration both the receptor conformations during molecular docking, screening allows to distinguish towards agonist and antagonist compounds: if a compound has ahigh score in the agonist conformation and not in the antagonist conformation, it can be speculated that it could act as an agonist compound; on the other side, if a molecule has a higher score in the open conformation compared to the agonistic one, it could probably act as an antagonist. Although not included in the OECD. guidelines, additional in vitro tests could be performed to analyse the capability of a compound to interfere with the endocrine system. The effects of EDCs on the expression of NR target genes can be also examined using real-time PCR (RT-PCR). Cell lines expressing the nuclear receptor under investigation are treated with different concentrations of a tested compound and incubated for a variable period of time. Total RNA content is then extracted, and the mRNA of specific genes transcribed by the NR is converted into cDNA. Different techniques could be used in this passage, but generally they allow to detect the mRNA conversion in real time. Since the conversion is a linear reaction, the methodology allows us to quantify the expression levels of mRNA transcribed and thus the ability of a compound to induce the nuclear receptor activation. For example, Dellafiora and colleagues have used quantitative RT-PCR to measure the transcriptional activity of oestrogen receptor- controlled genes (GREB1, growth receptor by oestrogen in breast cancer 1; PR, progesterone receptor) induced by two mycotoxin compounds, the well-known xenoestrogenic zearalenone (ZEN) and zearalenone-14-glucoside (ZEN14Glc), a metabolite produced by plants and is present in food intended for human and animal consumption {10}. They have found that ZEN14Glc can induce a more potent activation of ER target genes and thus supposedly a more potent oestrogenic

interference. The same experiment has been used by Yin and co-workers for evaluating different probable EDC compounds for their capability to activate the estrogenic activity showing that bisphenol A and bisphenol AF consistently can activate endogenous ER target genes [17].

4.3.3 Steroidogenesis Assay

Endocrine disruptor compounds can also affect steroid biosynthesis influencing the NR activity as an indirect effect. A range of in vitro models for steroidogenesis is available, and the H295R assay is the one accepted by OECD (OECD TG 456) and also included in the EPA Endocrine Disruptor Screening Program (EPA 640-C-09-003). The human adenocarcinoma H295R cell line expresses all enzymes needed to convert cholesterol to the key steroids. However, although the interaction of EDCs with steroidogenesis proteins can influence the production of different sex steroids such as oestrogens and androgens as well as progesterone, glucocorticoids and aldosterone, the assay was validated only to detect (estosterone and estradiol. In brief, H29SR cells are exposed to seven concentrations of the tested compound in at least triplicate for 48-72 h. At the end of the exposure period, the concentration of hormones secreted into the medium can be measured using a variety of methods, such as radioimmunoassay, ELISA (enzyme-linked immunosorbent assay) or chemical analysis. The results are expressed as fold changes in hormone concentration compared with the negative control. Chemicals that may induce steroidogenesis increase the production of estradiol and testosterone; rather, chemicals that inhibit the steroidogenesis decrease the concentration of the two hormones. However, the test does not provide specific information concerning the interaction of the test substance with the endocrine pathway, and thus the results cannot be correlated with in silico studies. Additionally, aromatase assay can be chemicals that may affect the endocrine system (e.g. steroidogenesis) by inhibiting the catalytic activity of aromatase, the enzyme responsible for the conversion of androgens to oestrogens. It is included in the EPA's EDSP Tier I screening protocol (EPA. 740-C-09-004), Human recombinant microsomes are incubated with radiolabelled androstenedione [3H]ASDN, an aromatase substrate, and an essential cofactor (NADPH) for the aromatase activity together with increased concentration of the tested compound. The rate of tritiated water (3H2O) released during the conversion of [3H] ASDN to estrone is quantified, and it is influenced by the activity of aromatase. If a chemical is able to interact and inhibit the enzyme binding to the binding pocket of the androstenedione, a decrease in the tritiated water

(3H2O) is reported. Thus, plotting the production of 3H2O as a percent of the solvent control versus the log of the concentration of the test chemical, it is possible to obtain the response curve that allows classifying a compound as an aromatase inhibitor or non-inhibitor.

Although in vitro studies are common usage for screening endocrine disruptor compounds, the huge amount of food contact chemicals highlights the importance of alternative methods (in silico) that can predict EDCs in a faster, safer and better way.

4.4 In Silico Methods for Screening Endocrine Disruptor Compounds

4.4.1 3D Protein Structure: The Starting Point of Computational Methods

Currently, over 700 nuclear receptors' structures have been solved using X-ray crystallography or NMR spectroscopy. When a structure is solved, it is deposited in various structural databases, such as PDB. This database, called Protein Data Bank (PDB), contains the experimental data of the protein structures. In the PDB database, protein 3D structures are represented as a set of coordinate triplets (x, y and z) that define the position of protein atoms. The quality of the PDB structure is defined by two parameters: the resolution (Å) value and the B-factor value.

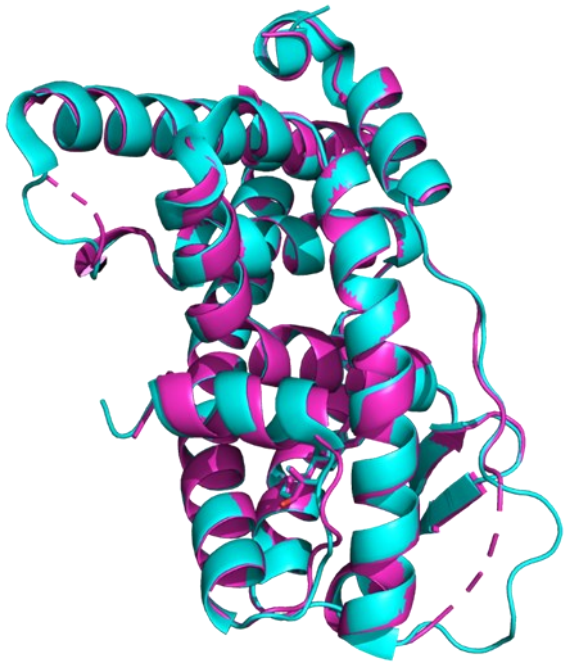


Fig. 3 PDB structures of oestrogen receptor alpha with two different resolution values. *In blue* the protein (PDB ID: 2YJA) with a high-resolution value (1.82 Å resolution) and *in magenta* the protein (PDB ID: 1ERE) with a low-resolution value (3.10 Å resolution) are shown. The box highlights the part of the protein resolute in 2YJA

The resolution value is influenced by how well the crystal diffracts and by the amount of time needed to collect resolution data. When a structure has a high resolution, the value is around 1 Å, whereas when it has a lower resolution is around 3 Å and above (Fig. 3).
The B-factor monitors the oscillation amplitudes of the protein atoms around their equilibrium positions, or it can be defined as a probability density function for the location of each atom in the protein [7]. The B-factor is defined according to the following equation:

$$B = 8\pi^2(u^2)$$

where u is the mean displacement of a scattering centre, measured in Angstroms, and it is an isotropic displacement parameter associated with the reference atom. Usually, an isotropic model is used to model protein motion characterized by a low resolution and a spherical shape, while an anisotropic or ellipsoid model is used to describe the protein motion of small organic crystals (Fig. 4) [23]. The latter provides both the magnitudes and the directions of each atom shift, and, thus, it allows a dynamic description of the protein structure.
However, as shown in Table 1, not all the nuclear receptors' structures are crystallized. In fact, the major limitation of the X-ray crystallography technique is that the molecules under study must be able to adopt sufficient compact and rigid structures to pack and form a crystal. Instead, nuclear receptors are very complex, both for their flexibility, characterized by an essential biological conformational transition under relatively mild conditions in a wide range of time and space scales, and for the millions mechanism of action given from the relationship between the receptor conformation and the ligand binding. Moreover, some of the structures of the nuclear receptors are unknown, both for the flexibility and the plasticity of the system than for the expenses, labour and time of the procedure. These gaps can be filled in by computational techniques, in particular, due to the use of homology modelling. Homology modelling is the most common and used techniques fundamental to predict the 3D structure of proteins. The basic principle of homology modelling is that proteins with similar sequences may display common structural features. It is for this reason that the accuracy of 3D structures obtained is highly dependent on the sequence identity to the reference structural models.

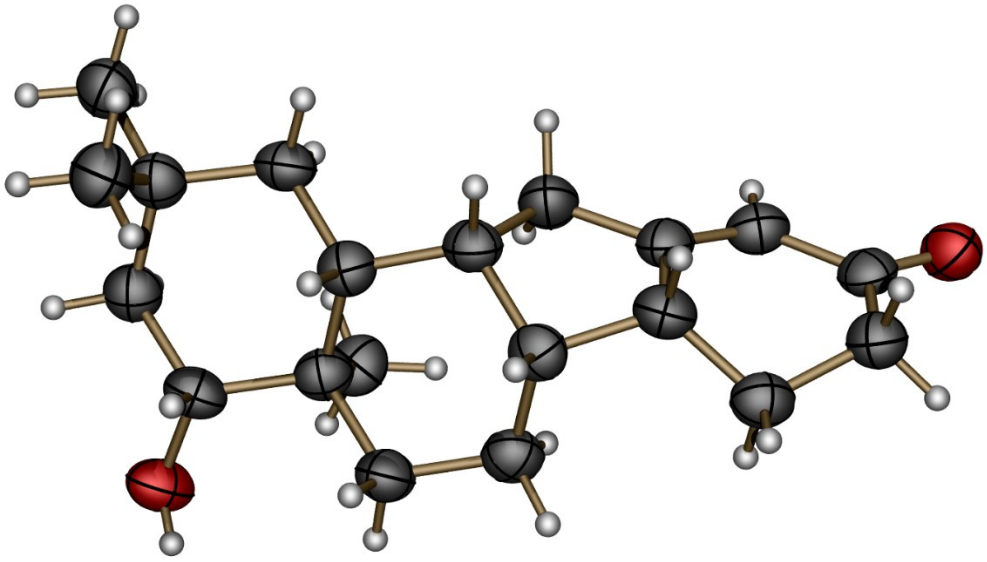


Fig. 4 The epitestosterone shown above, made by the program ORTEP [12], illustrates the thermal ellipsoid

In silico methods (Fig. 5) are widely used in the fields of computational chemistry, computational biology and material sciences to study molecular systems, ranging from a small system to large biological molecules. Virtual screening is a powerful tool to predict the activity of a huge number of chemicals in a reasonable time. Several databases of molecules are currently available for virtual screening campaigns, such as ZINC, a free database of commercially available compounds; ChEMBL, a database of bioactive molecules; and PubChem (https://pubchem.ncbi.nlm.nih.gov/), a database of chemical information [9, 18, 21]. Virtual screening approaches can be divided into ligand-based when the information of known ligands is used, and structure-based, when the information of the targeted protein-binding site is used.

The increasing number of chemical product synthesized and released in the market every year has necessitated the development of computational approaches to speed up the process of their food safety and security. Although the usage of computational approaches was started from the drug discovery field with the aim to identify new potential drug candidates, in recent years, the usage of virtual screening is becoming more important in the food risk assessment area, too. This is because on the molecular scale, interaction is an interaction, and thus from a chemical point of view, it is not important if a compound is a drug or a food contact chemical (FCC). Thus, in silico methods can be easily moved in the food safety field to screen the capability of FCCs to interact with target proteins interfering with their natural biological activity. In silico screening techniques of a large compound databases are commonly defined as virtual screening (VS), referring to those

computing techniques that use a complementary tool to identify potential binder compounds on a pool of chemicals.

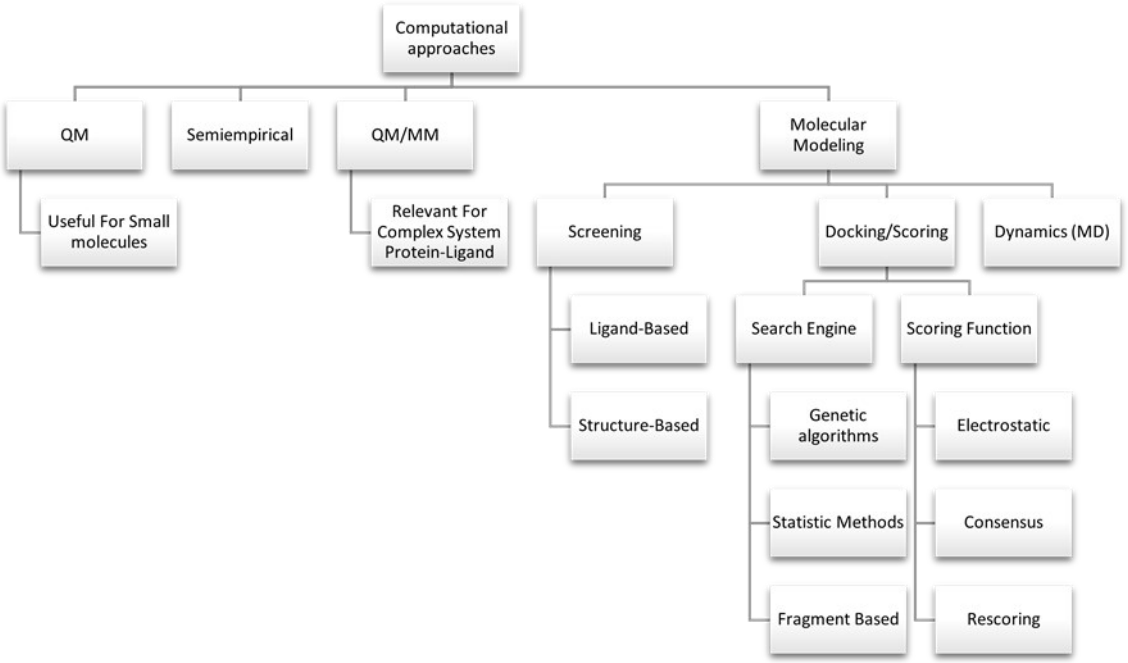


Fig. 5 Flowchart of computational (or in silico) approaches

Like high-throughput screenings (HTS), VS is used as a first step to process large libraries of compounds. The main advantage of VS compared to HTS is the rapidity of the screening method and the decreasing costs since it does not require compounds to be synthesized or purchased and tested. Since virtual screening methodologies are knowledge-based approaches, they require structural information about the binding site and/or the nature of ligand that should bind. Thus, based on the available information, virtual screening can be divided into ligand-based virtual screening and structure-based virtual screening. If the three-dimensional structure of target binding site is unknown, ligand-based virtual screening (LBVS) can be used since it faces the problem by the ligand point of view. In fact, based on known active molecules, this methodology searches for similar compounds. Ligand-based methods consider molecule dimensionality, with 1D or 2D methods being considered separately from 3D methods. The former searches for molecules' numerical descriptors that are independent by their molecular structure to attempt to relate them with their known biological activity, and they are mainly described as quantitative structure-activity relationship (QSAR). Instead, three-dimensional (3D) LBVS methods incorporate the molecular conformation and can be mainly divided in

subgroups based on the method used for the similarity search: (i) pharmacophore-based, (ii) shape-based, (iii) molecular field-based methods, (iv) fingerprint-based methods and (v) electrostatic potential similarity.

The concept of pharmacophore was introduced by Ehrlich in the nineteenth century based on the idea that specific groups within a molecule are responsible for its biological activity. The pharmacophore concept was developed over time reaching the modern IUPAC (International Union of Pure and Applied Chemistry) definition: "a pharmacophore is the ensemble of steric and electronic feature that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response" [26]. Thus, it describes the essential features that a molecule should have for the binding with the target protein and does not represent a real molecule. The three-dimensional (3D) structure of known active molecules is superimposed considering shared pharmacophore features in order to identify key interaction points for building a skeleton of abstract characteristics that define interaction type, such as hydrophobic and aromatic contacts, hydrogen-bond donors and acceptors and charged interactions.

Shape-based strategies, such as ROCS and shape screening (Schrodinger), are based on the concept that if a molecule has an overall similarity shape with a known binder, then it is likely to fit in the same binding pocket [19, 20]. Thus, they compare atomic radii instead of atom types and do not consider particular properties of the reference ligands. Therefore, shape-based methods are often used in combination with other approaches which consider some chemical properties.

Molecular field-based or grid-based methods, such as CoMFA (comparative molecular-field analysis), CoMSIA (comparative molecular similarity index analysis) and GOLPE based on GRID compared to molecules aligning the dataset compounds using different rules. Steric, electrostatic or hydrophobic potential fields (but also can be included hydrogen-bond donors and acceptor descriptors) are calculated at each grid point using a probe atom for identifying the similarity between the molecules [3].

Fingerprint-based methods are based on the concept to reduce the complexity of the molecular representation considering molecules as a sequence of bits which can then be easily compared. The similarity is then calculated using Euclidean distance or most commonly the Tanimoto coefficient. According to the nature of bits, fingerprints can be classified as sub-structure key-based, topological or path-based, circular and

pharmacophore fingerprints [6, 13]. Finally, since electrostatic interactions often play a critical role in ligand binding, another approach of LBVS uses the electrostatic potential of a reference ligand to collect compounds that have similar electrostatic distribution [13].

4.4.3 Molecular Docking

Two molecules can interact in several ways let alone the interaction of a protein and protein/small molecules. Molecular docking is a computational technique that involves finding the most favourable binding mode of a ligand to the target protein. First of all, to have an accurate docking prediction, a high resolution X-ray, NMR or homology-modelled structure is necessary. Molecular docking can be achieved through two steps: (i) the different conformations' prediction of the ligand in the active site of the protein and (ii) the conformations ranked via a scoring function. There are a huge number of binding modes between two molecules. For this reason, various sampling algorithms have been developed in molecular docking software (Table 3). These algorithms should be able to reproduce the experimental binding mode between two molecules.

As mentioned before, the nuclear receptors are flexible and plastic systems. The protein may adopt different conformations in the unbound and bound states and may adopt different conformations with different ligands. For these reasons, molecular docking methods can be divided into rigid docking where the bond angles, the bond lengths and the torsion angles of the ligand and the protein are not modified and flexible docking that permits conformation changes. The flexibility could be applied to the ligand and/or to the protein. If the flexibility is imposed on the ligand, it can be able to explore all the conformational space of the protein. The ligand flexibility is commonly considered in docking simulations, while the protein flexibility still remains a challenging goal, mainly because of the dynamic complexity and of the computational time required for running the simulations. A considerable option is to impose the flexibility only to a region of the protein. Then, a limited number of atoms are considered, for example, the pocket side chains.

Obtaining a huge number of ligands binding mode, scoring functions are fundamental to estimate and calculate the ligand-binding affinity between the protein and the ligand, to delineate the correct poses from incorrect poses. Two main aspects characterize a docking simulation and influence its results: (i) a search engine that defines the sampled conformational space and (ii) an empirical scoring function that is used to approximately predict the ligand-protein binding affinity and, in a virtual screening campaign, is used as

a measure to rank screened compounds. Scoring functions can be divided into force field-based, empirical and knowledge-based scoring functions. The first estimates the binding energy calculating the sum of the non-bonded interactions. The basis of the second scoring function is that the binding energies of the complex can be approximated by the sum of individual energy components: hydrogen bond, ionic interaction, hydrophobic bond and binding entropy. The knowledge-based scoring function uses statistical analysis of the ligand-protein complex to obtain the interatomic contact frequencies and/or a distance between the two components. As a technique that aims to furnish a quick result for the analysis of a complex biological process, the molecular docking has some limitations: (i) scoring functions are very sensitive to ligand size and are implemented mainly considering electrostatic contributions and underestimating the hydrophobic effect, and (ii) a docking simulation can be performed only between two molecules per time; it cannot predict the effect of water molecules and/or cofactors to the ligand binding. In such case studies, where the role and the position of a water molecule are well established, the water molecule can be explicitly considered even for docking simulations. However, it is challenging to determine the effect of waters in the binding when the experimental structure is not available. In order to deeply rationalize the ligand-protein binding process, molecular dynamics simulations can be used.

Table 3 The most used molecular docking programs with the respective algorithms

Molecular docking program	Algorithm
DOCK, LibDock	Matching algorithm (MA)
$DOCK_{4.0}$, SLIDE, FlexX	Incremental construction (IC)
AutoDock, $DockVision_{1.0.3}$	Monte Carlo (MC) technique
GOLD, FLIPDock	Genetic algorithm (GA)
GLIDE	Hierarchical method

The MA based on molecular shape maps a ligand into an active site of a protein in terms of shape features and chemical information. The IC fragments the ligand from rotable bonds into various segments. The MC modifies gradually the ligand using bond rotation and translation or rotation of the entire ligand. The GA is similar to the MC method, but it is used to find the global minima. The hierarchical method precomputes and aligns the low energy of ligand

4.4.4 Consensus Scoring

A solution to overcome the intrinsic limitations of a specific docking/scoring software is the consensus scoring. Because of any embedded force field used to score the docked solution that is intrinsically linked to the searching engine (the algorithm used to search

the possible positions of a ligand within a receptor cavity), a solution for a more reliable result is to use more than one package or more than one evaluation function. This is in order to achieve a “convergence”, a “consensus” to the best possible solution. We have three possible approaches: (i) one package with a different internal scoring function, not a great solution because the newest scoring function is, in general, an updated version of the previous one, and then it works better; (ii) two or more packages with the internal scoring function; and (iii) more packages with their internal scoring function plus a rescoring using one or more external independent scoring functions. Compared to a single scoring function, Wang and Wang have reported that using different scoring functions can reduce false positives and improve hit rates [24]. Moreover, Bissantz and co-workers have highlighted that using three different scoring functions allows to reduce the number of false positives and enhance the capability to reach hit rates from 10% up to 65–70% [4].

4.4.5 Molecular Dynamic Simulations

The power of the existing supercomputers allows us to carry out microsecond-scale MD simulations in a few days or a week depending on the architecture of the system. The atoms in a biomolecule are in constant motion, and both the molecular functions and the intermolecular interactions depend on the dynamics of the molecules involved. Molecular dynamic (MD) simulation is a computational technique used for analysing the physical movements of atoms and molecules and for investigating the structure, dynamics and thermodynamics of biological systems with the use of computer. The molecular dynamic simulation is based on Newton’s second law or the equation of motion, F = ma (F is the force exerted on the particle, m is the mass and a is the acceleration). From a knowledge of the force on each atom, it is possible to determine the acceleration of each atom in the system. Integration of the equations of motion then yields a trajectory that describes the positions, the velocities and the accelerations of the particles as they vary with time. From the trajectory, the average values of properties can be determined. MD trajectories provide a view of the motion of a molecular system in a time-space, allowing to consider the macro flexibility and the influence of the solvent. Water molecules solvate the protein but can also enter the cavity-binding site and influence its shape or, more importantly, mediate the ligand-receptor binding. There are different approaches to treat water molecules during the simulations. When water molecules play an important stabilizing effect, explicit water treatment should be used. In the case of HIV1 protease, a water

molecule (named W301) functions as a bridge between two lysins in the ligand-binding site and the ligand (i.e. the drug Saquinavir). Without this water, the ligand will not able to interact with the protein. With the computational cost of MD simulations, it is impossible to screen the huge number of FCCs with this technique. However, if the scope of the analysis is to study the mechanism of action (MoA) of an endocrine disruptor, molecular dynamics can be applied to a limited number of molecules. The analysis can give insight into how an EDC interacts with the NR, i.e. if it induces conformational changes compared to the endogenous ligand, the type of binding interactions inside the binding pocket, the effect of the compound in respect to the coactivator and corepressor binding, etc. Some kinds of parameters can be exploited to analyse the MD simulation results. The most commons are the use of the RMSD (root-main-square deviation) and RMSF (root-main-square fluctuation) values to monitor the stability of the system. Additionally, the hydrogen bond networks between the protein and the ligand and/or the protein and the coactivator/corepressor can be monitored during the simulation time to explore in more detail how ligand interacts with the NR compared to the endogenous ligand.

4.5 Case Studies

In this section, we illustrate some real case studies where in silico methods are applied together with the wet test (in vitro tests). Until a few years ago, the word computational in food science identified statistical applications, QSAR or COMFA applications. Taking into consideration what has been previously done in the medicinal chemistry field, screening, molecular docking and scoring functions can be used to discover new possible endocrine disruptors from a large dataset of food contact chemicals, such as food additives [1]. Starting from a joint FAO-WHO database of 1500 chemicals, Amadasi and colleagues screened 31 compounds predicting 13 of them as potential xenoestrogens towards oestrogen receptor alpha. Four of these compounds have been previously reported as well-known ER endocrine disruptors. Thus, the in silico analysis confirmed the prediction. For the other nine compounds, the binding affinity and oestrogenic effects were determined using in vitro assays. The most interesting result is propyl gallate that is a widely used antioxidant (in particular in the fish industry), and hexylresorcinols (www.fao. org/ag/agn/jecfa-additives) are predicted as oestrogen receptor binders both by in silico and in vitro analyses. It may be hypothesized that the latter has an indirect

effect and facilitates the interaction between unliganded ER and coactivators, inducing the transcription of the reporter.

Recently, EFSA considered “Safety and efficacy of propyl gallate for all animal species” paper important for the panel on additives and products or substances in animal feed [28].

Kenda and co-workers conducted a screening of 1046 US-approved and marketed small-molecule drugs for estimating their endocrine-disrupting properties [16]. Binding affinity to 12 nuclear receptors was assessed with a molecular docking program, Endocrine Disruptome. They identified 130 drugs with a high binding affinity to a nuclear receptor that is not their pharmacological target.

Another software, VitualToxLab, has been used to evaluate a subset of molecules, and the results have been compared with in vitro results from the Tox21 database. Another interesting approach of nonstatistical in silico prediction to screen oestrogenic and androgenic activity and to decipher the mechanism of binding (MOA) of substances of very high concern (SVHC) for the European Union is the case of bisphenols [5].

Bisphenol A (BPA) has been considered at first as toxic for reproduction and subsequently as an endocrine-disrupting chemical that interferes with the endocrine system mimicking the effects of oestrogen. Some European countries banned BPA from industrial production to avoid contact with the food and consequently with the human organism. Instead of BPA, they allowed the use of bisphenol S (BPS) as an alternative less active. The authors analysed a series of BPA alternatives and derivatives with similar physical-chemical properties that have been produced and used by companies for substituting it. They evaluated the oestrogenic and androgenic binding activity of 26 BPs against six different nuclear receptors using literature in vitro data for comparison. In this specific case, they propose a rough classification of the results, high binder, medium binder and low binder compared to bisphenol A as a reference. This rough ranking list could be useful and faster for massive screening instead of complex statistical analysis.

References

1. Amadasi A, Mozzarelli A, Meda C, Maggi A, Cozzini P. Identification of xenoestrogens in food additives by an integrated in silico and in vitro approach. Chem Res Toxicol. 2009;22(1):52–63. https://doi.org/10.1021/tx800048m.

2. Balaguer P, Delfosse V, Grimaldi M, Bourguet W. Structural and functional evidences for the interactions between nuclear hormone receptors and endocrine disruptors at low doses. C R Biol. 2017;340(9–10):414–20. https://doi.org/10.1016/j.crvi.2017.08.002.

3. Barril X. Physico-chemical and computational approaches to drug discovery; 2012. Retrieved from http://books.google.com/books?id=h7fWN9XtGx0C&pgis=1

4. Bissantz C, Folkers G, Rognan D. Protein-based virtual screening of chemical databases. 1.Evaluation of different docking/scoring combinations. J Med Chem. 2000;43(25):4759–67. https://doi.org/10.1021/jm001044l.

5. Cavaliere F, Lorenzetti S, Cozzini P. Molecular modelling methods in food safety: bisphenols as case study. Food Chem Toxicol. 2020;137(December 2019):111116. https://doi.org/10.1016/j.fct.2020.111116.

6. Cereto-Massagué A, Ojeda MJ, Valls C, Mulero M, Garcia-Vallvé S, Pujadas G. Molecular fingerprint similarity search in virtual screening. Methods. 2015;71(C):58–63. https://doi.org/10.1016/j.ymeth.2014.08.005.

7. Cozzini P, Kellogg GE, Spyrakis F, Abraham DJ, Costantino G, Emerson A, et al. Target flexibility: an emerging consideration in drug discovery and design. J Med Chem. 2008;51(20):6237–55. https://doi.org/10.1021/jm800562d.Target.

8. Cozzini P, Kellogg GE, Thurston DE. Computational approaches to nuclear receptors. WorldCat; 2012. Retrieved from http://web.a.ebscohost.com/ehost/ebookviewer/ebook/bmxlYmtfXzUxOTM5MF9fQU41?sid=d82f0efc-08f8-4571-aa23-9c5c190f2633@sessionmgr4001&vid=4&format=EB&rid=1

9. Davies M, Nowotka M, Papadatos G, Dedman N, Gaulton A, Atkinson F, et al. ChEMBL web services: streamlining access to drug discovery data and utilities. Nucleic Acids Res.2015;43(W1):W612–20. https://doi.org/10.1093/nar/gkv352.

10. Dellafiora L, Ruotolo R, Perotti A, Cirlini M, Galaverna G, Cozzini P, et al. Molecular insights on xenoestrogenic potential of zearalenone-14-glucoside through a mixed in vitro/in silico approach. Food Chem Toxicol. 2017;108:257–66. https://doi.org/10.1016/j.fct.2017.07.062.

11. DeSantis K, Reed A, Rahhal R, Reinking J. Use of differential scanning fluorimetry as a high-throughput assay to identify nuclear receptor ligands. Nucl Recept Signal. 2012;10:1–5. https://doi.org/10.1621/nrs.10002.

12. Farrugia LJ. WinGX and ORTEP for windows: an update. J Appl Crystallogr. 2012;45(4):849–54. https://doi.org/10.1107/S0021889812029111.

13. Gimeno A, Ojeda-Montes MJ, Tomás-Hernández S, Cereto-Massagué A, Beltrán-Debón R, Mulero M, et al. The light and dark sides of virtual screening: what is there to know? Int J Mol Sci. 2019;20(6) https://doi.org/10.3390/ijms20061375.

14. Hartung T. Toward a new toxicology – evolution or revolution? ATLA Altern Lab Anim.2008;36(6):635–9. https://doi.org/10.1177/026119290803600607.

15. Hass U, Scholze M, Christiansen S, Dalgaard M, Vinggaard AM, Axelstad M, et al. Combined exposure to anti-androgens exacerbates disruption of sexual differentiation in the rat. Environ Health Perspect. 2007;115(SUPPL1):122–8. https://doi.org/10.1289/ehp.9360.

16. Kenda M, Dolenc MS. Computational study of drugs targeting nuclear receptors. Molecules. 2020;25(7) https://doi.org/10.3390/molecules25071616.

17. Li Y, Luh CJ, Burns KA, Arao Y, Jiang Z, Teng CT, et al. Endocrine-disrupting chemicals (EDCs): in vitro mechanism of estrogenic activation and differential effects on ER target genes. Environ Health Perspect. 2013;121(4):459–66. https://doi.org/10.1289/ehp.1205951.

18. Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Félix E, et al. ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res. 2019;47(D1):D930–40. https://doi.org/10.1093/nar/gky1075.

19. Rush TS, Grant JA, Mosyak L, Nicholls A. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. J Med Chem. 2005;48(5):1489–95. https://doi.org/10.1021/jm040163o.

20. Sastry GM, Dixon SL, Sherman W. Rapid shape-based ligand alignment and virtual screening method based on atom/feature-pair similarities and volume overlap scoring. J Chem Inf Model. 2011;51(10):2455–66. https://doi.org/10.1021/ci2002704.

21. Sterling T, Irwin JJ. ZINC 15 – ligand discovery for everyone. J Chem Inf Model. 2015;55(11):2324–37. https://doi.org/10.1021/acs.jcim.5b00559.

22. Thouennon E, Delfosse V, Bailly R, Blanc P, Boulahtouf A, Grimaldi M, et al. Insights into the activation mechanism of human estrogen-related receptor γ by environmental endocrine disruptors. Cell Mol Life Sci. 2019;76(23):4769–81. https://doi.org/10.1007/s00018-019-03129-x.

23. Vitkup D, Ringe D, Karplus M, Petsko GA. Why protein R-factors are so large: a self-consistent analysis. Proteins Struct Funct Genet. 2002;46(4):345–54. https://doi.org/10.1002/prot.10035.

24. Wang R, Wang S. How does consensus scoring work for virtual library screening? An idealized computer experiment. J Chem Inf Comput Sci. 2001;41(3–6):1422–6. https://doi.org/10.1021/ci010025x.

25. Weikum ER, Liu X, Ortlund EA. The nuclear receptor superfamily: a structural perspective. Protein Sci. 2018;27(11):1876–92. https://doi.org/10.1002/pro.3496.

26. Wermuth CG, Ganellin CR, Lindberg P, Mitscher L a. Glossary for chemists of terms used in medicinal chemistry. Pure Appl Chem. 1998;70(5):1129–43. https://doi.org/10.1351/pac197951081725.

27. Zhang J, Li Y, Gupta AA, Nam K, Andersson PL. Identification and molecular interaction studies of thyroid hormone receptor disruptors among household dust contaminants. Chem Res Toxicol. 2016;29(8):1345–54. https://doi.org/10.1021/acs.chemrestox.6b00171.

28. Bampidis, V. et al. Safety and efficacy of propyl gallate for all animal species. EFSA J. 2020;18:1–16.

CHAPTER 5

Molecular modelling methods in food safety: Bisphenols as case study

Francesca Cavaliere, Stefano Lorenzetti, Pietro Cozzini

Published in Food and Chemical Toxicology
https://doi.org/10.1016/j.fct.2020.111116

Abstract

Bisphenol A (BPA), a synthetic compound widely used as a building block for polycarbonate plastics, has been declared in the European Union (EU) as a substance of very high concern (SVHC). A series of BPA alternatives and derivatives (bisphenols/BPs) with similar physical-chemical properties have been produced and used by companies for substituting it. To evaluate the estrogenic and androgenic binding activity of 26 BPs, a non-statistical in silico approach has been applied. The results of molecular docking analyses applied on six different nuclear receptors (NRs) have revealed that: i) some BPA metabolites could lower the harmful effects of BPA exposure; ii) BPS is a lower interactor for all NRs, but it does not appear safer at all for androgen receptor (AR), for which its binding activity is found similar to a pharmacological anti-androgen; iii) only a BP has been found as a safer compound for all NRs considered. Moreover, molecular dynamic simulation of three BPs on ERα have revealed that the presence of negative hydrophobic interactions could induce a decrease in receptor activity. Overall, the present results demonstrate that in silico methods could be a valid approach to screen estrogenic and androgenic activity of food contact materials (FCMs).

Keywords

Molecular docking, Bisphenol A, Bisphenols, Estrogen receptor, Androgen receptor, Molecular dynamics

Abbreviations

AD, AutoDock; AO, Adverse Outcome; AOP, Adverse Outcome Pathway; AR, Androgen Receptor; BP, bisphenol; CAR, Constitutive Androstane Receptor; ChemSec, International Chemical Secretariat; ECHA, European Chemical Agency; ED, Endocrine Disruptor; EDC, Endocrine-Disrupting Chemical; ER, Estrogen Receptor; ERRγ Estrogen-Related Receptor; FLAP, Fingerprint for Ligand and Protein; GR, Glucocorticoid Receptor; H, higher; H4, helix 4; H8, helix 8; HINT, Hydropathic INTeraction; KE, Key Event; L, lower; M, medium; MD, Molecular Dynamic; MIE, Molecular Initiating Event; MR, Mineralocorticoid Receptor; NR, Nuclear Receptor; PDB, Protein Data Bank; PPAR, Peroxisome Proliferator-Activated Receptor; PR, Progesterone Receptor; PXR, Pregnane-X-Receptor; REACH, Registration, Evaluation, Authorisation and Restriction of Chemicals; RMSD, Root Main Square Deviation;

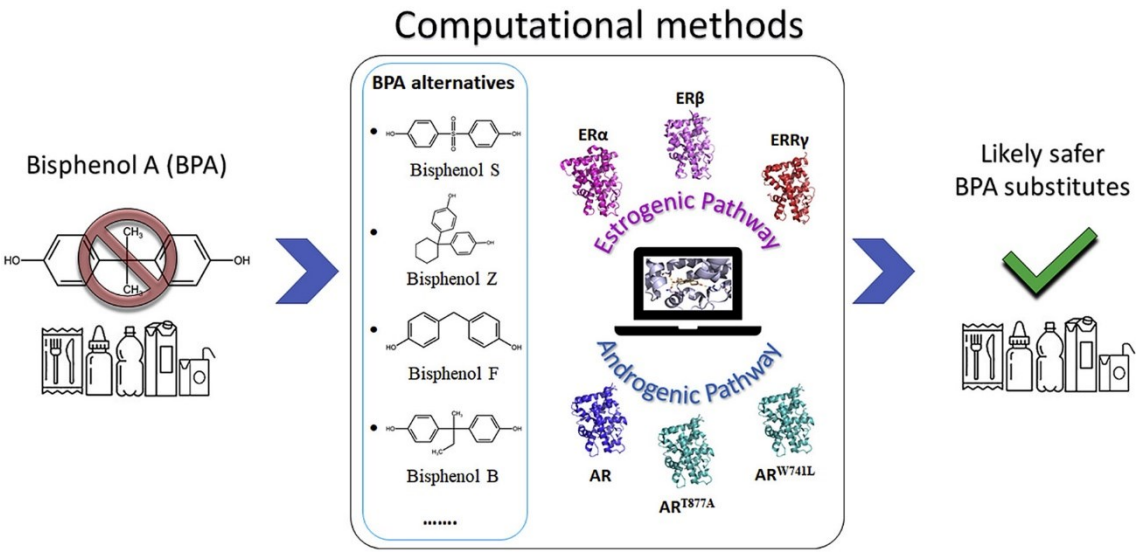
RMSF, Root Main Square Fluctuation; RPA, Relative Predicted Activity; SIN, "Substitute It Now!"; SVHC, Substance of Very High Concern; THR, Thyroid Hormone Receptor.

Author contribution

COZZINI and CAVALIERE contributed equally for the computational simulation and discussion. LORENZETTI contributed for the literature experimental data.

The manuscript has been wrote and discussed with the equal contribution of the three authors.

Graphical Abstract



5.1 Introduction

Bisphenol A (BPA) or 4,4′-isopropylidenediphenol, is a man-made chemical mostly known to be used to manufacture plastics and resins (Fenichel et al., 2013; EFSA, 2015; Pjanic, 2017). BPA is a building block in polycarbonate, a transparent and rigid plastic used to make food containers, including returnable beverage bottles, infant feeding (baby) bottles, tableware, mugs, and storage containers, is widely recognized. BPA is also present in epoxy resins used to make protective coatings and linings for food and beverage cans and vats (Fenichel et al., 2013; EFSA, 2015; Pjanic, 2017). Hence, dietary exposure to BPA can occur due to its migration in small amounts into food and beverages stored in food containers containing the chemical (EFSA, 2015).

Furthermore, thermal printers present in very common devices, such as adding machines, cash registers and credit card terminals, work with a special paper commonly known as thermal paper that also contains BPA. The latter usage accounts for dermal exposure to BPA. In the European Union (EU), under the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulation [Regulation (EC) No. 1907/2006], since January 2017 the European Chemical Agency (ECHA) classified BPA as a substance of very high concern (SVHC): nowadays, the alerts include both toxic for reproduction (REACH art.57c) and endocrine disrupting properties (REACH art.57f). Hence, being BPA included in the Candidate List for Authorization (REACH Annex XIV; https://echa.europa.eu/candidate-list-table), the effort to look for BPA alternative chemicals with a lower toxicological profile is increasing (Lorenzetti and Cozzini, 2017). Indeed, such a search for BPA substitutes became mandatory since the BPA usage as a dye developer in thermal paper will be prohibited due to its complete ban foreseen in 2020. Indeed, such a ban already lead to a partial substitution in which bisphenol S (BPS) usage as dye developer increased about 98% in just one year (from 2016 to 2017; https://echa.europa.eu/it/-/bpa-being-replaced-by-bps-in-thermal-paper-echa-survey-finds).

In reality, the search for BPA alternatives and derivatives (bisphenols/BPs) with similar physico-chemical properties started many years earlier and many of them have been produced and used to substitute it. Within the EU, 59 registered unique BPA derivatives are potentially marketed under REACH regulation. Of them, the most commonly used as BPA alternatives are: i) BPS or 4′-sulphonyldiphenol, used in plastics, food packaging, cosmetics, baby bottles and in BPA-free paper (Wu et al., 2018), ii) bisphenol F (BPF) or

4,4′-methylenediphenol, used in epoxy resins, food packaging, soda cans, cosmetics and hygiene products, iii) bisphenol AF (BPAF) or 4,4'-(hexafluoroisopropylidene)diphenol, used in electronic materials, high temperature copolymers and gas permeable membranes (Liao et al., 2012), iv) the halogenated derivative of BPA, tetrabromobisphenol A (TBBPA) or 4,4′-(propane-2,2-diyl)bis(2,6-dibromophenol), used as flame retardants (Darnerud, 2003; Chu et al., 2005; de Wit et al., 2010). Despite this, considering the EU market and taking into account the publicly available information about the ECHA registration dossiers, the declared annual tonnage range of BPA, one of the most abundant chemicals produced worldwide, is still very much higher than the sum of BPS, BPF, TBBPA and BPAF annual tonnages range (see Table 1). Indeed, as mentioned above, the REACH-banned BPA should be substituted by safer alternatives in 2020 and so far, it is expected that BPS, the most used alternative in thermal paper, will have a worldwide increase in production and usage.

Given the concerns and the restrictions about BPA use, BPs has been employed since many years as BPA substitutes, at least in some applications. Nowadays, BPs are recognized as novel environmental pollutants and suspected to have similar endocrine disruptor (ED)-like concern than BPA for both wildlife and human health (Duan et al., 2018; Wu et al., 2018). Indeed, searching within the International Chemical Secretariat (ChemSec) online tool, the "Substitute It Now!" (SIN) List database (https://chemsec.org/sin-list/), in which are present 919 hazardous chemicals likely to be banned or restricted in the future following the REACH advices, it is possible to find eight chemicals classified as bisphenols with an ED-like concern. The most commonly used BPs, namely BPA, BPS, BPF, and TBBPA, are all present in such a database.

As previously mentioned, BPA has been considered at first as toxic for reproduction and subsequently as an endocrine-disrupting chemical (EDC) that interferes with the endocrine system mimicking the effects of an estrogen. BPA is considered a weak hormone-like chemical, able to bind the estrogen receptors (ERs), whose role encompass adverse health effects on the reproductive, metabolic and cardiovascular tissues, on the immunological and central nervous system as well, and even on the triggering and development of hormone-dependent cancers (Pjanic, 2017; Prins et al., 2018; Patisaul, 2019). In recent years, several studies suggested BPA as a ligand interacting not only with ERs, but also with several other nuclear receptors (NRs). No animal testing studies, either in silico or in vitro (among the most recent: Grimaldi et al., 2019, Kojima et al., 2019,

MacKay and Abizaid, 2018, Usman and Ahmad, 2019, Wang et al., 2017), suggested to include as BPA targets also other NRs including androgen receptor (AR), estrogen-related receptor γ (ERRγ), glucocorticoid receptor (GR), progesterone receptor (PR), mineralocorticoid receptor (MR), pregnane-X-receptor (PXR), constitutive androstane receptor (CAR), peroxisome proliferator-activated receptors (PPARα, PPARβ and PPARγ) (Sharma et al., 2018), thyroid hormone receptor α (TRα), thyroid hormone receptor β (TRβ) and retinoic X receptors (RXRα, RXRβ and RXRγ) (Sharma et al., 2018). All of them, besides ERs, are further involved in modulating reproductive, metabolic and other hormone-dependent tissues along all stages of the life cycle (Diamanti-Kandarakis et al., 2009; Gore et al., 2015).

Notably, the existing no animal testing studies comparing BPA to its derivatives are already confirming that some NRs, different from the classical ERs, are common targets for several BPs (Grimaldi et al., 2019, Kojima et al., 2019, Usman and Ahmad, 2019, Zenata et al., 2017). Hence, the search for a reliable and fast computational approach to confidently screen for the endocrine-like toxicological properties of BPs is particularly relevant in order to avoid BPA substitution with chemicals potentially of equivalent concern as SVHC, such the above-mentioned BPS (Lorenzetti and Cozzini, 2017).

To date, different in silico methodologies have been previously reported to screen BPs against different NRs, showing their applicability to study BPs binding affinity. Since it is well known that BPA binds strongly ERRγ, Babu et al. (2012) have reported the binding capability of some chlorinated and nitrated BPA against ERRγ using molecular docking to understand BPA secondary products toxicodynamics and their estrogenic activity. All these derivates have shown higher binding affinity compared to the natural ligand (estradiol), but lower affinity compared to BPA (Babu et al., 2012). Using an in silico virtual screening approach, Delfosse et al. (2012) have studied the interaction of seven bisphenols (comprising BPA) against ERs (α and β isoforms), AR and ERRγ, correlating the data with in vitro studies and showing a good reliability of in silico method prediction (Delfosse et al., 2012). Sharma et al. (2018) have used molecular docking analyses to evaluate the interaction between 18 different bisphenol analogues and BPA against PPARs and RXRs, identifying BPA as a good binder and several BPs as higher interactors than BPA for both PPARs and RXRs (Sharma et al., 2018). Conroy-Ben et al. (2018) have used two approaches (in vitro/in silico) to study estrogen-like and anti-androgenic activity of different bisphenols. BPA, BPF and BPS and some bisphenols have been

predicted as EDs for ER and AR, highlighting that more hydrophobic bisphenols tended to have higher binding potencies for these two receptors (Conroy-Ben et al., 2018). Using molecular dynamics simulations, Lanlan et al. (2015) have shown the BPA mechanism of interaction with three different NRs, namely ERα, ERRγ and PPARγ, reporting that hydrogen bonds and hydrophobic interactions driven the binding process maintaining NRs in the active conformation (i.e. agnostic activity) (Lanlan et al., 2015). Also, Lu et al. (2018) have used molecular dynamics simulation to study the mechanism of binding of six different bisphenols against TRβ, combining the analysis with experimental data. By the results obtained from in vitro and in silico analyses, BPs seem to have anti-thyroid hormone activity onto TRβ and in particular brominated bisphenols (TBBPS and TBBPA) (Lu et al., 2018). Thus, based on the previously results, the main goal of this paper is to demonstrate that computational approaches could be used as in silico prioritization screening tools to predict the interaction and the mechanism of action between BPs and some of their targeted sex steroid NRs, including ERα, ERβ, ERRγ, AR and two adenocarcinoma-relevant AR mutated forms, ART877A and ARW741L. In order to reach this goal, molecular docking analyses have been used to study the interaction between twenty-six BPs (including 7 BPA metabolites) and the above-mentioned six different NRs.

Consequently, the obtained data sets will represent the most up-to-date comparative database of BPs-bound sex steroid NRs that could be used in parallel with the BPs-dependent, in vitro NR transactivation data sets produced by Kitamura and colleagues in 2005 to characterize step-by-step the structure-activity relationship of these NR-bound chemicals or, in other words, their ability of BPs to both bind and modulate gene transcription of a certain sex steroid NR.

In turn - using the vocabulary of the Adverse Outcome Pathway (AOP)-toxicological scheme (Lorenzetti et al., 2015) - the obtained data sets will represent also the most completed comparative data sets of BPs-associated Molecular Initiating Events (MIEs) to be directly linked to the currently used screening methods that are based on the first Key Event (KE) connecting BPA and BPs to their recognized Adverse Outcomes (AOs). The identification of different, multiple BPs-associated MIEs allows to consider a cumulative endocrine disrupting activity, the AO, that could be mediated by BPs.

Table 1
List of BPA alternatives, derivatives and metabolites used in this study.

Common name, acronym	IUPAC name	M.W. (g/mol)	CAS no.	EC no.	ECHA registration: total tonnage band (tonnes per annum)[1]
Bisphenol A, BPA	4,4'-isopropylidenediphenol	228.29	80-05-7	201-245-8	registered, 1 000 000 - 10 000 000
Bisphenol AF, BPAF	4,4'-[2,2,2-trifluoro-1-(trifluoromethyl)ethylidene]diphenol	336.23	1478-61-1	216-036-7	registered, 100 - 1 000
Bisphenol AP, BPAP	1,1-bis(4-hydroxyphenyl)-1-phenylethane	290.36	1571-75-1	433-130-5	registered, confidential
Bisphenol B, BPB	4,4'-(1-methylpropylidene)diphenol	242.31	77-40-7	201-025-1	not registered
Bisphenol BP, BPBP	4,4′-(diphenylmethylene)diphenol	352.43	1844-01-5		
Bisphenol C, BPC	4,4'-isopropylidenedi-o-cresol	256.34	79-97-0	201-240-0	registered, 0 - 10
Bisphenol C 2, BPC2	Bis(4-hydroxyphenyl)-2,2-dichloroethylene	281.13	14868-03-2		
Bisphenol E, BPE	4,4′-ethylidenediphenol	214.26	2081-08-5		
Bisphenol F, BPF	4,4′-methylenediphenol	200.23	620-92-8	210-658-2	not registered
Bisphenol G, BPG	4,4′-isopropylidenedi(2-isopropylphenol)	312.45	127-54-8		
Bisphenol M, BPM	4,4'-(1,3-phenylene-bis(1-methylethylidene))diphenol	346.47	13595-25-0	428-970-4	registered, 0 – 10 confidential
Bisphenol P, BPP	4,4′-(1,4-phenylenediisopropylidene)diphenol	346.46	2167-51-3		
Bisphenol PH, BPPH	5,5′-isopropylidenedi-2-biphenylol	380.48	24038-68-4		
Bisphenol S, BPS	4,4'-sulphonyldiphenol	250,27	80-09-1	201-250-5	registered, 10 000 - 100 000 intermediate use only
Bisphenol TMC, BPTMC	4,4'-(3,3,5-trimethylcyclohexane-1,1-diyl)diphenol	310.43	129188-99-4	404-140-7 603-320-4	
Bisphenol Z, BPZ	4,4′-cyclohexylidenediphenol	268.35	843-55-0	212-677-1	not registered
Tetrabromo BPA, TBBPA	2,2',6,6'-tetrabromo-4,4' isopropylidenediphenol	543.87	79-94-7	201-236-9	registered, 1 000 - 10 000
Tetrachloro BPA, TCBPA	2,2-bis-(3,5-dichloro-4-hydroxyphenyl)propane, 2,2',6,6'-tetrachloro-4,4'-isopropylidenediphenol	366.07	79-95-8	201-237-4	not registered
Tetramethyl BPA, TMBPA	4,4'-isopropylidenedi-2,6-xylol, 2,2',6,6'-tetramethyl-4,4'-isopropylidenediphenol	284.39	5613-46-7	227-033-5	registered, 10 - 100
BPA carboxylic acid	2,2-bis-(4-hydroxyphenyl)-1-propionic acid	258.27	92549-67-2	---	not registered
BPA catechol			79371-66-7		
BPA glucuronide	bisphenol A β-D-glucuronide, 4-[1-(4-hydroxyphenyl)-1-methylethyl]phenyl β-D-glucopyranosiduronic acid	404.41	267244-08-6		

BPA ol	2,2-bis-(4-hydroxyphenyl)-1-propanol	244.28	142648-65-5
BPA quinone	4,5-bisphenol-o-quinone	242.27	163405-36-5
BPA sulfate			-
MBP	4,4'-(4-methylpent-1-ene-2,4-diyl)diphenol	268.35	13464-24-9

[1] Data obtained from: https://echa.europa.eu/it/information-on-chemicals/registered-substances

Finally, in order to understand the differences in the mechanism of ERα binding of BPA with two main BPs, namely BPS and BPF, a more detailed analysis has been made using a molecular dynamics approach.

5.2 Materials and methods

5.2.1 Protein preparation

The crystallographic structures of human ERα, ERβ, ERRγ, AR, ART877A and ARW741L were taken from the RCSB Protein Data Bank (PDB) with the entry codes: 2YJA (R = 1.82 Å), 2YJD (R = 1.93 Å), 2E2R (R = 1.60 Å), 2AM9 (R = 1.64 Å), 2AX6 (R = 1.5 Å) and 2AX8 (R = 1.7 Å), respectively. All the NRs were processed using Sybyl software v8.1 (www.tripos.com): water molecules and ligands were removed, hydrogen atoms were added and energy minimized using the Powell algorithm with a coverage gradient of ≤0.5 kcal (mol Å)−1 and a maximum of 1500 cycles. However, for the docking with AutoDock (AD; see below), the receptors were further processed as followed: the AutoDockTools software was used to add polar hydrogen to the proteins and the Gasteiger charges were calculated for each atom and to assign AD4 type to the atoms.

5.2.2 Ligand preparation

The ligands were retrieved from the PubChem database in .sdf format. For the chemicals not available, the three-dimensional structures were built and energy minimized with Sybyl software v8.1 using the Powell algorithm with a coverage gradient of ≤0.5 kcal (mol Å)−1 and a maximum of 1500 cycles. Moreover, in order to assign the correct protonation state to the ligands (pH = 7.4), the software FLAP (Fingerprint for Ligand and Protein) was used.

5.2.3 GOLD docking

The GOLD software v5.2.2 (CCDC; Cambridge, UK; http://www.ccd.cam.ac.uk) was applied to dock bisphenols into the binding site of the wild type and mutated receptors.

For each compound and NRs, 50 binding poses were generated without any constraints. The centroid of the binding site was defined using the coordinates of the crystallographic complexes: ERα: #C9 of EST (x = 22,994 |y = 10,847 |z = 10,050); ERβ: #N13 of YJD (x = -14,459 |y = -18,473 |z = 2,843); ERRγ: #C2 of 2OH (x = -16,478 |y = -3,037 |z = -27,356); AR: #C9 of TES (x = 27,068 |y = 2,720|z = 4,565); ART877A (2AX6): #C1 of HFT (x = 27,651 |y = 3,467 |z = 3,875) and ARW741L (2AX8): #C6 of FHM (x = 27,692 |y = 3,578|z = 3,633).

Side chain flexibility was allowed for the amino acids: ERα: Phe404, Met421, Ile424, Phe425, His524, Leu525 (Cozzini and Dellafiora, 2012); ERβ: His475, Met340, Phe377; ERRγ: Glu275, Arg316, Leu345; AR: Leu701, Asn705, Gln711, Trp741, Met745, Met780, Thr877, Met895, Ile899; ART877A: Leu701, Asn705, Gln711, Trp741, Met745, Met780, Met895, Ile899; ARW741L: Leu701, Asn705, Gln711, Leu741, Met745, Met780, Thr877, Met895, Ile899.

For the genetic algorithm run, a maximum number of 100000 operations were performed on a population of 100 individuals with a selection pressure of 1.1. Operator weights for crossover, mutation, and migration were set to 95, 95, and 10, respectively. The number of islands was set to 5, and the niche size was set to 2. The default GOLD Score fitness function was applied for performing the energetic evaluations. The distance for hydrogen bonding was set to 2.5 Å, and the cut-off value for the van der Waals calculation to 4.0 Å. For ligand flexibility options, flip pyramidal N, flip amide bonds, and flip ring corners were allowed. After that, all the poses generated by GOLD were rescored using the scoring functions ChemScore and HintScore (HINT, Hydropathic INTeraction). The coupling of these three scoring function was chosen as: i) GoldScore allows to take into account factors such as H-bonding energy, van der Waals energy and ligand torsion strain; ii) ChemScore represents the total free energy change that occurs on ligand binding (trained by regression against binding affinity data for 82 complexes) and also incorporates a protein-ligand atom clash term and an internal energy term taking into account hydrophobic-hydrophobic contact area, hydrogen bonding and ligand flexibility; iii) HintScore provides a quantitative evaluation of protein-ligand interaction that allows to take into account both the enthalpic and entropic contributions to the ΔG of ligand-protein interaction, based on experimental protein and ligand Log Po/w values (Kellogg and Abraham, 2000; Kellogg et al., 2001).

5.2.4. AutoDock docking

The search space was included in a box of 24 × 24 × 24 Å, centred on the binding site of the ligands (as mentioned before in the Gold Docking paragraph 2.3). The side chain flexibility was allowed for the same residues defined in the Gold Docking. The ligand amide and backbone flexibility were allowed.

5.2.5 Protein-ligand complex preparation

The best HINT pose was chosen as starting point of the molecular dynamic simulations. The protein-ligand complex was prepared using the web-based graphical user interface CHARMM-GUI (http://www.charmm-gui.org/). Each complex was solvated in a rectangular 15 Å water box (TIP3S) and six phosphorus anions were added only to neutralize in three complexes. The final system contains, respectively, 64,530, 64,506 and 64,501 atoms, where around 60,500 were water molecules, for protein in complex with BPA, BPS and BPF, respectively.

5.2.6 Molecular dynamic simulations

Molecular dynamic simulations were performed using NAMD 2.0 software package. For each system, two rounds of energy minimization were performed, each comprising 0,1ns of conjugate gradient minimization. Firstly, a weak constraint (1.0 kcal/(mol·Å2)) was assigned to all heavy atoms for both the protein and the ligand, allowing the minimization of hydrogen atoms. Secondly, no restraints were employed in order to allow the minimization of the entire system. After that, three rounds of 0.1 ns in NVT mode were applied for gradually heating the system to 300 K (0–100 K, 100 K–200 K, 200 K–300 K) using the Langevin thermostat. The system was further equilibrated at a constant pressure (1.0 bar) for 1 ns (NPT). The molecular dynamic simulation has been running with a strong constraint (5.0 kcal/(mol·Å2)) to the atoms of the ligand for the first 5ns. After that, the ligand's constraint was gradually decreased of 1.0 kcal/(mol·Å2) every 5 ns for a total simulation run of 25 ns. After that, the simulation was performed for 75 ns without any constraint, allowing the movement of the entire system. Thus, the total simulation run is of 100ns for each complex.

5.3 Results and discussion

EDs can alter multiple endocrine pathways. However, as recently further highlighted "EDs particularly influence and perturb the steroidogenesis and the reproduction since most of the effects are exerted through disturbance of estrogen- or androgen-mediated

processes" (De Falco et al., 2015). Thus, a particular attention was made to these two pathways, since the imbalance in their regulation and/or function is related to a wide range of diseases, such as breast, endometrial and prostate cancer (Diamanti-Kandarakis et al., 2009; Gore et al., 2015; Rochefort, 2017; Sifakis et al., 2017). In this study we have focused our attention on six different NRs, related to the estrogen and androgen signaling pathways. In order to study the interference induced by BPs on estrogen processes, the two ER isoforms, ERα and ERβ, as well as ERRγ were taken into account. In the case of ERRs, we have only analysed the gamma isoform for these following reasons. Firstly, the structure of the beta isoform was not available in PDB database, and thus, in silico studies cannot be performed unless to use homology modelling. However, the lack of a reference compound (i.e. a substance that is a well-known binder of this receptor) could led to less accurate in silico analyses. Secondly, the volume of the binding pocket of ERRα in the agonistic conformation is very small (100 Å3) (Kallen et al., 2004) compared to the volume of bisphenols included in our dataset. In fact, the volume of ligands ranges from 170,2 Å3 to 334,5 Å3 and thus, it seems improbable that these class of BPs could bind to ERRα. For these two reasons, the ERRα and ERRβ isoforms were not taken into account. To study the BP-induced interferences on the AR-signaling pathway, three different forms of ARs were employed: the wild type conformation and the two AR mutated structures ART877A and ARW741L. These two AR mutations are commonly developed by patients with AR-dependent adenocarcinoma prostate cancer treated with AR antagonists (the most common being flutamide and bicalutamide); long-term AR-antagonist treatments convert the drug to an agonists behavior, making the patients unresponsive (Zhang and Zhong, 2010; McCrea et al., 2016). Moreover, it has been reported that although BPA has no agonistic activity on the wild type AR, it has an agonistic effect on the mutated ART877A, promoting prostate cancer cell proliferation (Wetherill et al., 2002; Hess-Wilson et al., 2007; Teng et al., 2013). Thus, we can speculate that multiple EDCs could interact with the AR in different ways based on its form, and for these reasons, we have included them in our study.

5.3.1 Molecular docking consensus scoring method

A total of twenty-six BPs (BPA, 18 BPs alternatives or derivatives and 7 BPA metabolites), listed in Table 1 and shown in Fig. 1), were computationally analysed for their capability to bind the above-mentioned six different NRs, namely ERα, ERβ, ERRγ, AR, ART877A and ARW741L.

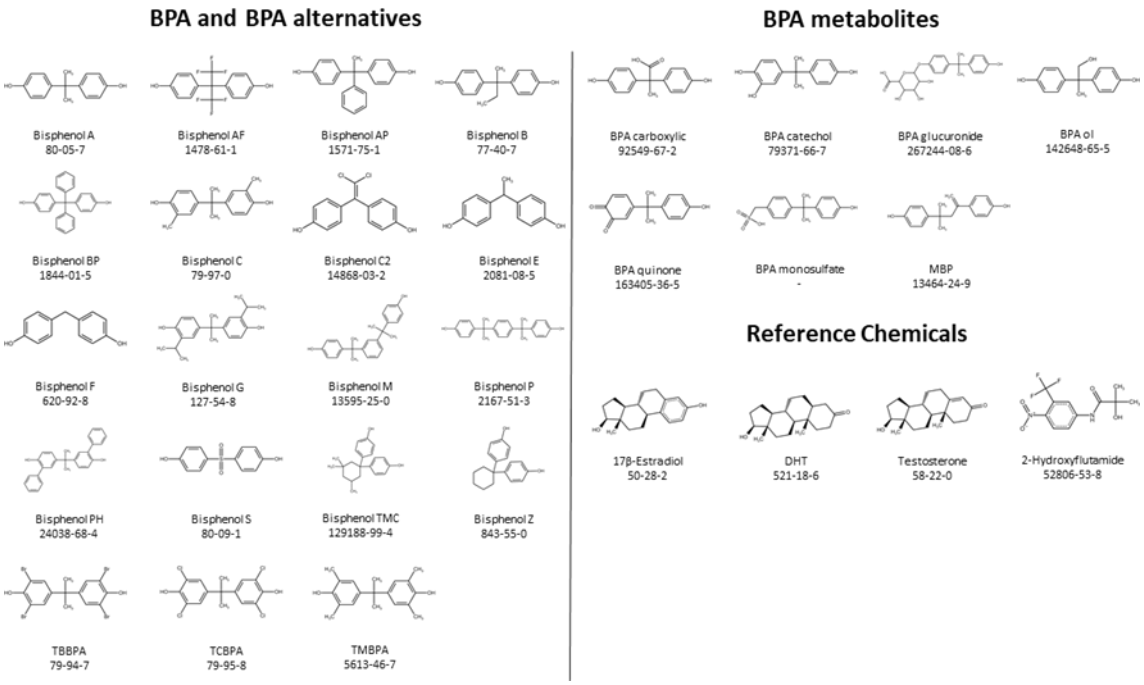


Figure 1: Chemical structure of BPA and other bisphenols and BPA metabolites used in this study. Reference chemicals were also included.

In our dataset, BPA has been considered as a reference compound due to its recognized role as EDC. We have performed the analyses with two different docking programs (GOLD and AutoDock) and four different scoring functions. In particular, the GOLD software was used to generate the poses of each chemical, which were scored with GoldScore, ChemScore, and HintScore; instead, the AutoDock software was used to generate and to score the poses with the internal scoring function. The obtained values for each chemical were used to calculate the BPA Relative Predicted Activity (RPA): every molecule has been scored in comparison to BPA, used as reference compound. In more detail, BPA RPA has been defined as follow:

$$BPA\ Relative\ Predicted\ Activity\ (RPA) = \frac{food\ contaminant\ score}{reference\ compound\ score} = \frac{BP\ score}{BPA\ score}$$

The chemicals with RPA greater and lower than 1 were considered respectively, as higher (H) and lower (L) EDC-like chemicals compared to BPA or, in other words, BPA-like chemicals. Thus, four different RPA values, related to each scoring functions, have been obtained for every single chemical.

A consensus scoring was applied based on the best score obtained by each scoring function. Since we have analysed different NRs that have different physico-chemical characteristics inside the binding pocket, the consensus scoring method was employed to

overcome the possible limitations derived by molecular docking software. The combination of different scoring functions allows to reduce the number of false positive across different targets, leading to more reliable results (Charifson et al., 1999), and in particular, it has been previously highlighted how the combination of three different scoring functions enhances the capability to reach hit rates from 10% up to 65–70% (Bissantz et al., 2000). Thus, to simplify the results we have distinguished three cases: a) higher (H) means a BPA-like EDC with at least three of the four used scoring functions predicting a chemical as a better ligand of a specific NR; b) medium (M) means a BPA-like EDC with two scoring functions predicting a chemical twice as a better ligands and twice as a worse one; c) lower (L) means a BPA-like EDC with at least three of the four scoring functions predicting a chemical as a worse ligand. In all cases in which one or more scoring functions predicted a chemical as a ligand similar to BPA (RPA = 1), more weight was conferred to the consensus between the remaining scoring functions (see Table 2).

Table 2:
BPA Relative Predicted Activity of BPs. A. BPA alternatives and derivatives. B BPA metabolites.

A. BPA alternatives and derivatives							
CAS	**Name**	**ERα**	**ERβ**	**ERRγ**	**AR_WT**	**AR_T877A**	**AR_W741L**
80-05-7	Bisphenol A	-	-	-	-	-	-
1478-61-1	Bisphenol AF	L	M	M	L	L	H
1571-75-1	Bisphenol AP	H	M	L	M	H	H
77-40-7	Bisphenol B	H	M	H	L	M	M
1844-01-5	Bisphenol BP	L	H	L	L	L	L
79-97-0	Bisphenol C	H	H	H	M	M	H
14868-03-2	Bisphenol C 2	H	H	L	L	H	H
2081-08-5	Bisphenol E	M	H	L	L	H	H
620-92-8	Bisphenol F	L	L	L	L	M	H
127-54-8	Bisphenol G	H	M	L	L	H	H
13595-25-0	Bisphenol M	H	H	M	H	H	H
2167-51-3	Bisphenol P	H	M	M	M	H	H
24038-68-4	Bisphenol PH	M	L	L	H	H	H
80-09-1	Bisphenol S	L	L	L	L	M	L
129188-99-4	Bisphenol TMC	H	H	L	L	L	H

843-55-0	Bisphenol Z	H	H	H	L	M	M
79-94-7	TBBPA	M	L	L	M	M	H
79-95-8	TCBPA	H	L	L	L	M	H
5613-46-7	TMBPA	H	L	L	M	L	H

B. BPA metabolites							
CAS	**Name**	**ERα**	**ERβ**	**ERRγ**	**AR_WT**	**AR_T877A**	**AR_W741L**
92549-67-2	BPA carboxylic	L	L	L	L	M	L
79371-66-7	BPA catechol	M	M	H	M	M	H
267244-08-6	BPA glucuronide	M	L	M	L	M	M
142648-65-5	BPA ol	L	L	L	L	L	L
163405-36-5	BPA quinone	L	L	L	L	M	H
-	BPA sulfate	M	L	H	H	H	H
13464-24-9	MBP	H	H	H	H	H	H

5.3.2 In silico prediction of endocrine disruptor activity of BPs against estrogen receptors (ERs)

As shown in Table 2, Table 3, twelve BPs have been predicted as higher interactors (better ligands) compared to the BPA, seven as medium and six as lower interactors (worse ligands) for ERα. In our dataset, 17β-estradiol (E2), the ERα endogenous ligand, has been also included as a reference chemical, and it was correctly predicted as a higher interactor compared to BPA (data not shown). In fact, it is well known by experimental data that the binding affinity of BPA to the ERα is weaker than its natural ligand. Even if the different sources about the ratio between BPA/E2 are quite dissimilar, almost all of them predict it as a lower interactor (Kuruto-niwa et al., 2005; Bolli et al., 2008; Grignard et al., 2012).

Table 3
Number of BPs predicted as higher, medium and lower ligand of the 6 studied NRs.

	ERα	**ERβ**	**ERRγ**	**AR**	**AR^{T877A}**	**AR^{W741L}**
H	12	8	6	4	9	18
M	6	6	4	6	11	3
L	7	11	15	15	5	4

In order to rank BPs computational binding affinities to sex steroids receptors, the HintScore value have been used since it takes into account both enthalpic and entropic contributions to the ligand-protein interaction. Moreover, the HintScore has been widely applied on both ERs and AR (Cozzini and Dellafiora, 2012; Ginex et al., 2014; Dellafiora et al., 2015, 2017; Ehrlich et al., 2015).

Hence, the computational ERα binding affinity of BPs has been ranked as follows: BPM > MBP > BPP > BPTMC > TBBPA > BPAP > BPG > E2 > TCBPA > TMBPA > BPC > BPB > BPA > BPPH > BPBP > BPZ > BPE > BPA quinone > BPAF > BPC2 > BPA catechol > BPF > BPA_ol > BPA carboxylic > BPA sulfate > BPS > BPA glucuronide.

To assess our procedure reliability, in silico results have been compared to some in vitro studies. It is fundamental to highlight that molecular docking predicts how strength is the interaction between a compound and a NR. On the contrary, most of experimental in vitro studies take into account a downstream effect, since they monitor the effect on the transcription of genes that are regulated by a NR. Other in vitro studies, instead, are focused on the general intracellular effect. A more grade of correlation could be obtained using data coming from in vitro competitive binding assays, which considered the direct compound binding affinity. Blair et al. (2000) have considered 188 natural and xeno-chemicals to study their capability to bind ERα. Between them, four BPs (BPB, BPA, BPS and BPA carboxylic) are in common with our study. Their rank list was BPB > BPA > BPS > BPA carboxylic: we obtained a similar prediction. In fact, although BPS and BPA carboxylic are exchanged in our study, if we consider the mean IC50 values obtained by Blair et al., BPS and BPA carboxylic show the same order of magnitude (1.05 x 10–4 ± 0.35 x 10–4 and 1.20 x 10–4 ± 0.30 x 10–4, respectively). Grignard et al. (2012) have focused their attention on BPA and BPS for their estrogenic activity. Two different cell lines have been used for a luciferase report assay: MELN (MCF-7 cell derivates) and BG1 (human ovarian cancer cell line). The results have highlighted the same holds, showing a higher sensitivity of the tests with BG-1 cell lines compared to MELN tests. However, mean EC50 values have reported slightly dissimilar results. Using MELN cell lines, BPA and BPS have similar estrogenic potency, but a 105-fold less estrogenic activity compared to E2. On the contrary, using BG-1 cell lines, BPA showed a 10-fold higher potency than BPS. Kurutu-Niwa et al. (2005) have found comparable estrogenic activity between BPS and BPA using GFP expression system, respect to MELN cells experiment. More detailed information about in vitro/in vivo studies on endocrine activities of some BPs (BPA,

BPAF, BPF, BPS and BPB) could be found in a recently published review (Skledar and Mašič, 2016).

Related to our in silico prediction method, E2 has been predicted as higher interactor of BPA, placing the two BPA metabolites (BPA sulfate and BPA glucuronide) as well as BPS among those ones in the lower ranking. Indeed, BPS is among the most used chemical replacements of BPA based on its experimental lower ERα binding affinity (Blair et al., 2000; Kitamura et al., 2005; Kuruto-niwa et al., 2005; Grignard et al., 2012; Rochester and Bolden, 2015), and as such it has been judged a safer compound for human health. Thus, our computational approach appears to be a good in silico predictor of the already published in vitro data.

Taking into account ERβ, eight BPs have been predicted as higher interactors, six as medium and eleven as lower interactors (Table 2, Table 3). As in the case of ERα, E2 was predicted as a higher interactor compared to BPA. Based on the HintScore values, the computational ERβ binding affinity of BPs has been ranked as follows: BPM > BPP > MBP > E2 > BPG > BPBP > BPTMC > BPAP > TBBPA > BPAF > BPE > TCBPA > BPZ > BPC > BPA > BPB > BPA catechol > BPA quinone > BPPH > TMBPA > BPC2 > BPF > BPA_ol > BPA sulfate > BPA carboxylic > BPS > BPA glucuronide.

5.3.3 In silico prediction of endocrine disruptor activity of BPs against Estrogen Related Receptor γ (ERR γ)

Considering the Estrogen Related Receptor γ (ERRγ) as a target, six BPs were predicted as higher interactors, four as medium and fifteen as lower interactors (Table 2, Table 3). Considering the HintScore values, the computational ERRγ binding affinity of BPs has been ranked as follows: TCBPA > TBBPA > BPTMC > BPA catechol > BPAF > BPE > MBP > BPF > BPA glucuronide > BPA > BPB > BPC2 > BPBP > BPC > BPAP > BPZ > TMBPA > BPM > BPA quinone > BPA sulfate > BPA carboxylic > BPPH > BPP > BPS > BPG > E2 > BPA_ol.

Interestingly, E2 was found at the end of the ranking list of our dataset and that was completely in agreement with previous experimental data in which E2 has shown no ERRγ binding affinity, also at high doses (10 μM; Takayanagi et al., 2006). This phenomenon was reflected also in the specific value of the best HintScore value that was in the order of −661,3301, that means that no positive interactions were established between the protein and the ligand.

5.3.4 In silico prediction of endocrine disruptor activity of BPs against androgen receptors (ARs)

Considering the wild type AR as a target and the AR ligands testosterone (T), DHT, and di-hydroxy flutamide (2OH-FTA) as reference chemicals, only four bisphenols were predicted as higher interactors, six as medium and fifteen as lower interactor (Table 2, Table 3). The HintScore-based ranking of the AR binding affinity prediction of BPs was the following one: BPPH > BPM > BPA sulfate > BPA > BPAF > TBBPA > BPB > MBP > BPG > BPC > BPA catechol > BPE > BPZ > BPAP > TCBPA > TMBPA > BPF > BPTMC > BPP > BPA quinone > BPC2 > DHT > BPA ol > T > 2OH-FTA > BPS > BPBP > BPA carboxylic > BPA glucuronide.

Hence, BPA seems to be a strong AR interactor, more than the endogenous ligands T and DHT and even more than the anti-androgenic, pharmacological molecule 2OH-FTA. Indeed, looking at both the HintScore ranking and the BPA Relative Predicted Activity of BPs (listed in Table 2, Table 3), two BPA-like molecules, bisphenol M (BPM) and bisphenol PH (BPPH), and one BPA metabolite, BPA sulfate, results all as stronger than BPA as AR binders and a second BPA metabolite, MBP, has an high probability to be an AR binder.

On the other hand, in the HintScore ranking BPS closely resembles the pharmacological anti-androgenic drug 2OH-FTA as well similar binding energies to the two endogenous androgens T and DHT; whether these could relate to an androgen-like activity for BPs has to be demonstrated in biologically-relevant experimental models. Moreover, BPAF has a similar HintScore value compared to BPA (1228,022 and 1195,678 for BPA and BPAF, respectively).

This was in good agreement with the experimental data reported by Teng et al. for which BPA and BPAF were shown to have AR binding affinity. However, it is worth mentioning that no animal testing and in vivo results related to BPA effects on AR activity showed some discrepancies (Bonefeld-Jørgensen et al., 2007; Luccio-Camelo and Prins, 2011; Rubin, 2011).

Taking into account the two mutated ARs, the obtained computational data gave a very different situation. In the case of the mutated ART877A as a target, nine BPs were predicted as higher interactors, eleven as medium and five as lower interactors. The BPA Relative Predicted Activity of BPs (see Table 2) includes five BPs as stronger ART877A binders, namely bisphenol AP (BPAP), bisphenol C2 (BPC2), bisphenol E (BPE),

bisphenol G (BPG), bisphenol M (BPM), bisphenol P (BPP), bisphenol PH (BPPH). Hence, the HintScore ranking supports the fact that many BPs and BPA metabolites might be stronger binders of ART877A than BPA and with a strength similar to BPA, BPS and even more interesting to the endogenous androgens (T and DHT) and pharmacological antiandrogen 2OH-FTA. Furthermore, based on the HintScore values, the computational ART877A binding affinity of BPs has been ranked as follows: BPB > BPA catechol > MBP > BPM > BPG > BPZ > BPE > BPAP > BPF > BPP > TBBPA > BPPH > BPA > BPC2 > BPA glucuronide > 2OH-FTA > BPC > BPTMC > TCBPA > DHT > BPA sulfate > T > TMBPA > BPS > BPAF > BPA quinone > BPA carboxylic > BPA ol > BPBP.

Finally, looking at the case of the mutated ARW741L, eighteen BPs were predicted as higher interactors, four as medium and three as lower interactor (Table 2, Table 3). In this second case of a mutated AR, the computational AR^{W741L} binding affinity of BPs showed a further different ranking: BPPH > BPP > BPA catechol > TBBPA > MBP > TCBPA > BPM > BPAF > BPE > BPA glucuronide > BPC > BPF > BPA sulfate > 2OH FTA > TMBPA > BPB > BPAP > BPG > BPA > BPZ > T > BPC2 > BPTMC > BPBP > BPA ol > DHT > BPA quinone > BPA carboxylic > BPS.

Besides the BPs identified as strong binders of AR and ART877A, the ARW741L set of stronger binders includes all tested BPs except four of them, and other 2 different BPA metabolites, namely BPA catechol and BPA quinone.

Overall, a greater number of higher and/or medium interactors with the two mutated ARs were identified in comparison to the wild type form with the mutated ARW741L being the best recognized one (Table 2, Table 3). Hence, considering that both AR mutated forms brings a point mutation in the ligand binding domain (LBD), the computational prediction support the notion that altering the ligand binding pocket of the AR allows the interactions of chemicals unable to bind the wild type as it was shown in vitro for the ART877A (, 2005, 2006).

5.3.5 In silico prediction of a wider endocrine disrupting activity

Our consensus score prediction was able to correctly score the BPS as a lower interactor of ERα and ERβ, for which it is well known by in vitro experimental data that it has a weaker binding affinity compared to BPA. However, BPS has been predicted as a lower interactor for all other NRs except for ART877A, for which it was predicted as a medium interactor. This same latter prediction was found for one BPA metabolite, the BPA carboxylic.

Contrariwise, two investigated BPs have been showing a higher alert. One BPA metabolite, 4,4'-(1,1-dimethyl-3-methylene-1,3-propanediyl)bisphenol (MBP), was predicted as a higher interactor for all the six NRs, whereas bisphenol M (BPM) was predicted as a higher interactor for all the NRs considered, except for ERRγ, for which the consensus scoring classified it as medium interactor (Table 2). Thus, these two BPs appear to possess a wider endocrine disrupting activity in the human body far beyond the mere estrogenic-like one.

On the other hand, only one of the studied BPS, the BPA metabolite 2,2-bis(4-hydroxyphenyl)propanol (BPA_ol), was predicted as a lower interactor for all six NRs.

In our dataset, we have also included seven metabolites of BPA (BPA glucuronide, BPA sulfate, MBP, BPA_ol, BPA carboxylic acid, BPA cathecol and BPA quinone), for which the estrogenic and anti-androgenic activities were previously analysed, except for BPA quinone (Kitamura et al., 2005). Again, our computational predictions fit well with previously published in vitro experimental data. For example, the BPA_ol, BPA carboxylic acid and BPA catechol, that were previously annotated as weakly estrogenic and weakly or not anti-androgenic chemicals by in vitro studies, were computationally predicted as lower or medium interactors for both ERs and AR when compared to BPA (Table 4).

Table 4. Comparison of the endocrine disrupting activity of BPA metabolites by ERs and AR between *in vitro* experimental data* and the *in silico* approach of the present study.

BPA metabolites						
CAS	Name	ERα	ERβ	ER *in vitro* prediction*	AR_WT	AR *in vitro* prediction*
92549-67-2	BPA carboxylic	L	L	weakly estrogenic	L	non anti-androgenic
79371-66-7	BPA catechol	M	M	weakly estrogenic	M	weakly anti-androgenic
267244-08-6	BPA glucuronide	M	L	non estrogenic	L	non anti-androgenic
142648-65-5	BPA_ol	L	L	weakly estrogenic	L	non anti-androgenic
163405-36-5	BPA quinone	L	L	-	L	-
-	BPA sulfate	M	L	non estrogenic	H	non anti-androgenic
13464-24-9	MBP	H	H	strongly estrogenic	H	-

**Kitamura et al., 2005*

However, the common metabolites of BPA, the bisphenol A sulfate and BPA glucuronide, have been predicted as medium and/or lower interactor of ERα, ERβ except for AR for which BPA sulfate was judged as higher interactor. Anyway, the in vitro studies have highlighted their non-estrogenic and non-antiandrogenic activity. We have to remind that computational approaches are able to predict the interaction between a

compound and a receptor and not considering its biological activity, except in some specific cases. It is also true that in silico methods in the food science and safety are normally applied to screen a lot of compounds in a very fast and economic way, allowing also avoiding and limiting animal testing. In this scenario, we prefer to consider false positive than false negative.

5.3.6 Molecular dynamics of BPA, BPF and BPS

To evaluate the stability and the mechanism of interaction between different BPs with ERα, 100 ns of molecular dynamic (MD) simulations were carried out on three different complexes: ERα-BPA, ERα-BPS and ERα-BPF. To monitor the stability of each system, the root-main-square-deviation (RMSD), with respect to the initial structures, were analysed during the total simulation run. The first RMSD comparison is toward the protein backbone atoms, the second one towards the heavy atoms of BPA, BPS or BPF, and the last, more deeply, towards the backbone of the residues 5 Å around the ligands. The RMSD value of protein backbone increased slowly during the first steps of MD, reaching the equilibrium at about 50 ns for all the systems. After that, the RMSD fluctuation values were lower than 1 Å (Fig. 2A) showing the complexes achievement stability.

Conversely, the analyses of ligand heavy atoms RMSD remarked the conformational changes of BPs inside the binding pocket during the simulation time. It is important to remind that the simulation was initially performed using a strong constrain on the ligands, that was gradually decreased every 5 ns to reach slowly the equilibration phase, avoiding misleading trajectories of MD results. Thus, very low RMSD values have been recorded during the first 25 ns (Fig. 2B). After that, BPA and BPF were able to remain stable until 45–50 ns, showing more flexibility after that. Instead, the RMSD profile of BPS changed quickly (after the constrain removal) and this fluctuation is retained during the remaining simulation run. In fact, while BPA and, to a greater extent, BPF had constant little movement inside the binding pocket during MD (Fig. 3A and C), the oscillation observed for BPS is mainly due to a great movement of one of its phenyl groups after 50 ns (Fig. 3B). However, the observed fluctuation values are lower than 2 Å for all the complexes.

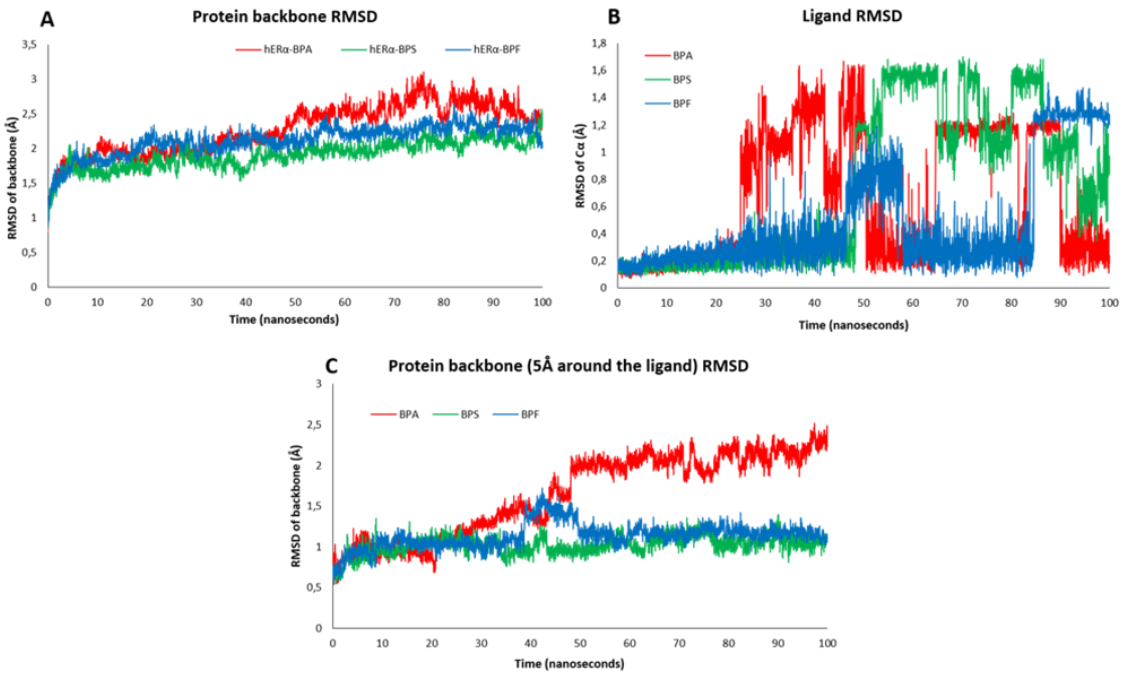


Figure 2: RMSD value of protein backbone (A), heavy atoms of the ligands (B) and the protein backbone 5Å around the ligands (C).

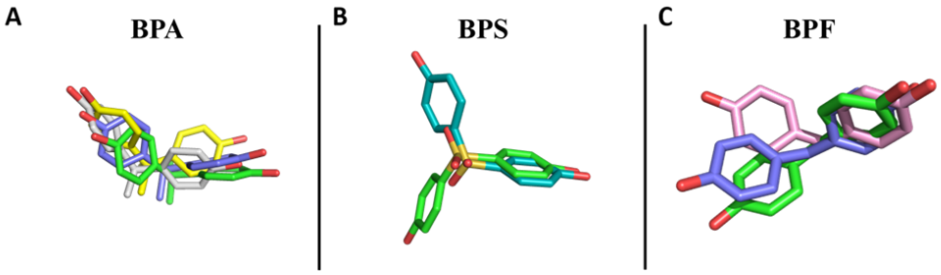


Figure 3: Movement of BPA (A), BPS (B), BPF (C) during the molecular dynamic simulations. Extracted snapshot: A) green, yellow, grey and violet refers to BPA at 0ns, 30ns, 50ns and 60ns, respectively; B) in green and in cyan is shown BPS at 0ns and 60ns, respectively; C) green, violet and pink refers to BPF at 0ns, 30ns and 50ns, respectively.

Moreover, as shown by the RMSD profile in Fig. 2C, the binding of BPS and BPF to ERα leads to an overall reduction in the conformational flexibility of the residues 5 Å around the ligand compared to BPA. This could be the reflection of more stable hydrogen bonds established between ERα and BPS or BPF than BPA (Fig. 4, Fig. 5 and Table 5).

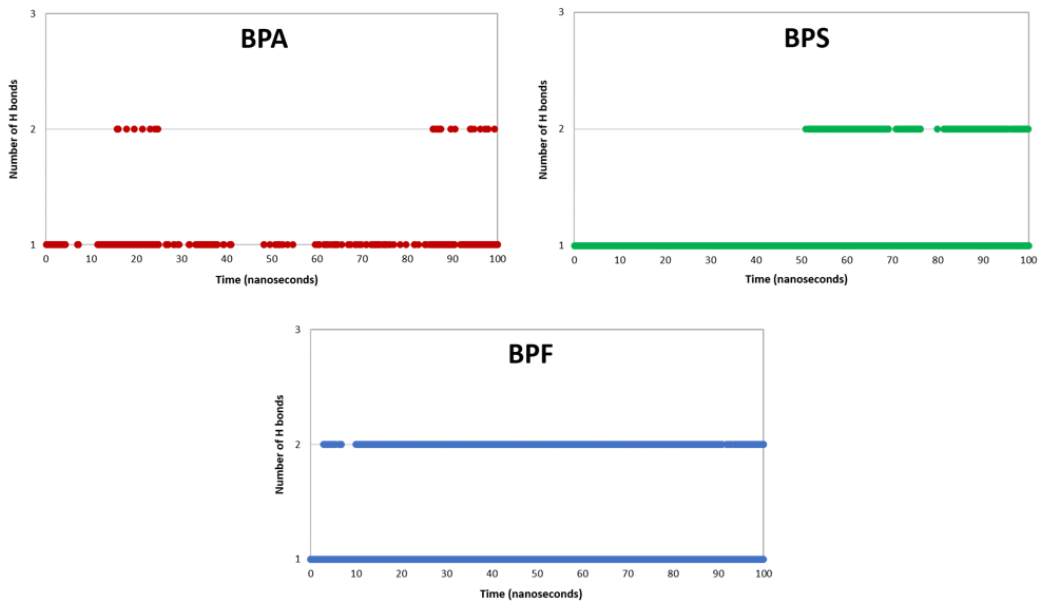


Figure 4: Numbers of hydrogen bonds established in the three complexes along the MD simulation time.

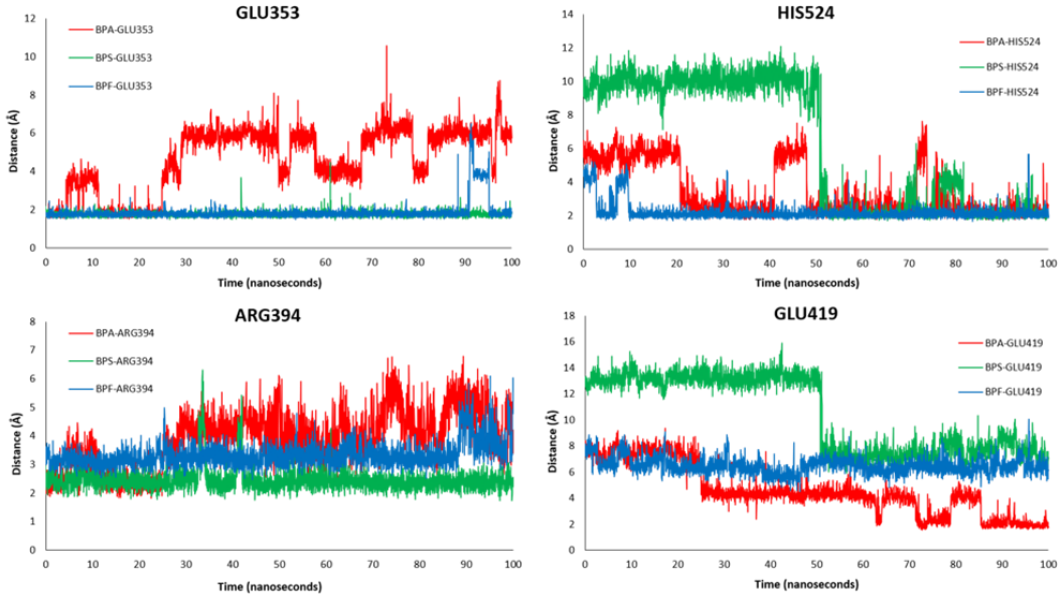


Figure 5: Distances between BPA, BPS or BPF and the residues of ERα involved in the hydrogen bond interactions with the ligands.

Table 5: Persistence time of hydrogen bonds interactions between the protein and BPA, BPS or BPF.

BPA		
Donor	*Acceptor*	*Residence time*
BPA	Glu353-side chain	7.30%
Arg394-side chain	BPA	0.28%
His524-side chain	BPA	3.73%
BPA	Glu419-main chain	5.22%

BPS		
Donor	*Acceptor*	*Residence time*
BPS	Glu353-side chain	71.52%
BPS	His524-side chain	19.60%
Arg394-side chain	BPS	0.02%
BPF		
Donor	*Acceptor*	*Residence time*
BPF	Glu353-side chain	69.85%
BPF	His524-side chain	34.38%

This phenomenon could be also the reflection of the hydrophobic contribution. In fact, the presence of the sulfone group in the BPS structure generates a higher number of negative hydrophobic interactions with the protein. Thus, this ligand could induce a more rigidity inside the binding pocket to avoid the formation of more negative interactions with the ligand. On the contrary, the absence of two methyl group in BPF (compared to BPA) decreases the number of positive hydrophobic interactions established with the receptor. Thus, since hydrogen bonds play a fundamental contribution for BPF binding, they should be stable during the simulation to avoid ligand unbinding. Again, this phenomenon could induce a more "freeze" conformation (inside the binding pocket) to maximize ligand interaction.

Comparing the formation and persistence of hydrogen bonds network between the ligands and the protein, it is worth to note that the interaction between BPA and ERα is characterised by two hydrogen bonds: one is almost distributed during the total simulation, and second one occurred only for few nanoseconds at the beginning and end of the simulation. Contrarywise, BPF was able to form two stable hydrogen bonds during the total simulation run as well as BPS after 50ns (Fig. 4).

In fact, the residues involved in the ligand binding during the simulation are different for the three complexes (Table 5). The interaction between the hydroxylic group of BPS and Glu353 is mostly stable for the total simulation run, contributing for the 71,52% of the time, followed by the interaction with His524 (19.60%), and to a lesser extent Arg394 (0.02%) and a similar persistence time of H bonds are established by BPF with Glu353 (69.85%) and His524 (34.38%). Instead, the hydrogen bonds established by BPA contributes for shorter period of time (Table 5), and an additional amino acid, Glu419, is contacted by BPA, establishing a hydrogen bond with 5.22% of residence time.

These interactions are perfectly explained by the graphs showing the distances between the residues of the protein and BPA, BPS and BPF (Fig. 5). It is evident that BPS and

BPF are able to form a stable hydrogen bond with Glu353 as opposed to BPA. Instead, the amino acid Arg394 shows a low contribution to the interaction for all the three complexes, and this trend is more relevant for BPA, since, as we can see in Fig. 5, the distance between BPA and Arg394 greatly increases after ~25 ns.

On the contrary, Glu419 is reached only by BPA and, by a visual inspection, this peculiar trend is induced by the different network of hydrogen bonds established by BPA, BPS or BPF with His524. In fact, while BPF is able to contact stably His524 and BPS after 50 ns, it is not the same for BPA: the latter form and break this bond during the evolution time. However, this induces an approach of BPA on the mainchain of Glu419, and in the last 15 ns, they form a stable hydrogen bond. It is important to note that Glu419 is a fundamental residue for establishing the correct hydrogen-bond network needed for maintaining the closed conformation of helix 12 (H12). In fact, when the endogenous ligand E2 is bound, His524 is positioned correctly for establishing two hydrogen bonds, one with the ligand and the other one with the carbonyl group of Glu419. This latter residue, in turn, forms a hydrogen bond network with the residues Glu339 (helix H3) and Lys531 (helix H11), allowing the formation of the active conformation. Probably, the closed conformation could be further stabilised by the formation of a hydrogen bond between Glu419 and BPA.

Thus, to evaluate if there are conformational changes in the position of helix 12, the initial coordinate (0 ns) and the final coordinate (100ns) of the receptor have been superimposed for each complex. As we can see in Fig. 6 A, B, C, no relevant conformational changes have been reported for both BPA, BPS and BPF. Only BPA complex (Fig. 6A) has shown a H12 little movement, approaching more closely to the receptor, in a more closed conformation. The differences in the helix 12 position have been shown in Fig. 6D, where we have reported the overlapping structures of BPA, BPS and BPF complex referring to the coordinates at 100 ns.

Given the capability of BPA to induce in vitro estrogenic activity, the results highlighted the importance of establishing hydrophobic interactions to bind and stimulate ERα in spite of the lower numbers of hydrogen bonds. In fact, taken into account the best Hint poses used as starting point structures, all three ligands are initially able to establish one hydrogen bond with one of their hydroxylic group and Glu353 and the differences into the ligand-binding mode are mainly due to the numbers of positive/negative hydrophobic interactions. In fact, while BPA is able to establish a wide number of positive hydrophobic

interactions, the presence of BPS sulfone group induces a great number of negative hydrophobic interactions, leading to a drastic drop of the total HintScore. Instead, the absence of two methyl group in the BPF structure decreases the number of positive hydrophobic interaction, and thus hydrogen bond and hydrophobic interactions contribute with the same extent to the HintScore. The capability to form negative and middle positive hydrophobic interactions of BPS and BPF, respectively, may provide an explanation of the estrogenic activity found experimentally (Kitamura et al., 2005; Fic et al., 2014). In fact, the greater number of negative hydrophobic interactions established by BPF and, to a greater extend, by BPS could be the cause of the greater reported binding site rigidity (Fig. 2C).

Based on these results and compared them with experimental data, we could speculate that the more rigidity induced by negative hydrophobic interactions could prevent the binding of different proteins (such as co-activators, co-repressors, translocators, etc) needed for the activity of nuclear receptor, or for its dimerization.

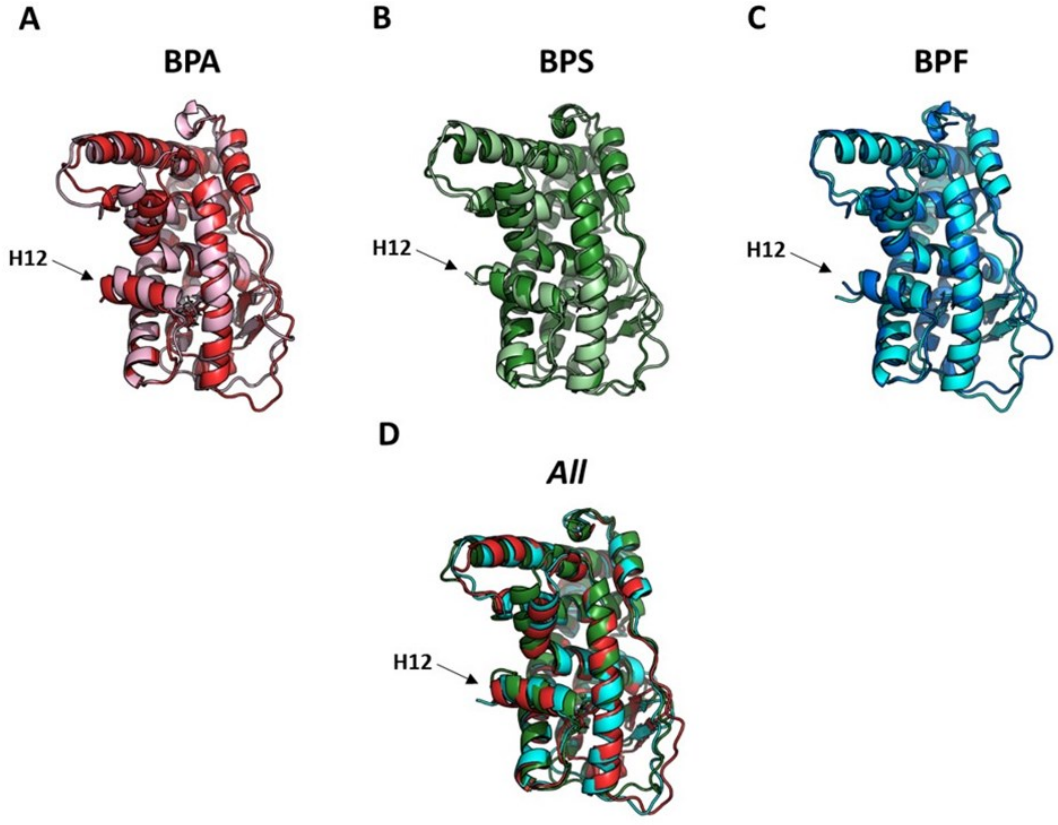


Figure 6: Alignment of the complex at 0 nanosecond and 100 nanoseconds for: (A) BPA (pink and red, respectively), (B) BPS (light green and dark green, respectively), and (C) BPF (blue and cyan, respectively). No relevant conformational changes of Helix 12 (H12) have been reported. Just BPA complex has a little movement of H12, that moved toward the receptor, in a more closed conformation. (D) Coordinate superimposition of BPA (red), BPS (green) and BPF (cyan) at 100 nanoseconds.

In addition to that, the root main square fluctuation (RMSF) of the three complexes was monitored to analyse the local mobility of protein residues. As shown in Fig. 7, the three complexes had a similar trend. However, in the region corresponding to the amino acids from 368 to 377 (helix 4, H4), a slightly greater fluctuation (close to 1 Å) was evident in BPA and BPF complexes compared to BPS.

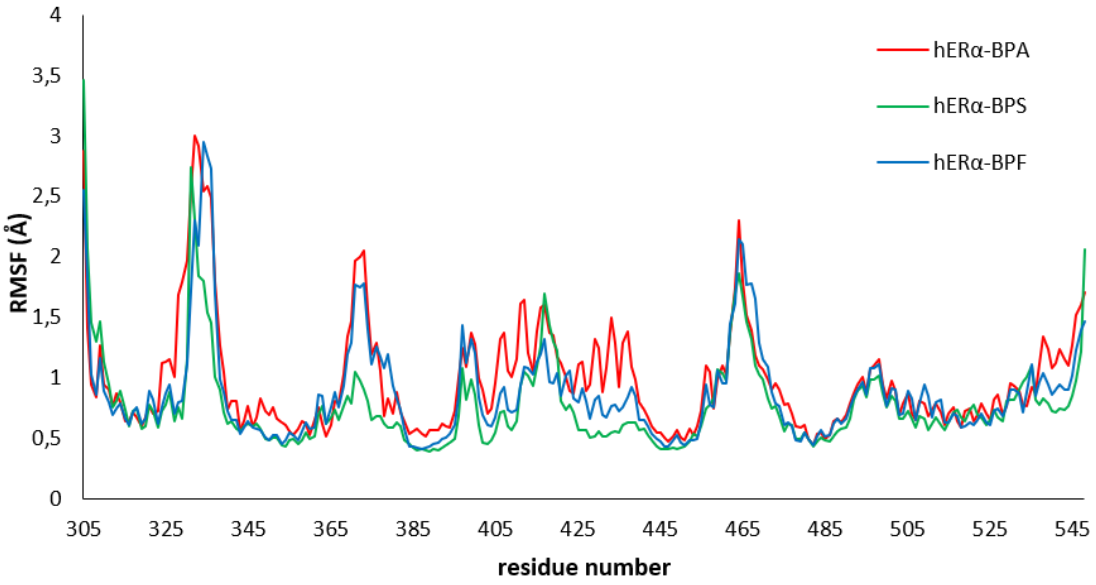


Figure 7: RMSF of the three complexes obtained by molecular dynamic simulation.

This greater value in the ERα-BPA complex is induced by the movement of almost four Å of H4, highlighting again that the complex with BPS had less flexibility than with BPA. The same goes for the range corresponding to the residues 421 to 440 (helix 8, H8).

5.4 Conclusions

The application of a non-statistical computational approaches, such as the simulation of a ligand-NR interaction by molecular docking, allowed us to demonstrate that using BPs and sex steroid receptors as case study, we could correctly predict the MIE and the relative binding affinities of BPA alternatives in comparison to the chemical to be substituted, BPA itself. The known data on BPs activities coming from in vitro experimental approaches (mainly gene transactivation or gene reporter assays) confirm the good reliability of our in silico prediction.

In particular, the estrogen-like in silico data obtained in this study could be summarized as follows: i) the computational ERα and ERβ binding affinities of BPs by molecular docking predicted BPS as a very worse interactor than BPA and E2, thus suggesting BPS as a safer alternative to BPA on the basis of the estrogen-like disrupting potential; ii) the two BPA metabolites, BPA sulfate and BPA glucuronide, also appeared as very worse interactors in comparison to both BPA and E2, suggesting that a fast metabolic turnover could lower the harmful effects of BPA exposure; iii) as expected, BPA resulted as a strong ERRγ interactor, whereas BPS closely resembled E2 that does not bind it at all; hence, BPS appeared as a safer chemical also in this case.

The apparent safer estrogen-like profile of BPS towards BPA is opposed by the obtained androgen-like data, in which we observed that: i) the computational AR binding affinities of BPs by molecular docking predicted BPA and some BPs as strong AR interactors, even better than endogenous androgens (a result in accordance with some but not all published in vitro data); ii) in accordance with previously published in vitro data, the mutated ARs are recognized by more BPs than the wild type AR; iii) BPS resulted as a weaker binder than BPA and androgens, although its binding affinity prediction closely resembles that one of the pharmacological anti-androgen 2OH-FTA. Hence, considering its androgen-like profile, BPS does not really appear a completely safer alternative to BPA.

Furthermore, the application of molecular dynamic simulations allowed to study the mechanism of binding of three different BPs (BPA, BPF and BPS) on ERα and to analyse the subsequent protein conformational changes. BPS and BPF form more stable hydrogen bonds compared to BPA, which binds the protein establishing mainly hydrophobic interactions. This phenomenon reduces protein and binding pocket flexibility in the presence of BPS and BPF. Taken together docking results and the information available by experimental data, BPs estrogenic activity is BPA > BPF > BPS. Thus, it seems that hydrophobic interactions are preferred to hydrogen bonds for the activation of the estrogenic activity. This could be probably related to the greater protein flexibility in presence of hydrophobic interactions, which could be necessary to allow the interaction with different proteins (such as co-activators) needed for the real activation of the receptor.

Overall, using molecular docking as an in silico approach for prioritization appears a cost-effective and timesaving screening tool for BPA alternatives. It also allows predicting NR

interactions not yet studied either in vitro or in vivo, focussing on the targets to be successively investigated.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

A special thanks to Prof. F.J. Luque, Barcelona University, for profitable comments on molecular dynamics.
This manuscript has been envisaged within the frame of the After-LIFE dissemination activities of the LIFE + granted project LIFE-EDESIA (LIFE12 ENV/IT/000633). FC is a recipient of a Ph.D. fellowship at University of Parma funded by the BIGDATA project of Regione Emilia-Romagna.

References

Babu, S., Vellore, N. A., Kasibotla, A. V., Dwayne, H. J., Stubblefield, M. A., & Uppu, R. M. (2012). Molecular docking of bisphenol A and its nitrated and chlorinated metabolites onto human estrogen-related receptor-gamma. Biochemical and Biophysical Research Communications, 426(2), 215–220. https://doi.org/10.1016/J.BBRC.2012.08.065

Bissantz, C., Folkers, G., Rognan, D. 2000. Protein-Based Virtual Screening of Chemical Databases. 1. Evaluation of Different Docking / Scoring Combinations. J. Med. Chem. 43, 4759-4767, doi: 10.1021/jm001044l.

Blair, R. M., Fang, H., Branham, W.S., Hass, B.S., Dial, S.L., Moland, C.L., Tong, W., Shi, L., Perkins, R., Sheehan, D.M., 2000. The Estrogen Receptor Relative Binding Affinities of 188 Natural and Xenochemicals : Structural Diversity of Ligands. Toxicol. Sci. 54, 138-153, doi: 10.1093/toxsci/54.1.138.

Bolli, A., Galluzzo, P., Ascenzi, P., Del Pozzo, G., Manco, I., Vietri, M.T., Mita, L., Altucci, L., Mita, D.G., Marino, M., 2008. Laccase Treatment Impairs Bisphenol A-Induced Cancer Cell Proliferation Affecting Estrogen Receptor a -Dependent Rapid Signals. IUBMB Life, 60, 843–852. doi: 10.1002/iub.130.

Bonefeld-Jørgensen, E.C., Long, M., Hofmeister, M.V., Vinggaard, A.M., 2007. Endocrine-disrupting potential of bisphenol A, bisphenol A dimethacrylate, 4-n-nonylphenol, and 4-n-octylphenol in vitro: new data and a brief review. Environ. Health Perspect. 115, Suppl 1, 69-76, doi: 10.1289/ehp.9368.

Charifson, P.S., Corkery, J.J., Murcko, M.A., Walters, W.P., 1999. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. J. Med. Chem. 42, 5100-5109, doi: 10.1021/jm990352k.

Chu, S., Haffner, G.D., Letcher, R.J., 2005. Simultaneous determination of tetrabromobisphenol A, tetrachlorobisphenol A, bisphenol A and other halogenated analogues in sediment and sludge by high performance liquid chromatography-electrospray tandem mass spectrometry. J Chromatogr A. 1097, 25-32, doi: 10.1016/j.chroma.2005.08.007.

Conroy-Ben, O., Garcia, I., & Teske, S. S. (2018). In silico binding of 4,4'-bisphenols predicts in vitro estrogenic and antiandrogenic activity. Environmental Toxicology, 33(5), 569–578. https://doi.org/10.1002/tox.22539.

Cozzini, P., Dellafiora, L., 2012. In silico approach to evaluate molecular interaction between mycotoxins and the estrogen receptors ligand binding domain: A case study on zearalenone and its metabolites. Toxicol. Lett. 214, 81-85, doi: 10.1016/j.toxlet.2012.07.023.

Darnerud, P.O., 2003. Toxic effects of brominated flame retardants in man and in wildlife. Environ. Int. 29, 841-853, doi: 10.1016/S0160-4120(03)00107-7.

De Falco, M., Forte M., Laforgia, V., Dietrich, D., Costa, M., 2015. Estrogenic and anti-androgenic endocrine disrupting chemicals and their impact on the male reproductive system. Frontiers in Environmental Science, 3, 1–12. doi: 10.3389/fenvs.2015.00003.

de Wit, C. A., Herzke, D. and Vorkamp, K., 2010. Brominated flame retardants in the Arctic environment — trends and new candidates. Science of The Total Environment. Elsevier, 408, 2885-2918. doi: 10.1016/J.SCITOTENV.2009.08.037.

Delfosse, V., Grimaldi, M., Pons, J.-L., Boulahtouf, A., le Maire, A., Cavailles, V., Balaguer, P. (2012). Structural and mechanistic insights into bisphenols action provide guidelines for risk assessment and discovery of bisphenol A substitutes. Proceedings of the National Academy of Sciences, 109(37), 14930 LP – 14935. https://doi.org/10.1073/pnas.1203574109.Dellafiora, L., Galaverna, G., Dall'Asta, C., Cozzini P., 2015. Hazard identification of cis/trans-zearalenone through the looking-glass. Food Chem. Toxicol. 86, 65-71, doi: 10.1016/j.fct.2015.09.009.

Dellafiora, L., Ruotolo, R., Perotti, A., Cirlini, M., Galaverna, G., Cozzini, P., Buschini, A., Dall'Asta, C., 2017. Molecular insights on xenoestrogenic potential of zearalenone-14-glucoside through a mixed in vitro/in silico approach. Food Chem. Toxicol. 108, 257-266, doi: 10.1016/j.fct.2017.07.062.

Diamanti-Kandarakis, E., Bourguignon, J.P., Giudice, L.C., Hauser, R., Prins, G.S., Soto, A.M., Zoeller, R.T., Gore, A.C., 2009. Endocrine-disrupting chemicals: an Endocrine Society scientific statement. Endocr. Rev. 30, 293-342. doi: 10.1210/er.2009-0002.

Duan, Y., Yao, Y., Wang, B., Han, L., Wang, L., Sun, H., Chen, L., 2018. Association of urinary concentrations of bisphenols with type 2 diabetes mellitus: A case-control study. Environ. Pollut. 243, 1719-1726, doi: 10.1016/j.envpol.2018.09.093.

EFSA CEF Panel (EFSA Panel on Food Contact Materials, Enzymes, Flavourings and Processing Aids), 2015. Scientific Opinion on the risks to public health related to the

presence of bisphenol A (BPA) in foodstuffs: Executive summary. EFSA J. 13, 3978. doi:10.2903/j.efsa.2015.3978.

Ehrlich, V.A., Dellafiora, L., Mollergues, J., Dall'Asta, C., Serrant, P., Marin-Kuan, M., Lo Piparo, E., Schilter, B., Cozzini, P., 2015. Hazard assessment through hybrid in vitro / in silico approach : the case of zearalenone. ALTEX. 32, 275-286. doi: 10.14573/altex.1412232.

Fenichel, P., Chevalier, N., Brucker-Davis, F., 2013. Bisphenol A: an endocrine and metabolic disruptor. Ann. Endocrinol. (Paris). 74, 211-220. doi: 10.1016/j.ando.2013.04.002.

Fic, A., Žegura, B., Gramec, D., Mašič, L. P., 2014. Estrogenic and androgenic activities of TBBA and TBMEPH, metabolites of novel brominated flame retardants, and selected bisphenols, using the XenoScreen XL YES/YAS assay. Chemosphere. 112, 362–369. doi: 10.1016/J.CHEMOSPHERE.2014.04.080.

Ginex, T., Dall'Asta, C. and Cozzini, P., 2014. Preliminary hazard evaluation of androgen receptor-mediated endocrine-disrupting effects of thioxanthone metabolites through structure-based molecular docking. Chemical Research in Toxicology, 27, 279–289. doi: 10.1021/tx400383p.

Gore, A. C., Chappell, V. A., Fenton, S. E., Flaws, J. A., Nadal, A., Prins, G. S., Toppari, J., Zoeller, R. T., 2015. EDC-2: The Endocrine Society's Second Scientific Statement on Endocrine-Disrupting Chemicals', Endocrine reviews. 2015/11/06. Endocrine Society, 36(6), pp. E1–E150. doi: 10.1210/er.2015-1010.

Gramec Skledar, Darja and Lucija Peterlin Mašič. 2016. "Bisphenol A and Its Analogs: Do Their Metabolites Have Endocrine Activity?" Environmental Toxicology and Pharmacology 47:182–99.

Grignard, E., Lapenna, S. and Bremer, S., 2012. Weak estrogenic transcriptional activities of Bisphenol A and Bisphenol S. Toxicology in Vitro. Elsevier Ltd, 26, pp. 727–731. doi: 10.1016/j.tiv.2012.03.013.

Hess-Wilson, J. K., Webb, S. L., Daly, H. K., Leung, Y., Boldison, J., Comstock, C. E. S., Sartor, M. A., Ho, S., Knudsen, K. E., 2007. Unique Bisphenol A Transcriptome in Prostate Cancer : Novel Effects on ER β Expression That Correspond to Androgen Receptor Mutation Status. Environmental Health Perspectives, 115, pp. 1646–1653. doi: 10.1289/ehp.10283.

Kallen, J., Schlaeppi, J., Bitsch, F., Filipuzzi, I., Schilb, A., Riou, V., Graham, A., Strauss, A., Geiser, M., Fournier, B., 2004. Evidence for Ligand-independent Transcriptional Activation of the Human Estrogen-related Receptor α (ERRα). Journal of Biological Chemistry, 279(47), pp. 49330–49337. doi: 10.1074/jbc.m407999200.

Kellogg, G. E. and Abraham, D. J., 2000. Hydrophobicity: is LogPo/w more than the sum of its parts?. European Journal of Medicinal Chemistry, 35(7-8), pp. 651-661. doi: 10.1016/S0223-5234(00)00167-7

Kellogg, G. E., Burnett, J., Abraham, D. J., 2001. Very empirical treatment of solvation and entropy: A force field derived from Log Po/w. Journal of Computer-Aided Molecular Design, 15(4), pp. 381-393. doi: 10.1023/A:1011136228678

Kitamura, S., Suzuki, T., Sanoh, S., Kohta, R., Jinno, N., Sugihara, K., Yoshihara, S., Fujimoto, N., Watanabe, H., Ohta, S., 2005. Comparative Study of the Endocrine-Disrupting Activity of Bisphenol A and 19 Related Compounds. Toxicological Sciences, 84(2), pp. 249–259. doi: 10.1093/toxsci/kfi074.

Kuruto-niwa, R., Nozawa, R., Miyakoshi, T., Shiozawa, T. and Terao, Y., 2005. Estrogenic activity of alkylphenols , bisphenol S , and their chlorinated derivatives using a GFP expression system. Environmental Toxicology and Pharmacology, 19, pp. 121–130. doi: 10.1016/j.etap.2004.05.009.

Lanlan, L., Wang, Q., Zhang, Y., Niu, Y., Yao, X., & Liu, H. (2015). The Molecular Mechanism of Bisphenol A (BPA) as an Endocrine Disruptor by Interacting with Nuclear Receptors: Insights from Molecular Dynamics (MD) Simulations. PLOS ONE, 10, e0120330. https://doi.org/10.1371/journal.pone.0120330.

Liao, C., Liu, F. and Kannan, K., 2012. Bisphenol S, a New Bisphenol Analogue, in Paper Products and Currency Bills and Its Association with Bisphenol A Residues. Environmental Science & Technology. American Chemical Society, 46(12), pp. 6515–6522. doi: 10.1021/es300876n.

Lorenzetti, S. and Cozzini, P., 2017. The Substitution Principle within the REACH Regulation: Nuclear Receptor-Bound Endocrine Disruptors. Nuclear Receptor Research, 4, pp. 1–11. doi: 10.11131/2017/101205.

Lorenzetti, S., Marcoccia, D. and Mantovani, A., 2015. Biomarkers of effect in endocrine disruption: how to link a functional assay to an adverse outcome pathway. Ann Ist Super Sanità, 51(2), pp. 167–171. doi: 10.4415/ANN_15_02_16 Biomarkers.

Lu, L., Zhan, T., Ma, M., Xu, C., Wang, J., Zhang, C., Zhuang, S. (2018). Thyroid Disruption by Bisphenol S Analogues via Thyroid Hormone Receptor β: in Vitro, in Vivo, and Molecular Dynamics Simulation Study. Environmental Science & Technology, 52(11), 6617–6625. https://doi.org/10.1021/acs.est.8b00776

Luccio-Camelo, D. C. and Prins, G. S., 2011. Disruption of androgen receptor signaling in males by environmental chemicals. The Journal of Steroid Biochemistry and Molecular Biology. Pergamon, 127(1–2), pp. 74–82. doi: 10.1016/J.JSBMB.2011.04.004.

McCrea, E., Sissung, T. M., Price, D. K., Chau, C. H. and Figg, W. D. , 2016. Androgen receptor variation affects prostate cancer progression and drug resistance. Pharmacological Research. Academic Press, 114, pp. 152–162. doi: 10.1016/J.PHRS.2016.10.001.

Patisaul, H. B., 2019. Achieving CLARITY on bisphenol A, brain and behaviour. Journal of Neuroendocrinology. John Wiley & Sons, Ltd (10.1111), 0(0), p. e12730. doi: 10.1111/jne.12730.

Pjanic, M., 2017. The role of polycarbonate monomer bisphenol-A in insulin resistance. PeerJ, 5, p. e3809. doi: 10.7717/peerj.3809.

Prins, G. S., Patisaul, H. B., Belcher, S. M. and Vandenberg, L. N., 2018. CLARITY-BPA academic laboratory studies identify consistent low-dose Bisphenol A effects on multiple organ systems. Basic & Clinical Pharmacology & Toxicology. John Wiley & Sons, Ltd (10.1111), 0(0). doi: 10.1111/bcpt.13125.

Rochefort, H., 2017. Endocrine disruptors (EDs) and hormone-dependent cancers: Correlation or causal relationship?. Comptes Rendus Biologies. Elsevier Masson, 340(9–10), pp. 439–445. doi: 10.1016/J.CRVI.2017.07.007.

Rochester, J. R. and Bolden, A. L., 2015. Bisphenol S and F : A Systematic Review and Comparison of the Hormonal Activity of Bisphenol A Substitutes. Environmental Health Perspectives, 123, pp. 643–650.

Rubin, B. S., 2011. Bisphenol A: An endocrine disruptor with widespread exposure and multiple effects. The Journal of Steroid Biochemistry and Molecular Biology. Pergamon, 127(1–2), pp. 27–34. doi: 10.1016/J.JSBMB.2011.05.002.

Sharma, S., Ahmad, S., Khan, M. F., Parvez, S. and Raisuddin, S., 2018. In silico molecular interaction of bisphenol analogues with human nuclear receptors reveals

their stronger affinity vs. classical bisphenol A. Toxicology Mechanisms and Methods, 28: 660-669. doi: 10.1080/15376516.2018.1491663

Sifakis, S., Androutsopoulos, V. P., Tsatsakis, A. M. and Spandidos, D. A., 2017. Human exposure to endocrine disrupting chemicals: effects on the male and female reproductive systems. Environmental Toxicology and Pharmacology. Elsevier, 51, pp. 56–70. doi: 10.1016/J.ETAP.2017.02.024.

Takayanagi, S., Tokunaga, T., Liu, X., Okada, H., Matsushima, A., Shimohigashi, Y., 2006. Endocrine disruptor bisphenol A strongly binds to human estrogen-related receptor γ (ERRγ) with high constitutive activity. Toxicology Letters, 167(167), pp. 95–105. doi: 10.1016/j.toxlet.2006.08.012.

Teng, C., Goodwin, B., Shockley, K., Xia, M., Huang, R., Norris, J., Merrick, B. A., Jetten, A. M., Austin, C. P., Tice, R. R., 2013. Bisphenol A affects androgen receptor function via multiple mechanisms', Chemico-Biological Interactions. Elsevier Ireland Ltd, 203(3), pp. 556–564. doi: 10.1016/j.cbi.2013.03.013.

Wetherill, Y. B., Fisher, N. L., Staubach, A., Danielsen, M., de Vere White, R. W., Knudsen, K. E., 2005. Xenoestrogen Action in Prostate Cancer: Pleiotropic Effects Dependent on Androgen Receptor Status. Cancer Research, 65(1), pp. 54 LP – 65. Available at: http://cancerres.aacrjournals.org/content/65/1/54.abstract.

Wetherill, Y. B., Hess-Wilson, J. K., Comstock, C. E. S., Shah, S. A., Buncher, C. R., Sallans, L., Limbach, P. A., Schwemberger, S., Babcock, G. F., Knudsen, K. E., 2006. Bisphenol A facilitates bypass of androgen ablation therapy in prostate cancer. Molecular Cancer Therapeutics, 5(12), pp. 3181 LP – 3190. doi: 10.1158/1535-7163.MCT-06-0272.

Wetherill, Y. B., Petre, C. E., Monk, K. R., Puga, A., Knudsen, K. E., 2002. The Xenoestrogen Bisphenol A Induces Inappropriate Androgen Receptor Activation and Mitogenesis in Prostatic Adenocarcinoma Cells. Molecular Cancer Therapeutics, 1, pp. 515–524.

Wu, L.H., Zhang, X., Wang, F., Gao, C., Chen, D., Palumbo, J. R., Guo, Y., Zeng, E. Y., 2018. Occurrence of bisphenol S in the environment and implications for human exposure: A short review. Science of The Total Environment. Elsevier, 615, pp. 87–98. doi: 10.1016/J.SCITOTENV.2017.09.194.

Zhang, N. and Zhong, R., 2010. Structural basis for decreased affinity of Emodin binding to Val66-mutated human CK2 α as determined by molecular dynamics. Journal of Molecular Modeling, pp. 771–780. doi: 10.1007/s00894-009-0582-2.

CHAPTER 6

Daphnia Magna and in silico prediction to understand the binding of pyriproxyfen and its metabolite on RXR

Manuscript in preparation

Preface

Molecular docking and molecular dynamic techniques are two strong computational supports that allow the screening of huge quantities of chemicals. The power of *in silico* methods is not limited to this, they provide strong support in the 3R context. The concept was introduced in 1959 with the aim to reduce and replace *in vivo* animals tests. In particular, the analyses carried out in the present chapter focused on the study of the endocrine-disrupting properties of some relevant pesticides on *Daphnia Magna*, which is a microcrustacean widespread in freshwater. It plays an important role in the aquatic ecosystem, and it is largely used as an organism model to evaluate the presence of water contaminants. The computational approach used in the present work provides precious insights on the mechanism of action of some pesticides on the retinoid X receptor of *Daphnia Magna*. In more detail, the pesticide pyriproxyfen and its metabolite have been predicted as strong interactors of the receptor. To better understand the outcome of this pesticide, *in vivo* analyses are carrying out. However, since the experimental analyses are still ongoing, their results are not included yet.

The following study has been done in collaboration with Prof. Annamaria Buschini of the University of Parma and with the Prof. F. Javier Luque of the University of Barcelona.

The results of the present study will be published as a manuscript. For that reason, the following paragraphs are presented as a scientific research paper.

6.1 Introduction

Daphnia Magna is a microcrustacean, commonly known as water flea, widespread in freshwater. It plays a key role in the aquatic ecosystem since it is a primary food source for fish and other invertebrate predators (Zhang et al., 2016). Furthermore, *Daphnia* is an important organism model, usually used to evaluate the presence of contaminants in the water, by *in vivo* experiments, due to its high sensibility to a wide range of pollutants and environmental stressors. Indeed, significant reproductive decline, aberrant vertical mobility, behavioral pattern, and lastly phenoplasticity in daphnids, can imply the presence of chemical substances, synthetic hormones, changes in the water environment (acidity, salinity, calcium levels), the presence of bacterial pathogens (Kim et al., 2015). Since 1944, this crustacean has been used to assess toxicity substances in industrial wastewater (Tkaczyk et al., 2021). Today, *Daphnia magna* is adopted as species test for acute and chronic ecotoxicity by the Organization for Economic Co-operation and Development (OECD), which proposed two different accepted guideline tests: the acute immobilisation test (Test No. 202) used to assess effect of chemicals on the daphnids mobility (*Test No. 202: Daphnia sp. Acute Immobilisation Test*, 2004), and the reproduction test (Test No. 211) used to evaluate the effect of chemicals on the reproductive output (*Test No. 211: Daphnia magna Reproduction Test*, 2012). Reproduction strategy changes adopted by *Daphnia magna,* during ecotoxicological tests, is considered an important endpoint as response to the environment stressor factors. Under suitable environments, this organism undergoes a parthenogenesis reproduction, leading to identical female progenies. However, in response to unfavorable environmental and biological factors (photoperiodic changes, low temperature, overpopulation, and lack of nutrients), *Daphnia magna* can switch to sexual reproduction heading to males' production in offspring. Some evidences (Abe et al., 2015; Olmstead and Leblanc, 2002) highlight a possible role of methyl farnesoate (MF), a precursor of juvenile hormone (JH), in stimulating male sex production . The exposure of daphnids to high concentrations of MF, results in a proportional increase in male offspring production (Camp et al., 2019). Initially, the retinoid X receptor (RXR), a nuclear receptor, was proposed as a target of MF. Recently, Kakaley et al. have demonstrated that MF binds and activates the Methoprene Tolerant (Met) protein, a transcription factor belonging to the basic helix-loop-helix (bHLH)/PAS family proteins (Kakaley et al., 2017).

Nuclear receptors (NRs) are transcription factor proteins composed by two different domains: a Ligand Binding Domain (LBD) and DNA Binding Domain (DBD). The LBD is the protein region that interacts with hormones. As a result, the protein-ligand interaction promotes some conformational changes leading to the agonistic form of the nuclear receptor. The complex migrates inside the nucleus where the DBD recognizes specific sequences in the promoter regions of genes. This binding recruits other proteins on the DNA which untimely determine the chromatin changes, needed to regulate gene transcription.

Normal activity of endocrine system can be modified by chemicals that work as endocrine disruptor compounds (EDCs). EDCs can affect the normal activity of the endocrine system in many organisms, leading to severe diseases, caused by changes at the gene transcription levels, by direct binding to nuclear receptors or indirectly modifying their activities.

This is possible because some EDCs have physical-chemical properties close to natural hormones of NRs and thus can directly interact with them, mimicking the natural ligand or occupying the ligand-binding pocket, inhibiting the interaction with the hormone. The interaction between NRs and ligands promotes the proteins activity inducing alteration in the physiological processes of the cells.

Many chemicals have been identified as endocrine disruptors, among them, several are pesticides (Mnif et al., 2011). Pesticides are undoubtedly important molecules to protect plants from weeds, fungus, insects, and rodents (Spaggiari et al., 2021). However, these compounds can last in time in the air, soil, and water, (Bilal et al., 2019) determining a constant exposure of the organisms to pesticides. Exposure to pesticides can often be dangerous for human health. A famous example is the case of DDT , used to fight malaria since 1939 and prohibited in many countries in 1972 because of its cancerogenic effects and the impact on the ecosystem.

Pyriproxyfen (PP) is an insect growth regulator (IGR) used as an active ingredient of a lot of insecticides. It is highly used in the agriculture, but the exact tonnage data are confidential (https://echa.europa.eu/it/substance-information/-/substanceinfo/100.102.814). PP, along with methoprene, is a juvenile hormone analog (JHA) closely related to the retinoic acid in mammals. It was firstly introduced to the US in 1996 to protect crops against whitefly and it is now used for different agricultural applications to control many insect species. By mimicking natural insect juvenile

hormones (JH), JHAs prevent the development of insect larvae into mature adults. In 2006 Oda et al. have evaluated the effect of pyriproxyfen, epofenonane, and fenoxycarb on three genetically different strains of *Daphnia magna*, finding that all these three pesticides were able to decrease the total number of neonates and/or to increase the production of males' offspring. Jordão et al. (Jordão et al., 2016) found that PP alters lipids storage levels in a concentration-dependent fashion, impairs molt of exposed females, decreases body length of exposed females, the size of their first brood, changes offspring sex, and increases offspring size. It was also highlighted that pyriproxyfen increases males production (Ginjupalli and Baldwin, 2013). From the studies performed, PP seems to be involved into two different pathways regulated by different nuclear receptors. Wang and LeBlanc found that PP is able to activate gene transcription associated with retinoid-X-receptor and ecdysone receptor which forms an RXR:EcR heterodimer (Wang and LeBlanc, 2009). However, PP is not able alone to activate the gene transcription and the presence of EcR seems to be fundamental. A possible role of PP in the Met pathway was suggested by Miyakawa et al., although the EC_{50} values obtained from gene transcription tests were not consistent between *Daphnia magna* and *Daphnia pulex*. Recently, it has also been predicted the interaction of pyriproxyfen with the retinoid X receptor in humans and *Apis mellifera* (Spaggiari et al., 2021). In addition, pyriproxyfen has low persistence and it is rapidly degraded in 4-OH'-pyriproxyfen ((EFSA), 2009). To date, no one has been analyzed the effect of this metabolite on *Daphnia magna*.

Retinoid X receptor is an important member of the NRs superfamily present in many species of the Metazoa. In humans, once bound to its natural ligand (9-cis retinoic acid), RXR interacts with other NRs forming homodimeric or heterodimeric complexes. It is involved in the regulation of gene transcriptions in many biological processes like metabolism, cellular differentiation, apoptosis, and development. In *Daphnia magna*, RXR acts as partners of another nuclear receptor, EcR, to form the heterodimeric complex RXR-ECR which is involved in the regulation of metamorphosis, embryonic development, and reproduction processes (Wang and LeBlanc, 2009). The sequence of RXR results really conserved. For example, the sequence of the ligand binding domain of *Daphnia magna* shares a rate of similarity of 75%.

The aim of the study is to shed light about the possible protein targets of some pesticides involved in male production. Although the role of some pesticides in interacting with Met

protein has been suggested, recently (see above), RXR has been proposed as an additional hypothetical target of pesticides. In fact, the effect of contaminants rarely is made by the interaction of these compounds with a single receptor. On the contrary, there is often a synergic effect with different proteins that leads to health injuries.

In this study, we have studied the interaction between eight different pesticides (Figure 1) against RXR using computational methods. In more detail, homology modelling, molecular docking, consensus scoring, and molecular dynamic techniques has been used.

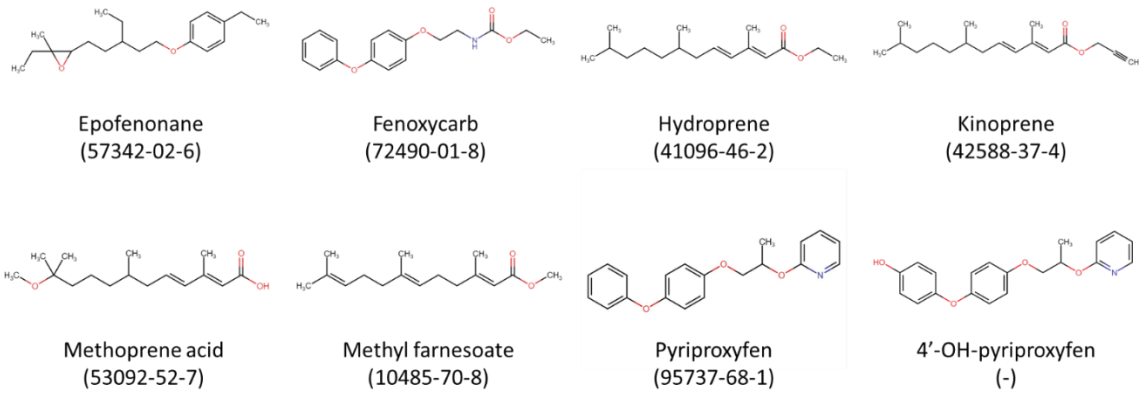


Figure 1: Two-dimentional structures of the compounds considering in the present work.

Experimental tests, like *in vivo* and *in vitro* tests, are often used to evaluate the effects caused by exposure to chemicals on the biomolecular pathway regulated by the interactions among molecules. Many techniques, used in the laboratories, allow defining a potential chemical's toxicity activity, but the use of these approaches is often complicated, expensive, and time-consuming. For these reasons, the use of this approach can be not efficient to assess a large number of chemicals. Otherwise, the use of computational approaches allows to perform virtual simulations and evaluate interactions between proteins and chemicals, predicting a potential endocrine disruptors activity. With the use of virtual predictions, it is possible to assess a great number of molecules, reducing the experimental tests required.

In this research, computational approaches have been used to investigate the interaction between RXR and agricultural pesticides. In particular, the potential interaction between *Daphnia magna* RXR with PP and its metabolite, 4-OH-pyriproxyfen has been investigated. The potential endocrine disruptors activity on *Daphnia magna* has been further evaluated by performing toxicity tests.

The use of computational approaches has allowed us to evaluate the potential endocrine disruptors activity of chemicals by virtual simulations, reducing the number of laboratory tests. The tests performed in laboratory can be very useful to evaluate the effect of chemicals at genes transcription level and to study the interactions among molecules, but are often complicated, costly, and time-consuming. For this reason, the use of in vivo test is not efficient to assess large numbers of agrochemicals. By using computational techniques, it has been possible to investigate the interaction of RXR protein with agricultural pesticides. Moreover, since pyriproxyfen is widely used in the agriculture, we have also evaluated its effect using toxicity test on *Daphnia magna*.

6.2 Material and Methods

6.2.1 Homology modeling

The *D. magna* amino acids sequence was retrieved from the Universal Protein Resource (UniProt)(UniProt Consortium, 2018) in the FASTA format (ID: A1XQQ1). Three different servers have been used to model the three-dimensional structure of the RXRα ligand binding domain of *Daphnia Magna*: SWISS-MODEL(Waterhouse et al., 2018), Phyre2 and I-TASSER.

SWISS-MODEL was used starting from the amino acid sequence in FASTA format considering the region of the protein corresponding to the ligand binding domain (amino acids: 174-397). Using two database searching methods (BLAST (Altschul et al., 1997; Camacho et al., 2009) and HHblits (Remmert et al., 2011)), the server has returned fifty different templates. Three different models have been built choosing the three best templates according to the template resolution and the sequence coverage: 1FBY (2.25 Å), 3OZJ (2.1 Å) and 1FM6 (2.1 Å), all of them belonging to the RXRα of Homo Sapiens. To give an estimation of model accuracy, SWISS-MODEL relies on two different scores: the QMEAN scoring function (Benkert et al., 2011) and the GMQE (Global Model Quality Estimation)(Biasini et al., 2014) values that have been used to evaluate the three models. The former value provides a geometrical quality evaluation using statistical potentials of mean force and generates both a global and a per residue value. QMEAN values below to zero indicate reliable models, instead values below or lower -4 are related to those structures with a low reliability. On the other side, the GMQE value estimates the tertiary structure accuracy and its value ranges from 0 to 1, where greater values are related to model with a robust reliability. Phyre2(Kelley et al., 2015) scans the input sequence against a sequence database using HHblits to predict the secondary structure

through PSIPRED (Lobley et al., 2009). This profile is converted to a hidden Markov model (HMM) to generate a query-template alignment against a precompiled database of HMMs of known 3D structures. The alignment is used to generate the structure of the backbone protein without examining side chains. Insertion and deletion regions are resolved considering a library of fragments (2-15 amino acids) of known protein structures looking at the geometry of flanking regions and distance between endpoints. These fragments are fitted to the rough model, ranked according to an empirical energy terms and the top scoring models are selected. After that, the program fits side chains to the backbone using a library of side chain rotamer to place side chains in their most probable rotamer in the input structure. I-TASSER server(Yang and Zhang, 2015) uses a hierarchical template-based method for protein structure prediction. In contrast to the other two servers used in the present study, it allows to specify the template from which the structure will be modelled. Thus, we have decided to model the receptor starting from the crystal structure 2P1T (R=1.8 Å) of human RXRα since it had the best resolution among the available PDB structures. I-TASSER provides two scores: a confidence score (C-score) and the template modeling score (TM-score)(Zhang and Skolnick, 2004) to estimate the models' accuracy. The value of C-score lies between [-5, 2], where models with greater values have better confidence. Instead, the value of the TM-score lies between [0, 1], where better templates have higher TM-scores. In more detail, values greater than 0.5 indicate a structure with a correct topology, whereas values >0.17 are for those structures with a casual similarity.

6.2.2 Evaluation of homology models

Five different models have been analysed using two different serves: PROCHECK (Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, 1993; Laskowski et al., 1996) and ProSA(Sippl, 1993; Wiederstein and Sippl, 2007). PROCHECK allows to assess the stereochemistry quality of a given protein structure considering both the backbone and the sidechains of residue's atoms. The server calculates the φ-ψ torsional angles of atoms and compared them with the Ramachandran plot obtained from a database of well refined high-resolution structures. A good model has a higher number of φ-ψ dihedral angles in favourable regions of the Ramachandran graph(Laskowski et al., 2012). The software also returns the overall G factor value: a low G-factor indicates a low-probability conformation since residues falls in disallowed regions of the Ramachandran plot. Generally, an acceptable G-factor value is greater than -0.5. The three models have

been also evaluated energetically using the ProSA-web service. The server returns the z-score value that is an overall score for evaluating protein structures. The software calculates the z-score value for the input protein and compared it with the z-score values of all experimentally protein chains in the PDB. A z-score value within the range values of the experimental protein scores typically found for native proteins of similar size indicates a good model. The software also returns a plot of local model quality by plotting energies as a function of amino acid sequence position to fast identify problematic or erroneous regions of the input structure corresponding to positive values (z-score greater than 0). Thus, the model was selected on the basis of the overall G-factor and the z-score values.

6.2.3 Structure minimization

The selected model (Model 1) was processed as followed. Hydrogen atoms were added, and energy minimized on Sybyl software v8.1 (www.tripos.com) using the Powell algorithm with a coverage gradient of $\leq$0.5 kcal (mol Å)$^{-1}$ and a maximum of 1500 cycles. After that, the model was prepared using the web-based graphical user interface CHARMM-GUI (http://www.charmm-gui.org/). The structure was solvated in a rectangular 15 Å water box (TIP3S) and NaCl atoms were added to neutralize the system. The overall structure was energetical minimized using NAMD 2.0 software package (Phillips et al., 2005) for 1ns using a conjugate gradient. The Root Main Square Deviation (RMSD) and the Root Main Square Fluctuation were calculated using the VMD 1.9.3 software (Humphrey et al., 1996). The model was energetically minimized for 1 nanosecond (ns) to remove potential structural distortion or clashes in the structure. The model reached an energetical local minimum quite quickly, after only 0.1 nanoseconds (ns). The average RMSD value was 0.715 and the RMSF value was lower than 0.12Å for all the residues of the model. The structure minimized has been used to evaluate side chains flexibility within the ligand binding pocket to run a flexible docking. Three residues have shown a certain degree of flexibility corresponding to the amino acids: Asn245, Arg255 and Cys371. The other residues in the binding pocket have not shown relevant side chains flexibility.

6.2.4 Protein and ligands preparation

The last frame of the molecular minimization was used as input structure for the molecular docking with GOLD (see below). However, for the docking with AutoDock

(AD; see below), the RXRα model was further processed as followed: the AutoDock Tools software was used to add polar hydrogen to the proteins, Gasteiger charges were calculated for each atom, and AD4 type were assigned to the atoms. Ligands were retrieved from PubChem database in *.sdf* format and the physiological pH was set using the software FLAP (Fingerprint for Ligand and Protein).

6.2.5 GOLD docking

The GOLD software v5.2.2 (CCDC; Cambridge, UK; http://www.ccd.cam.ac.uk)(Jones et al., 1997) was applied to dock the selected molecules into the binding site of *D. magna* RXRα. For each compound 50 binding poses were generated without any constraints. The centroid of the binding site was defined according to the binding site of the crystallographic structure from which the model has been built (1FBY). Side chains flexibility was allowed for the amino acids: Asn245, Arg255 and Cys371. For the genetic algorithm run, a maximum number of 100000 operations were performed on a population of 100 individuals with a selection pressure of 1.1. Operator weights for crossover, mutation, and migration were set to 95, 95, and 10, respectively. The number of islands was set to 5, and the niche size was set to 2. The default GOLD Score fitness function was applied for performing the energetic evaluations. The distance for hydrogen bonding was set to 2.5 Å, and the cut-off value for the van der Waals calculation to 4.0 Å. For ligand flexibility options, flip pyramidal N, flip amide bonds, and flip ring corners were allowed. After that, all the poses generated by GOLD were rescored using the scoring function HintScore (HINT, Hydropathic INTeraction). The coupling of these two scoring function was chosen as: i) GoldScore allows to take into account factors such as H-bonding energy, van der Waals energy and ligand torsion strain; ii) HintScore provides a quantitative evaluation of protein-ligand interaction that allows to take into account both the enthalpic and entropic contributions to the ΔG of ligand-protein interaction, based on experimental protein and ligand Log Po/w values (Cozzini et al., 2002; Eugene Kellogg and Abraham, 2000).

6.2.6 AutoDock docking

Autodock Vina(Trott and Olson, 2010) was used for docking molecules into the binding site of the receptor. The search space was included in a box of 24 × 24 × 24 Å, centred on the binding site of the ligands (as mentioned before in the Gold Docking paragraph).

The side chain flexibility was allowed for the same residues defined in the Gold Docking. The ligand amide and backbone flexibility were allowed.

6.2.7 Molecular dynamic simulations

The structural stability of the system bounded to the two ligands was performed by means of extended molecular dynamics (MD) simulations. The simulated systems were solvated in a pre-equilibrated octahedral box of TIP3P water molecules. The protonation state was set to the physiological pH for the ionizable residues. The final system contains 13246 water molecules and 5 chloride anions for the apo form, PPF11, PPF17 and 4OH, respectively. Simulations were performed in NPT ensemble for the equilibration phase and in the NVT mode for the production tuns. Periodic boundary conditions and Ewald sums (grid spacing 1 Å) were used for treating long-range electrostatics interactions. All simulations were performed using Amber version 18 with the Amber GAFF force field. Ligands were parametrized with the GAFF force field in conjunction with RESP charges.

6.3 Results and Discussion

6.3.1 Homology modeling

Since the structure of the Retinoic X Receptor LBD of D. magna is not yet solved, a 3D structure has been modelled. Three homology modeling servers (Swiss Model, Phyre2 and I-TASSER) were used to have a consensus model prediction that strengthen the model generated. Starting from the LBD amino acids sequence of D. magna, the best model generated by each server has been used. The three models have been evaluated from geometrical and energetical features using PROCHECK and ProSA and they showed a great number of φ-ψ torsional angles values in the allowed regions of the Ramachandran plot, an overall G-factor under the threshold value of -0.5, and energetically good values. Considering together the results, the 3D structure built by SWISS-MODEL has been used to carry out the research.

6.3.2 Molecular docking results

Eight pesticide compounds were docked within the ligand binding pocket of RXRα (Figure 1): epofenonane (57342-02-6), fenoxycarb (72490-01-8), hydroprene (41096-46-2), kinoprene (42588-37-4), methoprene acid (53092-52-7), methyl farnesoate (10485-70-8), pyriproxyfen (95737-68-1), 4'-OH-pyriproxyfen. All the three software have ranked the 4'-OH-pyriproxyfen as the most interactor compound. Considering the consensus prediction, the binding affinity for the RXRα was: 4'-OH-pyriproxyfen ≃

pyriproxyfen > methoprene acid > fenoxycarb > epofenonane > kinoprene > hydroprene > methyl farnesoate. It should be mentioned that all the values are quite high for each scoring function. Thus, all these compounds could be possible RXRα endocrine disruptors of *D. magna*. Analysing the binding mode of the best poses of Autodock and Gold, the compounds that have carboxylic and/or ester groups at the end of their structure were all docked in the same position in the binding site. These groups are oriented toward Thr211, Ala266 and Arg255 with which they establish hydrogen bonds and acid-base interactions. In the case of fenoxycarb, the compound in the best HINT and Gold poses were not docked in the same position. In fact, in the best Gold pose, the ligand established a hydrogen bond with the protein residue Thr211, while, in the best HINT pose, this interaction was not present and the ligand was predicted in an opposite position compared to the best Gold pose (***Figure 2***).

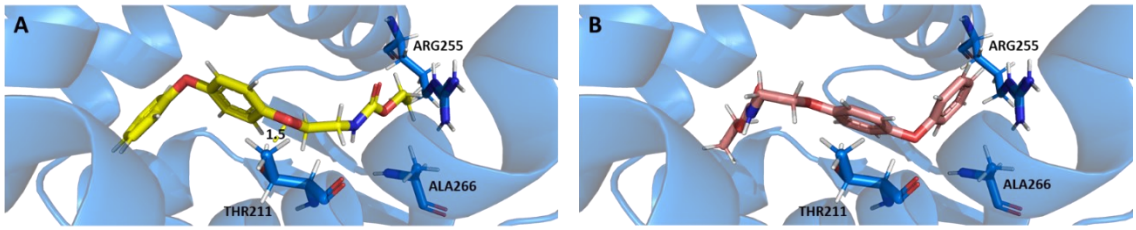


Figure 2: (A) Best Gold score binding pose of fenoxycarb: the ligand establishes a hydrogen bond with the residue THR211; (B) Best HINT score binding pose of fenoxycarb: the ligand was docked in the opposite position compared to the best Gold pose.

Since the pyriproxyfen has been predicted as one of the best interactors by all the software, a major attention has been paid to this compound. Although the three scoring functions have predicted different conformations, the ligand established mainly hydrophobic interaction with the receptor that did not constrain the pyriproxyfen in any specific conformation giving a certain degree of flexibility. However, it is well known that pyriproxyfen has a low to moderate persistence under aerobic conditions in soil. It rapidly undergoes to hydrolysis with the formation of its metabolite 4'-OH-pyriproxyfen ("Conclusion on pesticide peer review regarding the risk assessment of the active substance pyriproxyfen," 2009). Thus, molecular docking has been also applied to this compound to screen their capability to bind the RXRα. The main metabolite of pyriproxyfen has quite higher docking values compared to pyriproxyfen itself, ranking it in the first position. In contrast to the pyriproxyfen, the metabolite was docked in the same position for all the best binding poses of the three scoring functions. In fact, the presence

of the hydroxylic group at the end of the ligand locates the OH toward the Arg255 of the receptor. Moreover, while the phenol group of the 4'-OH-pyriproxyfen was constrained in the same position, the remain portion of the ligand was more flexible in the binding pocket showing a quite different position in the three best binding poses (***Figure 3***). Nevertheless, the three best poses showed that molecular docking predictions predicted a very close binding mode. This was clearly evident in the Figure 3D where the three best poses are overlapped. However, the best HINT and Gold poses were more similar compared to the Autodock best poses.

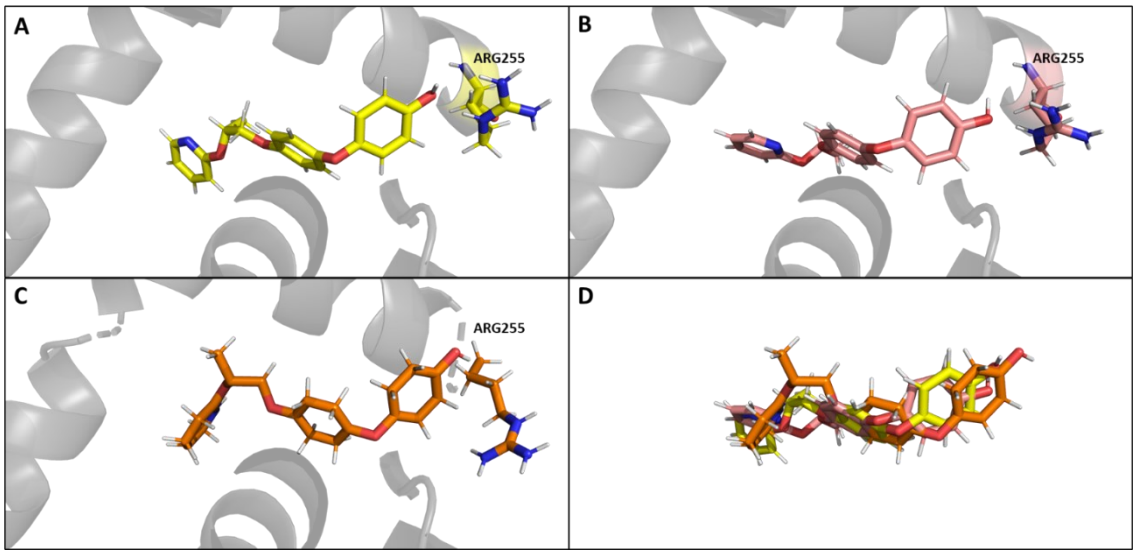


Figure 3: (A) best Gold Score pose (yellow); (B) best HINT score pose (pink); (C) best Autodock score pose (orange); (D) all three best binding poses overlapped: it is clearly evident that the three docking software converged toward a similar binding pose prediction.

The stability of pyriproxyfen and of its metabolite have been also analysed using molecular dynamics simulation to evaluate their stability during the simulation time and to find out relevant conformational changes (if any) in the protein structure due to the pyriproxyfen/4'-OH-pyriporxyfen binding.

6.3.3 Molecular Dynamic Simulations

As noted above, RXRα was simulated bound to the two ligands: (i) pyriproxyfen and (ii) 4'-OH-pyriproxyfen. Molecular dynamic (MD) simulation has been performed in triplicate. The analysis of the trajectories sampled by the MD of the two systems should evidence the differences induced by the binding of pyriproxyfen and by its metabolite on the dynamical properties of the protein. The results point out that the two systems reach the equilibrium very quickly (after only 50 nanoseconds). However, the binding of the

metabolite leads to a quite greater protein stability, as indicated by the time evolution of the root-mean square deviation (RMSD) profile (**Figure 4**). However, the main difference was more evident considering the ligands stability. As we can see in **Figure 4**, pyriproxyfen shows greater fluctuation in the RMSD values (~ 3 – 4 Å) than its metabolite (~ 2 - 2.5 Å) during all the simulation time and in all the three replicas.

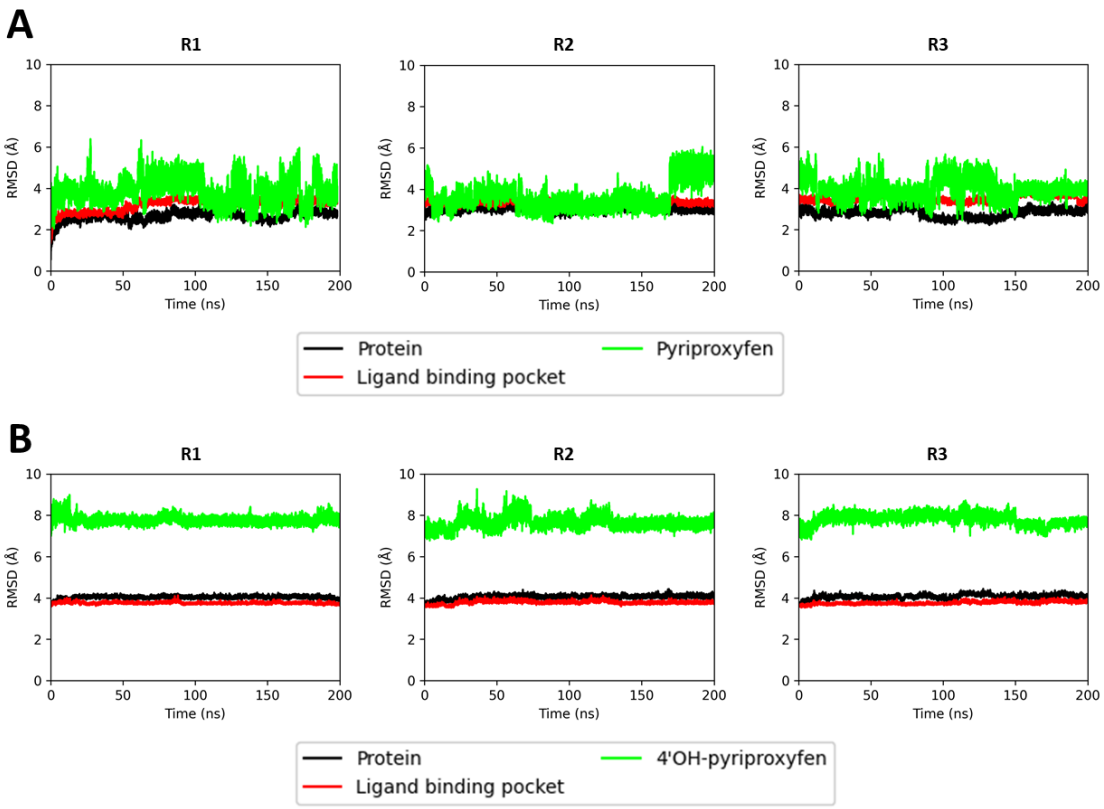


Figure 4: A) RMSD profile of the protein bound to pyriproxyfen in the three replicas (R1, R2, R3. B) The three replicas of MD simulation of the protein bound with the main metabolite of pyriproxyfen (4'-OH-pyriproxyfen).

This trend can be primarily attributed to the type of interactions that the two ligands established with the receptor. Pyriproxyfen mainly interacts with the protein using hydrophobic interactions, whereas the presence of an hydroxylic group in the 4'-OH-pyriproxifen allows to this compound the formation of a hydrogen bond with the residue Arg255. Moreover, during the simulation, 4'OH-pyriproxyfen forms a second hydrogen bond with a water molecule throughout its hydroxylic group. These two interactions can stabilize the ligand in the binding pocket maybe explaining why the metabolite has little movements. On the other side, little differences can be evidenced in the RMSD profile of the ligand binding pocket considering the two systems.

Moreover, since it is well-known that Helix 12 plays a significant role in nuclear receptors family, the RMSD trend of this protein region has been considered and it is shown in **Figure 5**. The binding mode of the metabolite confers a more stability to Helix 12 compared to the pyriproxyfen binding. This greater stability can be mainly induced by the lower movement exploited by the compound inside the binding pocket. In fact, whereas the metabolite is stable during the simulation time and does not diffuse in the binding pocket, the pyriproxyfen moves into the binding site. This ligand instability can induce more movements in the protein to allocate constantly the ligand also conferring a greater movement to Helix 12.

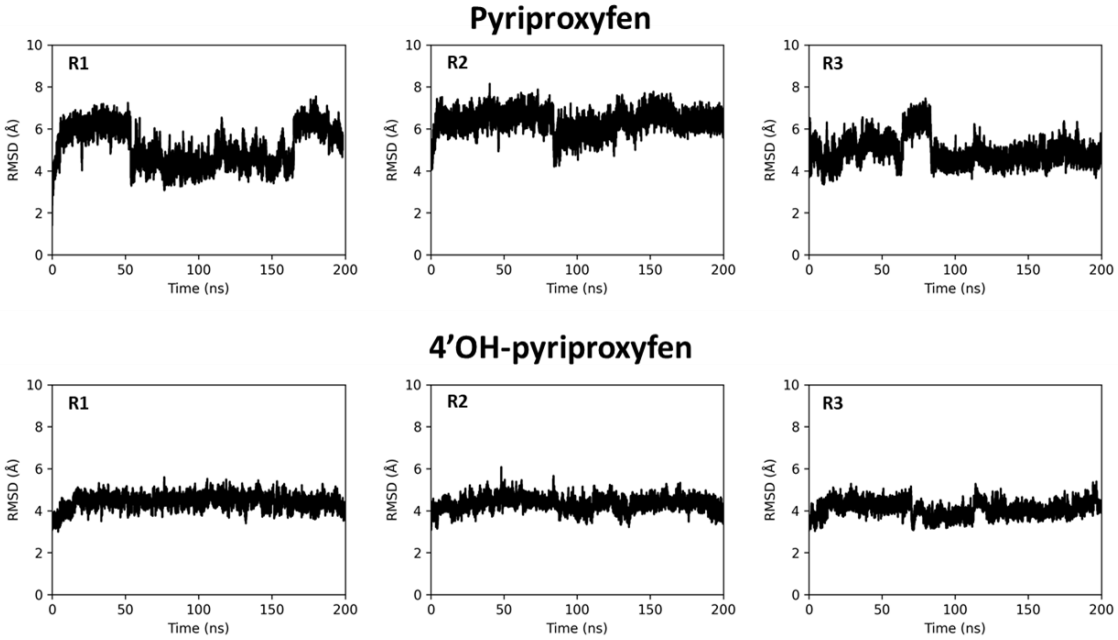


Figure 5: RMSD profile of MD simulations during the three replicas considering the helix 12 of Retinoic X Receptor alpha.

In **Figure 6** is shown the per-residue fluctuation averaged from the three replicas run for every simulated system. Generally, the trend of the RMSF is quite similar between the two complexes, but some differences can be evidenced.

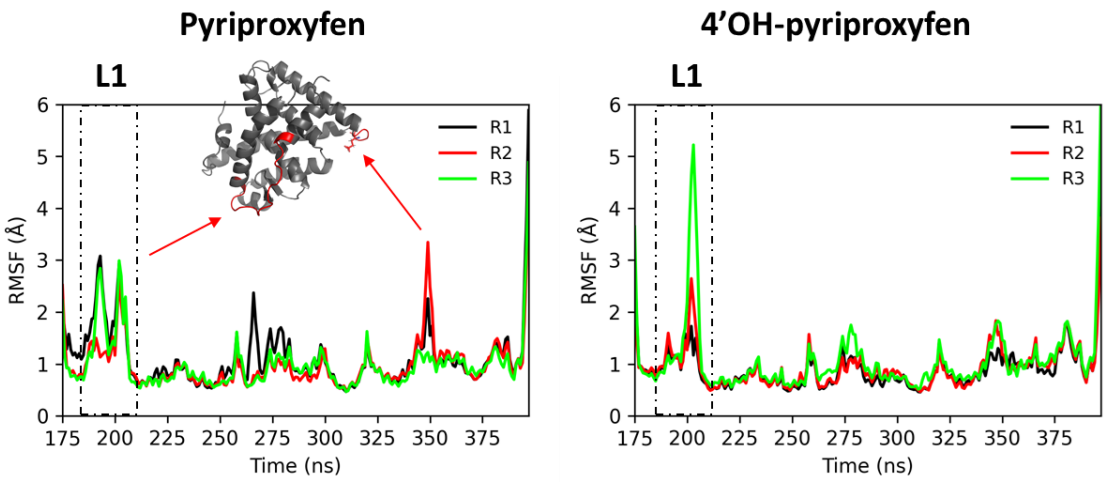


Figure 6: The RMSF profile of the full protein in complex with pyriproxyfen (left) and its metabolite (right). The results of the three replicas are shown in black, red, and green.

The regions with the highest differences correspond to the loop region that connects Helix 1 to Helix 2 (residues: 188-210) and below the residue 350 located in the loop region connecting Helix 9 to Helix 10. Although there is a peak in the RMSF below the residues 200-202 in the protein complexed with 4'-OH-pyriproxyfen, the magnitude of fluctuation interests a wider part of the protein in the complex with pyriproxyfen. Another difference can be highlighted near the residue Glu350 in the complex with the pyriproxyfen. Generally, the largest fluctuations are found in the complex with pyriproxyfen, and lower RMSF values are found in the complex with the metabolite, reflecting the increased stiffness arising by the formation of a hydrogen bond between the ligand and the binding pocket.

6.3.4 Essential Dynamics (ED)

As evidenced by the RMSD trends of the MD, the major difference highlighted between the two complexes is in the Helix 12 region. Moreover, its movement has a relevant role for the function of the protein. In order to gain a deeper insight about the major differences in the dynamic of this helix promoted upon binding of pyriproxyfen and its metabolite, essential dynamic (ED) was used. The first essential motion (EM) accounts for almost 33%, 29% and 20% (replica 1, 2, and 3, respectively) of the whole structural variance of the H12 Cα atoms for the complex with pyriproxyfen. The other two essential motions account only for a minor part of the structural variance since they are almost 3- and 8-fold lesser than the contribution of the first EM.

Comparison of the results obtained for the metabolite also reveals a similar trend. In fact, the first EM has a major contribution to the overall flexibility of the H12, as it accounts for almost 27%, 12% and 31% of the structural variance (replica 1, 2, and 3, respectively) and the EM2 and EM3 are lesser relevant for its structural variance. By the projection of the PC1 on PC2 and PC3, it is evident that the H12 exploits two different conformations during the simulation upon the binding of pyriproxyfen. Although it reveals that the extent of H12 conformational motion change observed in the complex with the metabolite is lesser valuable.

6.4 Conclusion

The use of pesticides is undoubtedly important to protect crops from pests, weeds, and plant diseases avoiding food loss in the field or in storage. In fact, it has been estimated that some 20-40% of the global crop production is globally lost because of the effects of pests. This has a greater impact on the worldwide economy, causing a cost of around $290 billion due to plant diseases and invasive insects (FAO). In addition, natural sources, such as energy, water, and soil, also contribute to food waste. However, although pesticides help to protect crops, sometimes there is an uncontrolled use. This rises very concerns about human and animal health which are exposed to pesticides through food and water. It is well-known that some of them can interact with nuclear receptors acting as endocrine-disrupting compounds.

Hazard identification analyses are normally performed to ensure the safety of all the compounds used in the entire food chain. These analyses are often carried out using organism models, like *Daphnia Magna*. Due to the large number of molecules that are currently used, it is often challenging to test all these compounds using *in vitro* and/or *in vivo* analyses.

Thus, the purpose of this paper was to use *in silico* techniques for the prediction of the endocrine interference of some pesticides on the *Daphnia magna* Retinoid X Receptor (RXR), a nuclear receptor that has been proposed as a hypothetical target of pesticides. Docking results predicted the pesticide pyriproxyfen (PP) as a strong interactor of the RXR. Since PP has low persistence and is rapidly degraded in 4-OH-pyriproxyfen ((EFSA), 2009), we also tested the protein-ligand interaction of this metabolite. The phenol group of 4-OH-pyrirpoxifen is able to interact with RXR establishing a hydrogen bond with the Arg255 of the receptor, which anchors the ligand in the ligand-binding pocket. To increase our knowledge about the interaction of pyriproxyfen and its

metabolite on RXR, we applied molecular dynamic simulations. The results showed that both compounds are stable during the simulation time. However, the hydroxylic group of the 4-OH-pyriproxyfen induced lower fluctuations in the protein-ligand complex compared to its parent pyriproxyfen and it is more stable than PP.

In conclusion, these *in silico* analyses identified pyriproxyfen and 4-OH-pyrirpoxifen as good interactors of the RXR, although almost all pesticides revealed great interaction values. The binding of these two compounds to the retinoid X receptor might affect the normal function of the protein, inducing an endocrine effect in *Daphnia magna*. Moreover, although *in silico* analyses have been performed considering *D. magna*, the high amino acids sequence identity of the retinoic X receptor alpha between *D. magna* and *Homo Sapiens* could suggest that all these compounds could also be good binders for the human receptor.

References

Abe, R., Toyota, K., Miyakawa, H., Watanabe, H., Oka, T., Miyagawa, S., Nishide, H., Uchiyama, I., Tollefsen, K.E., Iguchi, T., Tatarazako, N., 2015. Diofenolan induces male offspring production through binding to the juvenile hormone receptor in Daphnia magna. Aquat. Toxicol. 159, 44–51. https://doi.org/10.1016/j.aquatox.2014.11.015

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402. https://doi.org/10.1093/nar/25.17.3389

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

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

Bilal, M., Iqbal, H.M.N., Barceló, D., 2019. Persistence of pesticides-based contaminants in the environment and their effective degradation using laccase-assisted biocatalytic systems. Sci. Total Environ. 695, 133896. https://doi.org/https://doi.org/10.1016/j.scitotenv.2019.133896

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421. https://doi.org/10.1186/1471-2105-10-421

Camp, A.A., Haeba, M.H., LeBlanc, G.A., 2019. Complementary roles of photoperiod and temperature in environmental sex determination in Daphnia spp. J. Exp. Biol. 222. https://doi.org/10.1242/jeb.195289

Conclusion on pesticide peer review regarding the risk assessment of the active substance pyriproxyfen, 2009. . EFSA Sci. Rep. 336, 1–99.

Cozzini, P., Fornabaio, M., Marabotti, A., Abraham, D.J., Kellogg, G.E., Mozzarelli, A., 2002. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 1. Models without explicit constrained water. J. Med. Chem. 45, 2469–2483. https://doi.org/10.1021/jm0200299

Eugene Kellogg, G., Abraham, D.J., 2000. Hydrophobicity: Is LogP(o/w) more than the sum of its parts? Eur. J. Med. Chem. 35, 651–661. https://doi.org/10.1016/S0223-5234(00)00167-7

Ginjupalli, G.K., Baldwin, W.S., 2013. The time- and age-dependent effects of the juvenile hormone analog pesticide, pyriproxyfen on Daphnia magna reproduction. Chemosphere 92, 1260–1266. https://doi.org/https://doi.org/10.1016/j.chemosphere.2013.04.061

Humphrey, W., Dalke, A., Schulten, K., 1996. VDM: visual molecular dynamics. J. Mol. Graph. 14, 33–8, plates, 27–28. https://doi.org/10.1016/0263-7855(96)00018-5

Jones, G., Willett, P., Glen, R.C., Leach, A.R., Taylor, R., 1997. Development and validation of a genetic algorithm for flexible docking. J. Mol. Biol. 267, 727–748. https://doi.org/https://doi.org/10.1006/jmbi.1996.0897

Jordão, R., Garreta, E., Campos, B., Lemos, M.F.L., Soares, A.M.V.M., Tauler, R., Barata, C., 2016. Compounds altering fat storage in Daphnia magna. Sci. Total Environ. 545–546, 127–136. https://doi.org/https://doi.org/10.1016/j.scitotenv.2015.12.097

Kakaley, E.K.M., Wang, H.Y., LeBlanc, G.A., 2017. Agonist-mediated assembly of the crustacean methyl farnesoate receptor. Sci. Rep. 7, 45071. https://doi.org/10.1038/srep45071

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., Sternberg, M.J.E., 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10, 845–858. https://doi.org/10.1038/nprot.2015.053

Kim, H.J., Koedrith, P., Seo, Y.R., 2015. Ecotoxicogenomic approaches for understanding molecular mechanisms of environmental chemical toxicity using aquatic invertebrate, Daphnia model organism. Int. J. Mol. Sci. 16, 12261–12287. https://doi.org/10.3390/ijms160612261

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 1993;26(2)283-291 26, 283–291. https://doi.org/doi:10.1107/S0021889892009944

Laskowski, R.A., MacArthur, M.W., Thornton, J.M., 2012. PROCHECK: validation of protein-structure coordinates. Int. Tables Crystallogr., Major Reference Works. https://doi.org/https://doi.org/10.1107/97809553602060000882

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., Thornton, J.M., 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR 8, 477–486. https://doi.org/10.1007/BF00228148

Lima, D., Castro, L.F.C., Coelho, I., Lacerda, R., Gesto, M., Soares, J., André, A., Capela, R., Torres, T., Carvalho, A.P., Santos, M.M., 2015. Effects of Tributyltin and Other Retinoid Receptor Agonists in Reproductive-Related Endpoints in the Zebrafish (Danio rerio). J. Toxicol. Environ. Heal. Part A 78, 747–760. https://doi.org/10.1080/15287394.2015.1028301

Lobley, A., Sadowski, M.I., Jones, D.T., 2009. pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. Bioinformatics 25, 1761–1767. https://doi.org/10.1093/bioinformatics/btp302

Mnif, W., Hassine, A.I.H., Bouaziz, A., Bartegi, A., Thomas, O., Roig, B., 2011. Effect of endocrine disruptor pesticides: a review. Int. J. Environ. Res. Public Health 8, 2265–2303. https://doi.org/10.3390/ijerph8062265

Olmstead, A.W., Leblanc, G.A., 2002. Juvenoid hormone methyl farnesoate is a sex determinant in the crustacean Daphnia magna. J. Exp. Zool. 293, 736–739. https://doi.org/https://doi.org/10.1002/jez.10162

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L., Schulten, K., 2005. Scalable molecular dynamics with NAMD. J. Comput. Chem. 26, 1781–1802. https://doi.org/10.1002/jcc.20289

Remmert, M., Biegert, A., Hauser, A., Söding, J., 2011. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. Nat. Methods 9, 173–175. https://doi.org/10.1038/nmeth.1818

Sippl, M.J., 1993. Recognition of errors in three-dimensional structures of proteins. Proteins Struct. Funct. Bioinforma. 17, 355–362. https://doi.org/https://doi.org/10.1002/prot.340170404

Spaggiari, G., Iovine, N., Cozzini, P., 2021. In silico prediction of the mechanism of action of pyriproxyfen and 4′-oh-pyriproxyfen against a. Mellifera and h. sapiens receptors. Int. J. Mol. Sci. 22. https://doi.org/10.3390/ijms22147751

Test No. 202: Daphnia sp. Acute Immobilisation Test, 2004. , OECD Guidelines for the Testing of Chemicals, Section 2. OECD. https://doi.org/10.1787/9789264069947-en

Test No. 211: Daphnia magna Reproduction Test, 2012. , OECD Guidelines for the Testing of Chemicals, Section 2. OECD. https://doi.org/10.1787/9789264185203-en

Tkaczyk, A., Bownik, A., Dudka, J., Kowal, K., Ślaska, B., 2021. Daphnia magna model in the toxicity assessment of pharmaceuticals: A review. Sci. Total Environ. 763, 143038. https://doi.org/https://doi.org/10.1016/j.scitotenv.2020.143038

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

UniProt Consortium, T., 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 46, 2699. https://doi.org/10.1093/nar/gky092

Wang, Y.H., LeBlanc, G.A., 2009. Interactions of methyl farnesoate and related compounds with a crustacean retinoid X receptor. Mol. Cell. Endocrinol. 309, 109–116. https://doi.org/https://doi.org/10.1016/j.mce.2009.05.016

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296–W303. https://doi.org/10.1093/nar/gky427

Wiederstein, M., Sippl, M.J., 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 35, W407–W410. https://doi.org/10.1093/nar/gkm290

Yang, J., Zhang, Y., 2015. I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res. 43, W174–W181. https://doi.org/10.1093/nar/gkv342

Zhang, Y.-N., Zhu, X.-Y., Wang, W.-P., Wang, Y., Wang, L., Xu, X.-X., Zhang, K., Deng, D.-G., 2016. Reproductive switching analysis of Daphnia similoides between sexual female and parthenogenetic female by transcriptome comparison. Sci. Rep. 6, 34241. https://doi.org/10.1038/srep34241

Zhang, Y., Skolnick, J., 2004. Scoring function for automated assessment of protein structure template quality. Proteins Struct. Funct. Bioinforma. 57, 702–710. https://doi.org/https://doi.org/10.1002/prot.20264

CHAPTER 7

Computational Methods on Food Contact Chemicals: Big Data and In Silico Screening on Nuclear Receptors Family

Published in Chemosphere

https://doi.org/10.1016/j.chemosphere.2021.133422

Computational Methods on Food Contact Chemicals: Big Data and *In Silico* Screening on Nuclear Receptors Family

Pietro Cozzini[a]+*, Francesca Cavaliere[a]+, Giulia Spaggiari[a]+, Gianluca Morelli[b], Marco Riani[b]

[a] Molecular Modelling Lab, Department of Food and Drug
University of Parma
Parco Area Delle Scienze 17/A, I-43124, Parma, Italy
E-mail: francesca.cavaliere@unipr.it pietro.cozzini@unipr.it
giulia.spaggiari@unipr.it

* Corresponding author phone number: Italy, (+39) 0521 905669

[b] Department of Economics and Management and Interdepartmental Center of Robust Statistics
University of Parma
Via J. F. Kennedy 6, 43100, Parma, Italy
E-mail: gianluca.morelli@unipr.it marco.riani@unipr.it

Highlights

• Molecular docking and robust consensus scoring are useful to identify possible food and water dangerous molecules.

• Endocrine disruptor prediction using in silico methods to save time and cost.

• Database and big data approaches to accelerate hazard identification.

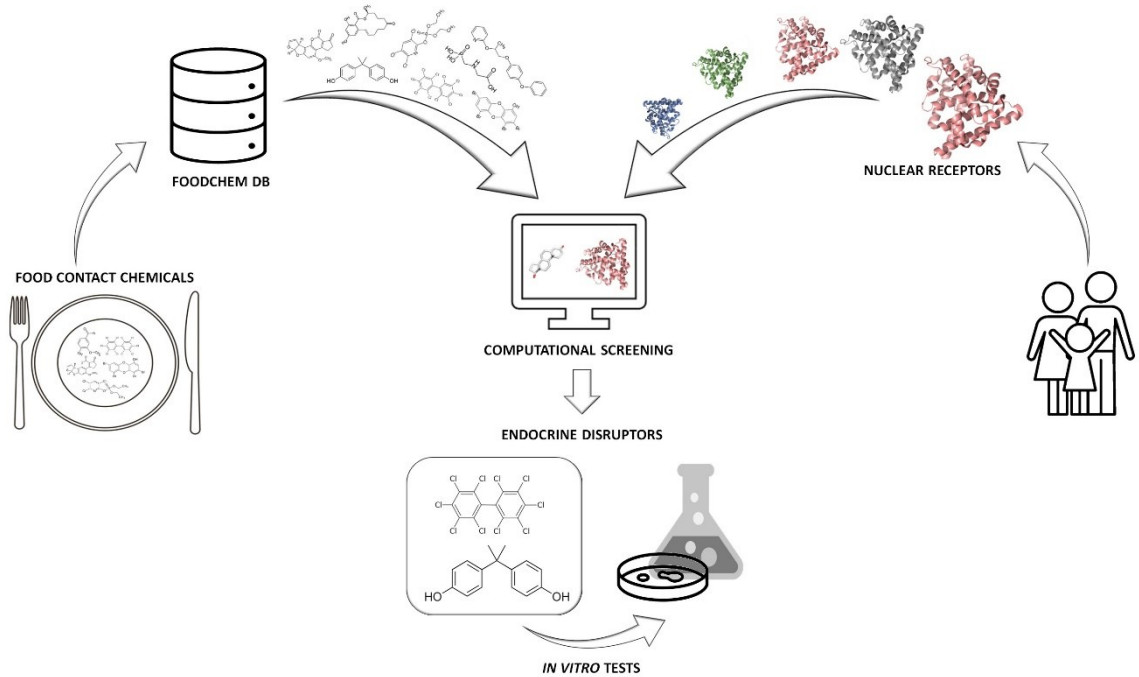
Abstract

According to Eurostat, the EU production of chemicals hazardous to health reached 211 million tonnes in 2019. Thus, the possibility that some of these chemical compounds interact negatively with the human endocrine system has received, especially in the last decade, considerable attention from the scientific community. It is obvious that given the large number of chemical compounds it is impossible to use in vitro/in vivo tests for identifying all the possible toxic interactions of these chemicals and their metabolites. In addition, the poor availability of highly curated databases from which to retrieve and download the chemical, structure, and regulative information about all food contact chemicals has delayed the application of in silico methods. To overcome these problems, in this study we use robust computational approaches, based on a combination of highly curated databases and molecular docking, in order to screen all food contact chemicals against the nuclear receptor family in a cost and time-effective manner.

Keywords

Computational chemistry; Consensus prediction; Database; Nuclear receptors; Toxicology

Graphical Abstract



7.1 Introduction

A research project starts with a question. The main question of this project is: how we can evaluate all the possible food contact chemicals against a protein family to discover potential endocrine disrupting activity. It is obvious that, given the large number of chemical compounds and their metabolites existing and developed every year, it is impossible to use *in vitro* (or *in vivo*) tests for identifying all possible toxic interactions. The solution is to use computational approaches to reduce the number of wet tests, seeking only the most probable interactors.

Endocrine disrupting chemicals (EDCs) are exogenous substances that can interfere with the synthesis, secretion, transport, binding, and elimination of natural hormones in the body that are responsible for the maintenance of homeostasis, reproduction, and behavior (Kavlock et al., 1996). Human exposure to EDCs occurs through oral consumption of food and water, contact with skin, inhalation, or intravenous, route (Kabir et al., 2015). These molecules are highly heterogenous and include pesticides, plasticizers (i.e., phthalates, bisphenols), persistent organic pollutants (POPs) (i.e., dioxins, polychlorinated biphenyls), but also chemicals added to food to enhance some characteristics (i.e., flavourings, food additives), or naturally occurred, such as mycotoxins. EDCs can act through different mechanisms: mimicking the action of a naturally produced hormone, blocking hormone receptors in cells, interacting indirectly by influencing the biosynthesis or availability of normal hormones. Between them, the most privileged route is the interaction with nuclear receptors (NRs). Nuclear receptors are a superfamily of 48 ligand-activated transcription factors, including estrogen receptor (ER), androgen receptor (AR), mineralocorticoid receptor (MR), glucocorticoid receptor (GR), progesterone receptor (PR), and thyroid receptor (TR). NRs share a common structural organization composed of an *N*-terminal region (A/B domain), a conserved region DNA-binding domain (DBD), and a ligand-binding domain (LBD) responsible for ligand recognition. The alteration of nuclear receptors pathways is correlated to many pathologies, such as breast cancer, prostate cancer, and testicular cancer, infertility, cardiovascular complications, disturbances in energy metabolism, immune responses, impairment of cognitive functions and the regulation of cell proliferation and differentiation, hypertension, obesity, and so on (Dall'Asta, 2016) (De Coster and Van Larebeke, 2012) (Desvergne et al., 2009) (Fucic et al., 2012) (Luccio-Camelo and Prins,

2011) (Odermatt and Gumy, 2008) (Petrakis et al., 2017) (Safe, 2004) (Schug et al., 2011) (Gore et al., 2015). In order to prevent human diseases, in the past decades, different regulatory and policy approaches were made even if the identification and safety assessment of potential EDCs is complicated both by the observed low-dose effects and the often long-term exposure or exposure during a critical window early in development. One of these is the REACH (Registration, Evaluation, Authorisation, and Restriction of Chemicals) legislation that is committed to protecting human health and the environment from hazardous chemicals. However, testing all the possible EDCs against all the potential targets is very important but also an expensive, long and difficult task (e.g., the nuclear receptors family contains 48 members). In fact, these tests are still mainly based on biological and animal experimentations (toxicity tests), very time- and cost-intensive, and which cause millions of animals' death every year. In this context, *in silico* methods, already well-established tools in drug discovery, can be good tools either in the identification of new EDCs or pointing in the right direction when finding the mechanism of action for already known EDCs. Computational approaches produce predictive models that are more rapid and less costly than *in vitro* and *in vivo* tests, allowing a large amount of data concerning numerous chemical substances to be generated and analysed in a short time without the use of test animals (F. Cavaliere et al., 2020). A key prerequisite for the successful application of computational modeling techniques is the quality of the input data. The availability of open access databases offers the capability to retrieve a huge amount of information from different data sources. The CAS Registry Number (RN) has been chosen, long time ago, as a unique and unambiguous numeric identifier for a specific chemical compound. It is developed by the American Chemical Society to help scientists to retrieve and use information from different data sources. Since it may be unique, validated, and internationally recognized, the governmental agencies rely on CAS RNs for substance identification. However, CAS RNs are often used improperly by the scientific community and there is no check made by the American Chemical Society. Thus, it is really common to find some errors and this wrong information propagates easily across the Internet (Grulke et al., 2019). In fact, conflicts in the chemical identifier are not so rare in public resources and these errors propagate quickly and easily across the internet. These undermine the effort of *in silico* methods. So far, much attention has been paid to structure normalization to ensure the detection and the correction of three-dimensional errors and a variety of public and commercial toolkits exist to address this problem. However, less attention is often given to the consistency of the association

between chemical identifiers (CAS RN and name) and chemical structures. For example, the compound classified as flavouring having the CAS N: 563187-91-7 and the common name “l-Menthone-1,2-glycerol ketal” in the EFSA list is a typical example of CAS:Name wrong association. In fact, this CAS actually corresponds to “DNA (mouse strain C57BL/6 J clone 5430425J12 EST (expressed sequence tag))” and the correct CAS RN of the compound “l-Menthone-1,2-glycerol ketal” is 67785–70-0. Moreover, although CAS RN is commonly used as an identifier of the majority of databases, in several databases molecules are classified using different identifiers and thus there is often a lack of standardisation (Hersey et al., 2015). Although data quality is undoubtedly important for every database, they may have been developed with different aims and scope, and it is unreasonable to expect the same degree of curation. The increasing amounts of compounds released every year (500–1000 new molecules) and that are in contact with food, along with the different sources of data, have made it difficult to check manually the reliability of data. In view of this, it is essential to design and implement a data curation pipeline into an automated procedure.

A wide number of computational applications (tools) specifically for the analysis of EDCs are available in the literature in order to determine the relationship between one compound and its toxic effect. In particular, the molecular docking technique is a well-establish application to study protein-ligand interaction, which means analysing if the ligand has the suitable physical-chemical characteristic, shape, the volume to fit properly into the binding cavity of the receptor. Molecular docking is mainly composed of two main parts: an algorithm that is used to predict different binding poses of a molecule in the protein binding site, and a scoring function used to evaluate the strength of ligand-protein interaction, i.e., to predict its binding affinity. Different algorithms and scoring functions exist but answering the question of which algorithm or scoring function is the best one, is a complicated task (Morris and Lim-Wilby, 2008). In fact, each docking software (that is the sum of algorithm and scoring function) has been trained with different proteins and ligands. Thus, before starting a molecular docking analysis, it should be advisable to identify the more appropriate software based on the trained protein-ligand complexes that best fit with the proteins and ligands under investigation. However, in the present work, 31 different nuclear receptors with different binding pocket characteristics and a huge number of heterogeneous molecules from a chemical and structural point of view were considered. Thus, it is unthinkable to identify a single docking program that may have the same performance for all nuclear receptors and for

all food contact molecules. For that reason, we used a robust consensus scoring approach using two different docking software and four different scoring functions. The combination of more scoring functions allows to reduce the number of false-positive and to obtain more reliable results by compensating the deficiencies of each scoring function, leading to an improvement of the performances (Teramoto and Fukunishi, 2007) (Wang et al., 2003). Such as Bissantz and co-workers have highlighted, the use of three different scoring functions enhances the capability to reach hit rates from 10% up to 70% (Bissantz et al., 2000).

The goal of this work is to predict a possible endocrine disrupting activity of a huge set of molecules that can contact the food as a base for further *in vitro/in vivo* tests using computational methods that do not consider the intake dose. The following approach takes into consideration the interaction between a ligand (i.e. the endocrine disruptor compound) and the binding site of a receptor (i.e. the nuclear receptor) that is considered the molecular initiate event (MIE). This event is fundamental from a biological point of view because it is the first mechanism that, in most cases, initiates a biological effect based on the occurrence of conformational changes, signaling cascade as well as interaction with other proteins. However, molecular docking does not predict the binding affinity of a ligand to a protein unless a correlation analysis was made using known experimental binding affinity (Kd, Ki, Ka, etc.) and the docking score value. In fact, molecular docking and scoring function refers to the binding interaction of a ligand for a protein that means analysing if the ligand has the suitable physical-chemical characteristic, shape, the volume to fit properly into the binding cavity of the receptor. Thus, it predicts how strong is the interaction. Although it can be wrongly thought that if a ligand interacts more tightly with a receptor, it should have a high binding affinity, this concept is not so obvious since other mechanisms are involved in determining the binding affinity. Lower binding force doesn't mean low in take dose, it means a lower DG° of binding.

7.2 Material and Methods

7.2.1 Database resources

Different databases and web sources have been used to identify the molecules that come into contact with food: European Food Safety Authority (EFSA) (www.efsa.europa.eu), United Stated Environmental Protection Agency (EPA) (www.epa.gov), Food Packaging

Forum (www.foodpackagingforum.org), and European Chemicals Agency (ECHA) (www.echa.europa.eu).

7.2.2 Data Quality

The entire procedure described below has been implemented as two different Python procedures, with a common part used to check CAS RN validity. In fact, most public databases use Chemical names and CAS RNs as substance identifiers. CAS RN is widely used across scientific literature, Internet resources, and the chemical regulatory domain. Data are often stored using CAS RN as the primary key of the database and chemical names and synonyms as secondary identifiers. A CAS RN can be considered valid if it fulfils two rules: 1) it is composed by 3-numeric parts separated by hyphens (##... - ## - #); 2) it satisfies the "checkdig" validation formula developed by CAS (www.cas.org/support/documentation/chemical-substances/checkdig). CAS numbers are preliminarily checked for the presence of leading zeros and zeros are removed. After that, the checkdig formula has been used on CAS numbers to verify their correctness.

First procedure

Using CAS RN as input query, the entire procedure can retrieve and check data congruence of the InChIKey extracted from three different servers: PubChem (www.pubchem.ncbi.nlm.nih.gov), ChemIDPlus (www.chem.nlm.nih.gov/chemidplus) from the National Institute of Health (NIH), and CompTox Chemistry Dashboard (www.comptox.epa.gov/dashboard) from EPA. Since the manual curation part of incongruent data and/or unfound CAS RN took a great amount of time, a second procedure has been developed.

Second procedure

Starting from the CAS RN information, it has been converted into fixed URLs to automatically extract the correct InChIKey information within the CAS database (www.commonchemistry.cas.org), which is the official repository of CAS RN. In this step, the presence of salt and mixture was also checked. At the end, the InChIKey information was used as input query for extracting other information from PubChem ("CAS", "CID", "Common_name", "IUPAC_Name", "MolecularFormula", "MolecularWeight", "CanonicalSMILES", "InChI", "InChIKey").

7.2.3 Database descriptors

The foodchem DB stores 27 different fields that can be divided in six different subgroups:

a) Chemical names: CAS, CID, EC number, common name, IUPAC name;
b) 1D chemical information: molecular formula, canonical SMILES, InChI, InChIKey;
c) Chemical information: molecular weight, volume, logP value, number of acceptor atoms, number of donor atoms, number of chiral atoms, number of hydrophobic atoms, atom count, bond count, ring count, rotational bond count, positive charge atoms, negative charge atoms, total charge;
d) Regulative information: EFSA and ECHA link;
e) Three-dimensional structure in *.mol2* format;
f) Classification: it classifies the molecule based on its use in the food industry: flavouring, pesticide, dioxin, etc.

The detail information about where and how these data have been obtained is explained below.

PubChem information

PubChem database has been used to retrieve some food contact chemical data, as explained in the previous procedures: "CID", "Common_name", "IUPAC_Name", "MolecularFormula", "MolecularWeight", "CanonicalSMILES", "InChI", "InChIKey".

ECHA number and ECHA link

A python script has been developed to convert CAS RNs into fixed URLs to automatically retrieve EC numbers and to provide the corresponding link to the ECHA website' Substance Infocard.

3D structures

The three-dimensional structures (in *.sdf* format) of molecules that passed the previous steps have been retrieved from PubChem using a third python script.

Calculated chemical information

To store additional chemical information, other data have been calculated using two software:

- Sybyl v.7.: Acceptor, Donore, Hydrophobe, AtomCOunt, BondCount, RingCount, RotBonds, Chiral, logP value, Volume ($Å^3$);
- FLAP: number of Charge – and Charge + and the Total Charge.

Moreover, the FLAP (Fingerprint for Ligand and Protein) software was also used to convert the *.sdf* file into a *.mol2* file.

7.2.4 SQL and NoSQL

The data have been organized into two different databases, MariaDB and Elasticsearch, written implementing SQL and Bigdata technology (NoSQL – Not only SQL) respectively. We decided to implement two versions of the same database to answer two requirements. An SQL DB storing structural data of the selected molecules, more suitable for docking and molecular dynamics analysis, and a Big Data version able to store a different kind of information, not only structural information but also *in vitro*/*in vivo* tests, regulatory reports, etc. The specification of the structure/mapping used in the present work is explained in more detail in Table 1.

Table 1. SQL and NoSQL database structure definition.

Field	Data type (SQL)	Mapping (NoSQL)	
CAS	CHAR(16)	keyword	
CID	CHAR(20)	keyword	
EC number	CHAR(50)	keyword	
Common name	TEXT	text	keyword
IUPAC name	LONGTEXT	text	keyword
Molecular Formula	CHAR(100)	text	
Canonical SMILES	TEXT	keyword	
InChI	LONGTEXT	keyword	
InChIKey	CHAR(254)	keyword	
MW	FLOAT	double	
Volume	FLOAT	double	
logP	FLOAT	double	
Acceptor	INT(3)	byte	
Donor	INT(3)	byte	
Chiral	INT(3)	byte	
Hydrophobe	INT(3)	byte	
Atom Count	INT(3)	short	
Bond Count	INT(3)	byte	
Ring Count	INT(3)	byte	
Positive Charge	INT(3)	byte	
Negative Charge	INT(3)	byte	
Total Charge	INT(3)	byte	
EFSA link	CHAR(254)	keyword	
ECHA link	CHAR(254)	keyword	
.mol2	LONGTEXT	keyword	
Classification	CHAR(100)	text	

7.2.5 Protein preparation

The crystallographic structures of 31 nuclear receptors of *Homo sapiens* were downloaded from the Protein Data Bank (PDB) (www.rcsb.org). Among them, only 26 structures with high reliability and quality are available. For this reason, the nuclear receptors (3) with fragmented portions, such as constitutive androstane receptor (CAR), nuclear receptor-related 1 protein (NURR1), and estrogen-related receptor alpha (ERRα), were built and minimized for 1 ns with NAMD 2.13 software package. In addition, the mutated amino acids present in glucocorticoid receptor (GR) (F602S) and steroidogenic factor 1 (SF-1) (C247S and C412S) crystallographic structures were replaced. The receptor structures were processed using Sybyl software v8.1 (www.tripos.com). Water molecules and ligands were removed, and hydrogen atoms were added. Energy was minimized using the Powell algorithm with a coverage gradient of ≤ 0.5 kcal (mol Å)$^{-1}$ and a maximum of 1500 cycles. For the molecular docking with AutoDock (*see below*), the receptors were further processed: using AutoDockTools software polar hydrogens are added to the proteins and the Gasteiger charges were calculated to assign AD4 type to each atom.

7.2.6 Ligand preparation

Structural coordinates of the endogenous and putative ligands were retrieved from the NCBI PubChem compound database. Software FLAP was used to assign the correct protonation state to each ligand (pH=7.4).

7.2.7 Molecular docking with GOLD software

The GOLD software v5.8.1 (CCDC; Cambridge, UK; www.ccd.cam.ac.uk) was applied in order to dock ligands into the binding site of the 31 nuclear receptors. For each compound and receptor, 30 binding poses were generated. The binding site centroid of each receptor was defined using the coordinates of the crystallographic complexes. The side chain flexibility was allowed for each receptor amino acid. For the genetic algorithm run, a maximum number of 100000 operations were performed on a population of 100 individuals with a selection pressure of 1.1. The number of islands and the niche size were set to 5 and 2, respectively. The default GoldScore fitness function was applied for performing the energetic evaluations. The distance for hydrogen bonding and the cut-off value for the van der Waals calculation were set to 2.5 Å and 4.0 Å, respectively. Flip pyramidal N, flip amide bonds, and flip ring corners were allowed for ligand flexibility

options. After that, all the poses generated by GOLD software were rescored using the scoring functions ChemScore and HintScore (HINT, Hydropathic INTeraction).

7.2.8 Molecular docking with Autodock Vina Software

Molecular docking experiments were performed with Autodock Vina 1.1.2 using default settings (Trott and Olson, 2009). The search space was included in a box of 24 × 24 × 24 Å, centred on the binding site of the ligands as mentioned before. The side chain flexibility was allowed for the same residues defined in the GOLD docking. The ligand amide and backbone flexibility were allowed.

7.3 Results and Discussion

The foodchem DB has been also designed to accelerate computational applications since it stores not only regulative information but also chemical-physical properties and three-dimensional structures. Very careful attention has been made to ensure the correctness of the 3D structure to the CAS RN. Thus, it has been conceived for a different purpose compared to the FPF database which does not contain all the chemical-physical information used in the foodchem DB and it does not store the three-dimensional structure. Moreover, our database has been written in SQL and NoSQL language with the purpose to make it available to the scientific community through a website interface where the user can make searches and extract information. Using our database, the three-dimensional structures of 8091 substances, belonging to different sub-classes (Table 2), has been extracted and all these molecules have been screened using a molecular docking approach in order to identify the compounds having the capability to bind the thirty-one nuclear receptors. This method allows to screen the substances which have the most probable physical-chemical characteristics to act as endocrine disruptors.

Table 2. The total number of food contact chemicals falling in each subclass. Food contact chemicals are divided into 11 subclasses: dioxins, acrylamide, flavourings, food additives, furans, mycotoxins, pesticides, phthalates, bisphenols, polychlorinated biphenyls (PCBs), and food contact chemicals contained in the database of Food Packaging Forum (FCCDB).

Classification	Total number (8091)
Dioxins	75
Acrylamide	1
Flavourings	2091
Food Additives	110
Furans	133
Mycotoxins	327
Pesticides	465
Phthalates	361
Bisphenols	51
PCBs	209
FCCDB	4268

Two different docking software and four different scoring functions have been used as in our previous papers (Francesca Cavaliere et al., 2020) (Spaggiari et al., 2021). Thus, for each receptor and for each food contact chemical, four values have been obtained. In humans, there are 48 nuclear receptors, but many of these remain "orphans" as their endogenous ligands are yet to be determined. For this reason, if the endogenous ligand is known, the relative binding affinity (RBA) of each molecule was calculated using it as a reference compound. On the other hand, all the endogenous and no-endogenous co-crystallized ligands were docked against the respective nuclear receptors to obtain a reference value. A cut-off value was selected for each four docking values: i) a cut-off of 50 for GoldScore; ii) a cut-off of 30 for ChemScore; iii) a cut-off of −7 for Autodock (affinity); and iv) a cut-off of 500 for HintScore.

To reach a consensus scoring prediction, a robust statistical method has been used and it is explained in more detail below.

As training dataset, the crystallographic structures available from PDB of all ligand-NR complexes were considered. All ligands bound to the corresponding receptor were extracted and docked into the ligand-binding pocket to obtain the corresponding four scoring values. As for the food contact chemical data, every single value was used to calculate the relative binding activity considering the natural ligand as a reference compound:

$$Relative\ Binding\ Affinity\ (RBA)_n = \frac{food\ contact\ chemical\ score}{reference\ compound\ score}$$

where *n* is the number of scoring functions.

However, since the distribution data is non-normal for the potential presence of some outliers, a robust multivariate method was used to detect atypical values. In fact, it is well-known that the presence of atypical values can affect the results of any statistical analysis

especially when the number of observations is large. Using a confidence level of simultaneous 1%, we removed only values that were very far from the general bulk of the data. After the outlier removal, the values were rescaled in the domain [0 1] setting a score equal to 1 when it was larger than the value of the natural ligand. The degree of dispersion of the four rescaled values ($X_1, \dots, X_4$) has been considered by normalizing them in order to obtain four new variables ($Z_1, \dots, Z_4$) with 0 mean and variance equal to 1. After that, a principal component analysis was used on the four new variables to identify a weight coefficient for each scoring function (w_1, w_2, w_3, w_4) in such a way that the explained variance of the original variable is as large as possible ($w_j \geq 0$ and $\sum_{j=1}^{4} w_j^2 = 1$). We obtained a weight value of 0.12 – 0.94 – 0.14 – 0.29, for GoldScore, HintScore, ChemScore, and Autodock (affinity), respectively.

As for the training dataset, the relative binding affinity of each molecule and scoring function has been rescaled in the [0 1] domain after the outlier removal. To consider the different degrees of dispersion of the new rescaled variables, we standardized them to obtain four new variables. Since the purpose of the analysis was to combine the four scores into a single consensus score prediction, the final scores for the *i*-th food contact chemical have been obtained as:

$$\frac{\sum_{j=1}^{4} x_{ij} w_j}{\sum_{j=1}^{4} w_j} \; i = 1,2, \dots, n$$

where *n* is the total number of food contact chemicals.

At the end, the results have been divided into three cases based on their score: i) the molecules with a score between 0.0 and 0.3 are considered weak ligands since they interact with the corresponding nuclear receptor with a binding affinity that is 70% (or more) lower than the natural ligand (Figure 1A); ii) the molecules with a score between 0.3 and 0.8 are considered medium interactor compared to the natural ligand (Figure 1B); iii) the molecules with a score between 0.8 and 1.0 are judged as high interactor since they are able to bind the corresponding nuclear receptor with a binding affinity that is more than 80% of the natural ligand (Figure 1C). This latter case also includes the molecules that can interact with the nuclear receptor with a binding affinity greater than the natural ligand. Thus, all food contact chemicals falling in this class may be considered as substances of very high concern and should be the first compounds to analyse with further experimental methods in order to re-evaluate their use in the food industry.

Fig. 1 shows the percentage of molecules that can interfere with the endocrine system receptors highlighting their abundance in each specific nuclear receptor. If we focus on the single nuclear receptor, we can underline that more than 50% of food contact chemicals are good interactors of liver X receptor β (LXRβ), pregnane X receptor (PXR), progesterone receptor (PR), farnesoid X receptor (FXR), retinoic acid-related orphan receptor γ (RORγ), and peroxisome proliferator-activated receptor α (PPARα). In fact, LXRβ is the nuclear receptor with the highest number of food contact chemicals that fall in the high interactor group, and thus, it is likely the receptor most affected by the presence of these compounds in our body.

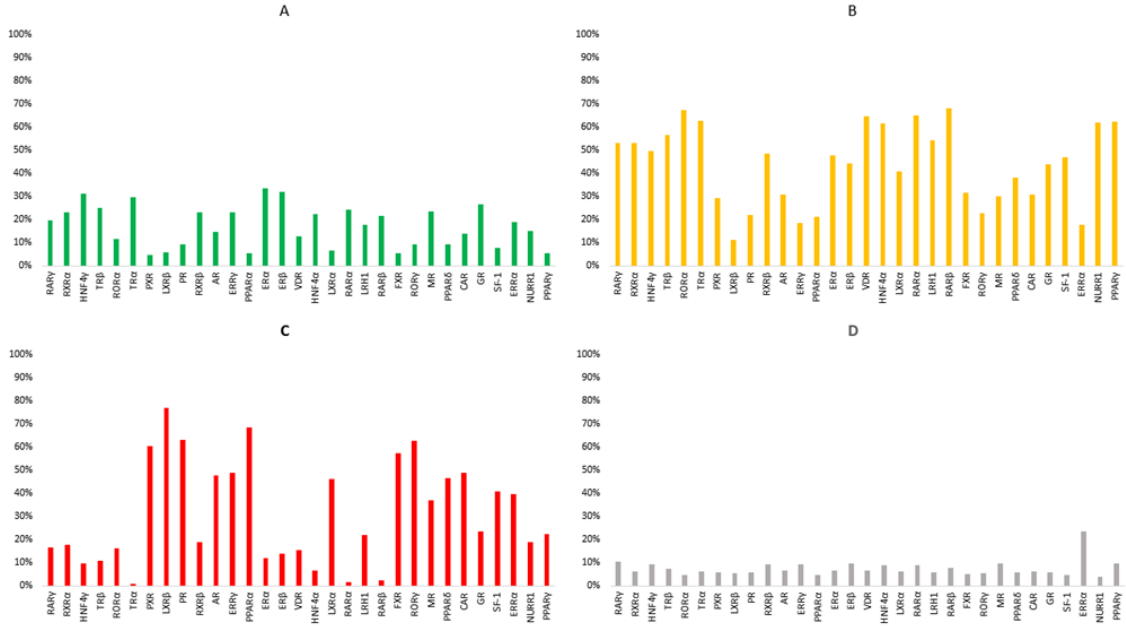


Fig. 1. Results obtained from the robust multivariate statistical procedure. The 31 NRs are on the x-axis, while the number of the molecules (%) is on the ordinate. The molecules with a score smaller than 0.3 are highlighted in green (A), the molecules with a score between 0.3 and 0.8 are highlighted in yellow (B), the molecules with a score greater than 0.8 are highlighted in red (C), while the outliers are highlighted in grey (D). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Considering Fig. 1D, we found almost the same number of outlier molecules in each nuclear receptor except for the estrogen-related receptor α (ERRα). This is not surprising since outlier molecules were generally substances having a high volume compared to the ligand-binding pocket of nuclear receptors. In fact, due to atom-atom clashes, the molecular docking scores were far away from the normal trend. Thus, considering that the volume of the ligand-binding pocket of ERRα is only about 80 $Å^3$ (against the ~300 $Å^3$ of the most nuclear receptor, excluding the PPAR family), it may be plausible to find a higher number of outliers.

As the second step of our analysis, we turned our attention on which class of food contact chemicals have the greater number of molecules able to interfere with the endocrine system. Thus, we counted the number of molecules belonging to each class that can interact with more than 50 percent of nuclear receptors with high, medium, and low binding affinity. As we can see in Fig. 2, almost the totality of dioxins, furans, and PCBs molecules can interact with more than 15 nuclear receptors with high binding affinity, following by the pesticides and phthalates sub-classes.

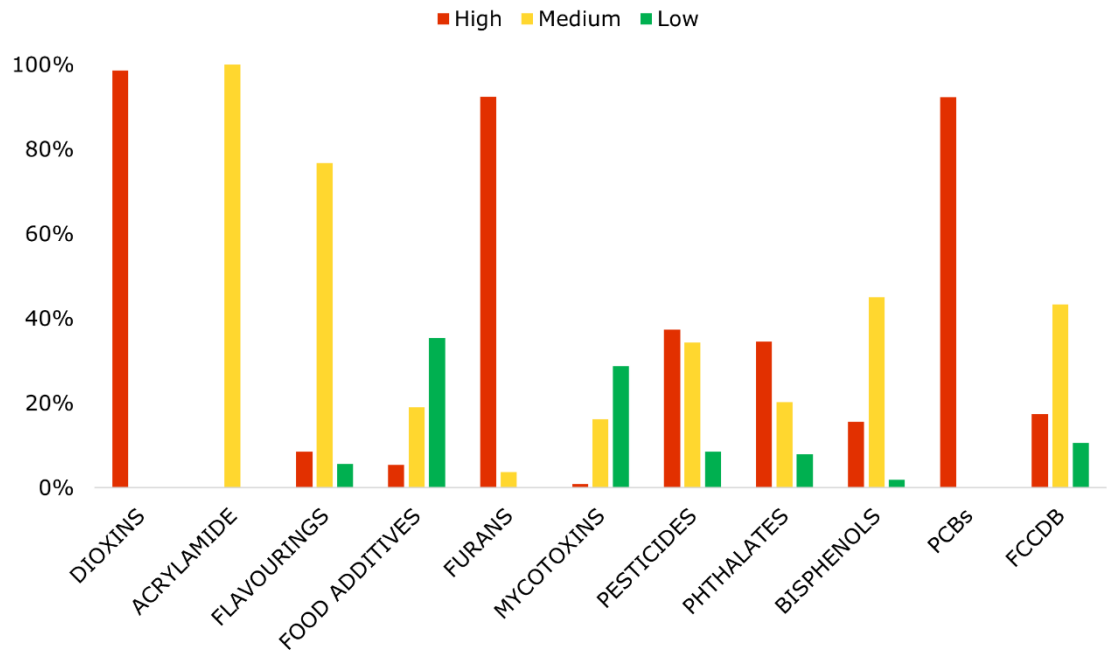


Figure 2. The percentage of molecules able to bind more than 15 nuclear receptors with high (≥0.8), medium (0.3-0.8), and low binding affinity (<0.3) considering each class of food contact chemicals.

The impact of this finding highlights the potential capability of these molecules to cause a very broad endocrine effect on the human body.

Considering the medium interactors, a great number of flavourings, bisphenols, and FCCDBs fall in this group. The single compound in the acrylamide class is also able to interact with more than fifteen nuclear receptors with medium binding affinity. On the other site, food additives and mycotoxins are more selective in their interaction with nuclear receptors, and just a few numbers of molecules can interact with high affinity to more than 50 percent of NRs.

7.4 Conclusion

One of the reasons that undermine *in silico* approaches is the availability of highly curated databases from which to retrieve and download the three-dimensional structure. This is most relevant in the food context due to the presence of salt and mixture components. In

fact, it is frequent on the web to find mixture or salt substances associated with the CAS RN of the main compound. In the present work, we created a database with a high level of data curation from which to retrieve chemical, structure, and regulative information about all food contact chemicals.

Using our foodchem database, we screened 8091 food contact chemicals against 31 nuclear receptors with the aim to identify the molecules that require major attention about their safety for the human body. In the food context, wet experiments are the most used and accepted methods and, thus, there is often a mistrust about the reliability of computational techniques. However, dry experiments also have their drawbacks. For example, the compound 4′-Methoxyacetophenone (CAS RN: 100-06-1), which is used as an additive and flavouring compound, and it is also included in the Food Contact Chemical DB (FCCDB), has two different predicted activities for its capability to act as an agonist for the estrogen receptor α. In fact, in the Tox21 project (Richard et al., 2021), the quantitative high-throughput screening assay (qHTS) identifies 4′-Methoxyacetophenone both as active and inactive for its agonist activity on ERα. In light of this, we think that there is not an approach that can be judged as better than another, but all are equally valid and should be considered together. Thus, the present work should not be seen as an opposing method to classical *in vitro* and *in vivo* tests, but it should be considered as a useful and preliminary method to screen a huge number of molecules in a cost and time-effective manner. In fact, using our robust computational method, we screened a large volume of molecules against the nuclear receptor family in a relatively short time when compared to the time needed for *in vitro* and *in vivo* experiments.

Author contribution statement

Pietro Cozzini – Conceptualization, Methodology, Project administration, Resources, Supervision, Writing reviewing/editing, Giulia Spaggiari – Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Francesca Cavaliere – Data curation, Formal analysis, Investigation, Methodology, Software Development, Validation, Writing – original draft.Marco Riani & Gianluca Morelli – Statistical methods and software development

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Bissantz, C., Folkers, G., Rognan, D., 2000. Protein-based virtual screening of chemical databases . 1 . Evaluation of different docking/scoring combinations. J. Med. Chem. 43, 4759–4767. https://doi.org/10.1021/jm001044l. Cavaliere, F., Lorenzetti, S., Cozzini, P., 2020. Molecular modelling methods in food safety: bisphenols as case study. Food Chem. Toxicol. 137 https://doi.org/10.1016/j. fct.2020.111116.

Cavaliere, Francesca, Spaggiari, G., Cozzini, P., 2020. Molecular Docking for Computer-Aided Drug Design: Fundamentals, Techniques, Resources and Applications. Dall'Asta, C., 2016. Mycotoxins and nuclear receptors: a still underexplored issue. Nucl. Recept. Res. 3 https://doi.org/10.11131/2016/101204.

De Coster, S., Van Larebeke, N., 2012. Endocrine-disrupting chemicals: associated disorders and mechanisms of action. J. Environ. Public Health 2012. https://doi. org/10.1155/2012/713696.

Desvergne, B., Feige, J.N., Casals-Casas, C., 2009. PPAR-mediated activity of phthalates: a link to the obesity epidemic? Mol. Cell. Endocrinol. 304, 43–48. https://doi.org/ 10.1016/j.mce.2009.02.017.

Fucic, A., Gamulin, M., Ferencic, Z., Katic, J., Krayer Von Krauss, M., Bartonova, A., Merlo, D.F., 2012. Environmental exposure to xenoestrogens and oestrogen related cancers: reproductive system, breast, lung, kidney, pancreas, and brain. Environ. Heal. A Glob. Access Sci. Source 11, 1–9. https://doi.org/10.1186/1476-069X-11-S1-S8.

Gore, A.C., Chappell, V.A., Fenton, S.E., Flaws, J.A., Nadal, A., Prins, G.S., Toppari, J., Zoeller, R.T., 2015. EDC-2: the endocrine society's second scientific statement on endocrine-disrupting chemicals. Endocr. Rev. 36, E1–E150. https://doi.org/ 10.1210/er.2015-1010.

Grulke, C.M., Williams, A.J., Thillanadarajah, I., Richard, A.M., 2019. EPA's DSSTox database: history of development of a curated chemistry resource supporting computational toxicology research. Comput. Toxicol. 12, 100096 https://doi.org/ 10.1016/j.comtox.2019.100096.

Hersey, A., Chambers, J., Bellis, L., Patrícia Bento, A., Gaulton, A., Overington, J.P., 2015. Chemical databases: curation or integration by user-defined equivalence? Drug Discov. Today Technol 14, 17–24. https://doi.org/10.1016/j. ddtec.2015.01.005.

Kabir, E.R., Rahman, M.S., Rahman, I., 2015. A review on endocrine disruptors and their possible impacts on human health. Environ. Toxicol. Pharmacol. 40, 241–258. https://doi.org/10.1016/j.etap.2015.06.009.

Kavlock, R.J., Daston, G.P., DeRosa, C., Fenner-Crisp, P., Gray, L.E., Kaattari, S., Lucier, G., Luster, M., Mac, M.J., Maczka, C., Miller, R., Moore, J., Rolland, R., Scott, G., Sheehan, D.M., Sinks, T., Tilson, H.A., 1996. Research needs for the risk assessment of health and environmental effects of endocrine disrupters: a report of the U.S. EPA-sponsored workshop. Environ. Health Perspect. 104, 715–740. https:// doi.org/10.1289/ehp.96104s4715.

Luccio-Camelo, D.C., Prins, G.S., 2011. Disruption of androgen receptor signaling in males by environmental chemicals. J. Steroid Biochem. Mol. Biol. 127, 74–82. https://doi.org/10.1016/J.JSBMB.2011.04.004.

Morris, G.M., Lim-Wilby, M., 2008. Molecular docking. Methods Mol. Biol. 443, 365–382. https://doi.org/10.1007/978-1-59745-177-2_19.

Odermatt, A., Gumy, C., 2008. Glucocorticoid and mineralocorticoid action: why should we consider influences by environmental chemicals? Biochem. Pharmacol. 76, 1184–1193. https://doi.org/10.1016/j.bcp.2008.07.019.

Petrakis, D., Vassilopoulou, L., Mamoulakis, C., Psycharakis, C., Anifantaki, A., Sifakis, S., Docea, A.O., Tsiaoussis, J., Makrigiannakis, A., Tsatsakis, A.M., 2017. Endocrine disruptors leading to obesity and related diseases. Int. J. Environ. Res. Public Health 14, 1–18. https://doi.org/10.3390/ijerph14101282.

Richard, A.M., Huang, R., Waidyanatha, S., Shinn, P., Collins, B.J., Thillainadarajah, I., Grulke, C.M., Williams, A.J., Lougee, R.R., Judson, R.S., Houck, K.A., Shobair, M., Yang, C., Rathman, J.F., Yasgar, A., Fitzpatrick, S.C., Simeonov, A., Thomas, R.S., Crofton, K.M., Paules, R.S., Bucher, J.R., Austin, C.P., Kavlock, R.J., Tice, R.R., 2021. The Tox21 10K compound library: collaborative chemistry advancing toxicology. Chem. Res. Toxicol. 34, 189–216. https://doi.org/10.1021/acs. chemrestox.0c00264.

Safe, S., 2004. Endocrine disruptors and human health: is there a problem. Toxicology 205, 3–10. https://doi.org/10.1016/j.tox.2004.06.032.

Schug, T.T., Janesick, A., Blumberg, B., Heindel, J.J., 2011. Endocrine disrupting chemicals and disease susceptibility. J. Steroid Biochem. Mol. Biol. 127, 204–215. https://doi.org/10.1016/j.jsbmb.2011.08.007.

Spaggiari, G., Iovine, N., Cozzini, P., 2021. In silico prediction of the mechanism of action of pyriproxyfen and 4′-oh-pyriproxyfen against a. Mellifera and h. sapiens receptors. Int. J. Mol. Sci. 22 https://doi.org/10.3390/ijms22147751.

Teramoto, R., Fukunishi, H., 2007. Supervised consensus scoring for docking and virtual screening. J. Chem. Inf. Model. 47, 526–534. https://doi.org/10.1021/ci6004993.

Trott, O., Olson, A.J., 2009. Software news and update AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31, 455–461.

Wang, R., Lu, Y., Wang, S., 2003. Comparative evaluation of 11 scoring functions for molecular docking. J. Med. Chem. 46, 2287–2303. https://doi.org/10.1021/ jm0203783. P. Cozzini et al.

CHAPTER 8

A focus on androgen receptor (AR) and estrogen receptors (ERα and ERβ)

A focus on androgen receptor (AR) and estrogen receptors (ERα and ERβ)

The food-contact chemical database was used to retrieve the three-dimensional structure of molecules which are used to screen the endocrine-disrupting properties against estrogen and androgen receptors using computational methods. In total, the database stores 11059 different molecules, belonging to the classes of additives, bisphenols, dioxins, flavourings, furans, mycotoxins, PCBs, pesticides, phthalates, and the molecules of the food packaging forum (FPF) DB. A molecular docking approach has been applied using four different scoring functions with the aim to use a consensus score (CS) prediction. To reach this goal, a robust statistical approach has been used to calculate the consensus scoring. Based on the value of the consensus scoring, the molecules have been categorized in high (≥ 0.8), medium ($0.3 \leq CS < 0.8$), and low interactors (< 0.3). As explained in Chapter 7, the robust analysis allows the identification of outliers' values, which are removed from the following analyses.

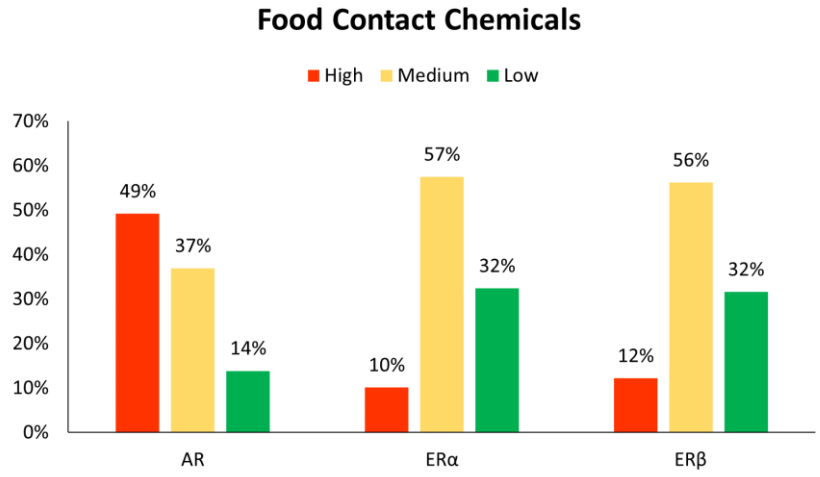


Figure 1: Percentage of all food contact chemicals predicted as high (red), medium (yellow), and low interactor (green) for the androgen receptor (AR) and the estrogen receptors (ERα and ERβ).

As can be seen in **Figure 1**, almost half of all food contact chemicals highly interact with the androgen receptor (49%) and only a small amount has been predicted as low interactor (14%). As might be expected, the two isoforms of estrogen receptors have similar results since they share a high level of sequence identity. Contrary to AR, the two estrogen

receptors have a lower number of molecules predicted as higher interactors (10% and 12% for ERα and ERβ, respectively). Nevertheless, more than half have been identified as medium interactors (57% and 56%). Considering the number of food contact chemicals, we are daily exposed to a mixture of these substances. The results show that almost 50% of food contact chemicals have potentially the chemical-physical properties to act as high or medium endocrine-disrupting compounds.

The term food contact chemicals refer to all molecules that come into contact with food. As explained in Chapter 1, it can be distinguished different classes of food contact chemicals. The foodchem DB, which has been developed in the present thesis, store ten different subclasses: additives, bisphenols, dioxins, flavourings, furans, mycotoxins, PCBs, pesticides, phthalates, and the molecules of the food packaging forum (FPF) DB (Groh et al., 2021). Thus, in order to have a deeper insight about which class of molecules mainly interferes with the three nuclear receptors, the high, medium, and low interactors have been evaluated for each food class and each receptor. The results are shown in Figure 2. As expected from the previous results, the androgen receptor is the NR which has a higher number of high interactors. The two isoforms of estrogen receptor show a similar trend, and the number of high, medium, and low interactors is similar. The only difference has been identified for the bisphenols class. The molecules belonging to this class seem to have a higher specificity for the alpha isoform than the beta one. In fact, there are more compounds predicted as higher interactors in ERα (41%) than in the ERβ (27%), whereas they have a similar number of medium interactors.

Three different classes of food contact chemicals can be considered as substances of very high concern (SVHC): dioxins, furans. and PCBs. In fact, all molecules are high interactors for the androgen receptor, and none of them has been predicted as low interactors for the three nuclear receptors. On the other side, the class of mycotoxin has the higher number of low interactors in the three nuclear receptors (54%, 69%, and 72% for AR, ERα, and ERβ, respectively).

The class of pesticides is also worthy of note because although the androgen receptor is the most affected with 68% of molecules predicted as high interactors, the estrogen receptors are also a target of these molecules.

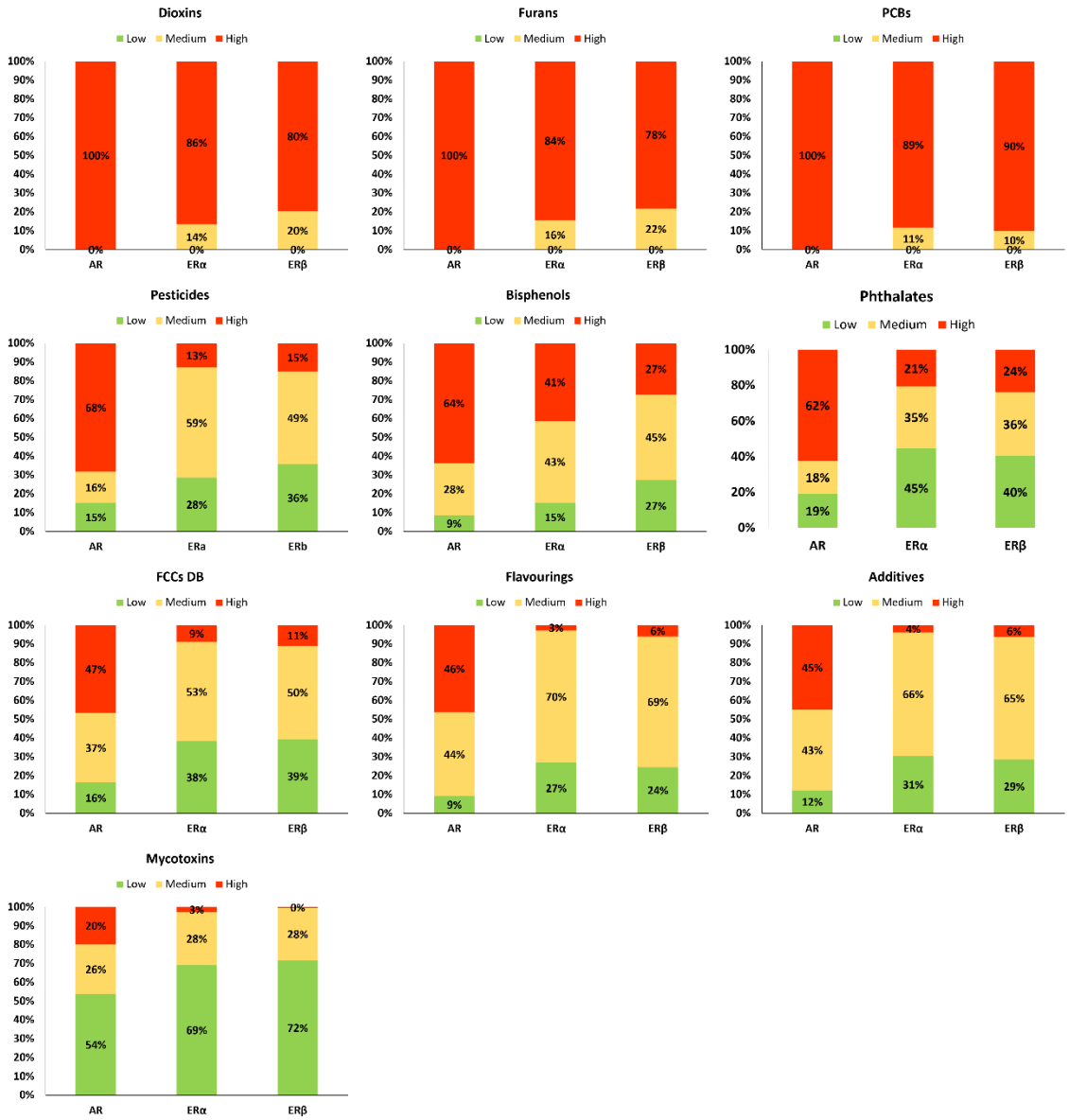


Figure 2. Percentage of molecules predicted as high (red), medium (yellow), and low interactor (green) for the androgen receptor (AR) and the estrogen receptors (ERα and ERβ) for each subclass of food contact chemicals

The following study evidences that the androgen receptor is the nuclear receptor with a higher number of molecules acting as high interactors. However, as explained before, people are daily exposed to a mixture of food contact chemicals, which can determine an additive harmful effect in the human body. For example, in the class of pesticides, 23 substances have been predicted as high interactors for all the three nuclear receptors, suggesting that they may interfere with both the estrogenic and androgenic pathways. Generally, 648 food contact chemicals are high interactors for all three nuclear receptors. Thus, considering their capability to interact with both the androgen and estrogen

receptors, these substances are of very concern for human health and should be considered as the first molecules to screen in future *in vitro*/*in vivo* experiments.

Additionally, to compare computational results and experimental data, the Comparative Toxicogenomics Database (CTD) (Davis et al., 2021) has been considered as a source of experimental data. The 648 food contact chemicals identified as higher interactors for all the three nuclear receptors have been searched on the CTD database. The analyses reports 321 experimental data. However, some chemicals are tested more than once for a specific nuclear receptor. Thus, only 43 different molecules have experimental data in the CTD database. In **Table 1** has been reported the molecule CAS RN, the experimental data and the consensus scoring values for each nuclear receptor. To make the table easy to understand, when experimental data reporting that the molecule "increases^activity" or "affects^binding" of the nuclear receptor was present, the "YES" entry has been added in the columns of experimental data.

Table 1. *In silico* prediction values and experimental data referring to the CTD database.

		Experimental Data			***In silico* Prediction**		
CLASSIFICATION	**CAR RN**	**AR**	**ERα**	**ERβ**	**AR**	**ERα**	**ERβ**
ADDITIVE	645-56-7		YES		0,88	0,80	0,87
BISPHENOLS	96-69-5		YES		0,93	0,86	0,86
DIOXINS	57653-85-7	YES			0,98	0,93	0,82
FCCs DB	129-00-0	YES	YES		0,97	0,81	0,88
FCCs DB	115-86-6	YES	YES	YES	0,95	0,96	0,98
FCCs DB	85-68-7	YES	YES	YES	0,94	0,88	0,96
FCCs DB	104-40-5	YES	YES	YES	0,94	0,95	0,96
FCCs DB	117-81-7	YES	YES	YES	0,93	0,88	0,85
FCCs DB	3648-21-3		YES	YES	0,93	0,91	0,90
FCCs DB	94-18-8		YES	YES	0,93	0,92	0,95
FCCs DB	92-69-3	YES	YES		0,92	0,90	0,93
FCCs DB	1806-26-4	YES	YES	YES	0,92	0,93	0,95
FCCs DB	1137-42-4		YES	YES	0,91	0,92	0,95
FCCs DB	80-46-6		YES	YES	0,89	0,89	0,88
FCCs DB	98-54-4		YES		0,88	0,87	0,87
FCCs DB	60207-90-1	YES			0,88	0,87	0,86
FCCs DB	94-26-8	YES	YES	YES	0,88	0,84	0,89
FCCs DB	140-66-9	YES	YES	YES	0,88	0,93	0,92
FCCs DB	99-89-8	YES			0,87	0,85	0,86
FCCs DB	4191-73-5		YES	YES	0,87	0,83	0,88

FCCs DB	59-50-7	YES			0,87	0,81	0,83
FCCs DB	94-13-3	YES	YES	YES	0,84	0,87	0,88
FURANS	57117-31-4	YES			0,97	0,92	0,91
PCBs	35065-28-2	YES			0,98	0,91	0,83
PCBs	38380-08-4	YES	YES		0,98	0,89	0,89
PCBs	38411-22-2	YES			0,98	0,96	0,95
PCBs	37680-73-2	YES			0,98	0,95	0,97
PCBs	38379-99-6	YES			0,98	0,95	0,95
PCBs	35065-27-1	YES	YES	YES	0,97	0,85	0,89
PCBs	52663-58-8	YES			0,97	0,92	0,90
PCBs	41464-40-8	YES			0,97	0,90	0,84
PCBs	35693-99-3	YES			0,97	0,95	0,96
PCBs	33284-54-7	YES			0,97	0,94	0,98
PCBs	38444-81-4	YES			0,96	0,83	0,96
PCBs	35693-92-6	YES			0,96	0,87	0,96
PCBs	13029-08-8	YES			0,95	0,82	0,94
PCBs	33146-45-1	YES			0,95	0,83	0,94
PCBs	34883-41-5	YES		YES	0,94	0,82	0,90
PESTICIDES	66063-05-6	YES			0,96	0,93	0,98
PESTICIDES	96489-71-3	YES	YES	YES	0,95	0,93	0,88
PESTICIDES	66332-96-5		YES		0,92	0,80	0,91
PESTICIDES	80844-07-1	YES			0,92	0,90	0,89
PHTHALATES	28553-12-0	YES	YES		0,97	0,81	0,82

Just to consider some of these molecules in more detail, the pesticide pyridaben (CAS RN: 96489-71-3) is a pyridazinone used as insecticide and acaricide which inhibit the mitochondrial NADH:ubiquinone reductase. This pesticide has been predicted by the *in silico* methods as a high interactor for all three nuclear receptors. In accordance with the computational prediction, experimental evidence shows that it increases the transcriptional activity of AR, ERα, and ERβ. Experimental studies have found that the pesticide flutolanil (CAS RN: 66332-96-5) affects the estrogen receptor alpha activity and it is correctly predicted as a high interactor of this receptor by the *in silico* prediction. In addition, it is also predicted as a high interactor of the estrogen receptor β and the androgen receptor. The diisononyl phthalate (CAS RN 28553-12-0) binds to AR protein and increases the activity of ERα according to *in silico* prediction. The triphenyl phosphate (CAS RN: 115-86-6), a food contact material used as a plasticizer and a fire

retardant, binds and inhibits the androgen receptor and increases the activity of the two isoforms of estrogen receptors in experimental studies. According to that, consensus scoring has ranked it as high interactors for all the three NRs. The above examples are just to give an overview about the accordance of consensus scoring and experimental data in order to highlight the utility of computational methods to screen large databases of compounds in a limited time and at low costs. Thus, considering that the 648 food contact chemicals have been predicted as high interactors of the androgen receptor and the estrogen receptors, a great attention to these molecules should be paid.

Reference

Davis, A.P., Grondin, C.J., Johnson, R.J., Sciaky, D., Wiegers, J., Wiegers, T.C., Mattingly, C.J., 2021. Comparative Toxicogenomics Database (CTD): update 2021. Nucleic Acids Res. 49, D1138–D1143. https://doi.org/10.1093/nar/gkaa891

Groh, K.J., Geueke, B., Martin, O., Maffini, M., Muncke, J., 2021. Overview of intentionally used food contact chemicals and their hazards. Environ. Int. 150, 106225. https://doi.org/https://doi.org/10.1016/j.envint.2020.106225

CHAPTER 9

Understanding the mechanism of action of R-bicalutamide on the W741L androgen receptor through molecular dynamic

Francesca Cavaliere and Pietro Cozzini

Under Review to

Toxicology and Applied Pharmacology

Preface

Molecular docking is a useful computational method that allows the screening large dataset of molecules against target proteins. Moreover, with the advent of high-performance computers (HPCs) these analyses can be carried out in a time-effective way. It has proven effective to study the interaction between endocrine-disrupting compounds and nuclear receptors, such as the estrogen receptors (α and β) and the androgen receptor. However, deciphering the mechanism of action using a molecular docking approach is not a simple task. In fact, while it is well-known that the estrogen receptor exists in two different states, a close and an open conformation, which is driven by the different position of helix 12, the same has only been suggested for the androgen receptor. If the purpose of a study is to understand the agonistic or antagonistic effect of an EDCs, molecular docking should be applied using the close and the open conformation of the receptors. However, the same cannot be performed for the androgen receptor for which an open conformation has not been solved. Thus, although molecular docking is a useful tool to screen the binding affinity of EDCs toward the androgen receptor, it does not give any information about their effect on the receptor in terms of agonist or antagonistic outcome. In the study reported below it has been studied the mechanism of action of antiandrogens which seem to affect the homodimer stability instead of inducing the opening of helix 12, as occurs in the estrogen receptor. The study may be seen in a perspective way to study the effect of food contact chemicals on the androgen receptor.

Understanding the mechanism of action of R-bicalutamide on the W741L androgen receptor through molecular dynamic

Francesca Cavalierea and Pietro Cozzinia*

a Molecular Modeling Lab, Department of Food and Drug, University of Parma, Parco Area Delle Scienze 17/A, 43124 Parma, Italy

*Corresponding author:

pietro.cozzini@unipr.it

Abstract

The androgen receptor (AR) plays an important role in the development and the progression of prostate cancer (PCa), and antiandrogens are used as first-line therapy. After a protracted drug administration, many patients develop a form of androgen insensitivity syndrome since some mutations occur in the ligand-binding domain. In this study, we explore the structural and dynamic effects of the antagonist R-bicalutamide in the homodimer stability, also considering the W741L mutation. When AR is bound to the natural ligand dihydrotestosterone, great homodimer stability can be evidenced. On the contrary, the binding of R-bicalutamide determines a general instability in each monomer of the AR dimer, suggesting its dissociation. The mutation W741L changes the effect of R-bicalutamide, and the system has a behavior that is halfway between an agonist and antagonist effect. These observations may explain the effect of the W741L in drug insensitivity from a structural point of view.

Keywords: androgen receptor; prostate cancer; androgen receptor insensitivity; molecular dynamic

Highlights

- Molecular dynamic simulation sheds light on the mechanism of action of the antiandrogen R-bicalutamide
- R-bicalutamide seems to affect the androgen receptor homodimer stability
- The induced W741L mutation in the androgen receptor restores homodimer stability even though it is bound to R-Bicalutamide

9.1 Introduction

Prostate cancer (PCa) is one of the most common cancer among men and it accounts for nearly one-third of all male cancers worldwide (https://gco.iarc.fr/). Additionally, PCa is the fifth leading cause of cancer death [1].

The androgen receptor (AR) belongs to the steroid receptor transcriptional factor of nuclear receptors (NRs). AR consists of three main domains: i) the N-terminal transcriptional regulation domain (NTD); ii) the DNA binding domain (DBD); iii) and the ligand-binding domain (LBD) that comprises the activation function 2 (AF2) region [2]. After the binding of its native ligand (testosterone or dihydrotestosterone (DHT)) to the LBD, the receptor translocates into the nucleus, where it likely forms a homodimer. The receptor binds to the androgen-response elements (ARE) of promoter regions and activates the transcription of targeted genes. Alterations in its function can have a profound effect on carcinogenesis and tumor growth [3] leading to the development and the progression of several diseases, such as androgen insensitivity syndrome (AIS) and PCa [4]. In fact, AR is highly expressed in prostate cancer cells [1] and its activation can promote the development and the progression of PCa. Therefore, prostate cancer is often treated with drugs acting as antiandrogens which block the action of the endogenous ligand [5].

There are two main classes of antiandrogens, steroidal and non-steroidal molecules. This latter class of antiandrogens has been developed for avoiding the off-target effects of the steroidal agents and seems to bind specifically the androgen receptor [6]. R-bicalutamide belongs to the class of non-steroidal first-generation antiandrogen used to treat prostate cancers (PCs). They prevent AR downstream signaling through the inhibition of AR translocation to the nucleus by competitive binding to the receptor [7]. Although they are very effective at the beginning, patients become frequently resistant to drug administration after several years, developing a metastatic form of castration-resistant PCa (CRPC). Point mutations in the AR ligand-binding domain (LBD) and/or the expression of active AR splicing variants result in constitutive activation of AR signaling, switching first-generation anti-androgens to become agonists or AR partial agonists [8]. For example, R-bicalutamide acts as an agonist in presence of the single point mutation W741L in the LBD of AR.

To investigate the mechanisms of drug resistance and to explore the effect of this point mutation on R-bicalutamide interaction, Liu et al. [5] have applied molecular dynamics

simulations and molecular mechanics generalized Born surface area calculation considering both the wild-type and the mutated protein. They found that in the presence of W741L mutation, R-bicalutamide interacts with the receptor with a lower free binding energy. Moreover, the substitution of the tryptophane with leucine generates a greater space in the binding pocket, promoting the helix 12 (H12) movement close to the LBD as the agonist conformation. It is well-known that helix 12 has an important role in the regulation of agonist/antagonist function. The estrogen receptor shows two different conformations: a close (agonist) conformation where the helix 12 is close to the body of the LBD, and an open (antagonist) conformation where H12 is found away from the LBD body. Although the crystal structure of the wild-type androgen receptor in the antagonist conformation is not yet solved, it has been suggested that helix 12 of AR undergoes the same movement after antagonist binding. However, the androgen receptor has an additional C-terminal region, arranged as a β-sheet, at the end of H12 that seems firmly anchor the helix on the receptor, leading to significant doubts about the H12 movement. Different authors have attempted to understand the effect of antiandrogens on helix 12 using molecular dynamics, but only slight modification in its position has been highlighted [5,9,10]. After the crystallization of the androgen receptor homodimer, the hypothesis that antiandrogens can destabilize the dimer complex has emerged. Homodimer is fundamental for androgen receptor activity: when the natural ligand binds to the androgen receptor, AR forms a homodimer, and it can migrate inside the nucleus. More than forty mutations associated with androgen insensitivity syndrome (AIS) and PCa are located to the dimer interface supporting the important role of dimer formation on the activity of the receptor. After that, Shizu et al. [11] have evaluated the effect of Pro767Ala mutation on dimer formation using both *in vitro* and *in silico* analysis. Pro767 residue is one of the fundamental amino acids in the homodimer interface and its mutation in Ala767 decreases the capability of the androgen receptor to form a homodimer. Consequently, they found a lower concentration of AR translocation from the cytoplasm to the nucleus and hence a lower transcriptional activity. Although the work of Shizu et al. has highlighted the important role of the dimer formation for the receptor activity, they have not addressed the effect of point mutations on the dimer stability in presence of an antagonist. We suggested that the antagonist binding may affect the dimer stability instead of inducing a helix 12 movement. To understand what happens from a protein point of view, we have applied molecular dynamic simulations on the AR homodimer in complex

with the natural ligand (DHT), and the R-bicalutamide. It has been also analysed the effect of the point mutation W741L on the dimer stability when it is bound to R-bicalutamide. In fact, AR^{W741L} is insensitive to the inhibition effect of R-bicalutamide, which acts as an agonist or partial agonist in this case.

In this paper, we suggested that the antagonistic effect may be induced by the homodimer dissociation. Moreover, we have also suggested a possible explanation of the changed effect performed by R-bicalutamide after the induced W741L mutation.

9.2 Materials and Methods

9.2.1 System setup

Three different systems were studies:

1) Dimer AR - DHT: the crystallographic structure (PDB: 5JJM, R = 2.15Å) [12] of the androgen receptor dimer in complex with dihydrotestosterone (DHT) was obtained from the Protein Data Bank (PDB) [13,14] and the core-dimer conformation was used.

2) Dimer AR - R-bicalutamide: the crystallographic structure 1Z95 (R = 1.80Å) [15] refers to the androgen receptor with the point mutation W741L bound to the R-bicalutamide. Thus, to simulate the position of R-bicalutamide into the wild-type dimer protein (5JJM), the crystallographic R-bicalutamide (1Z95) was extracted and superimpose into the binding pocket of each monomer. As it could be expected, R-bicalutamide clashes with the W741 residue. To also consider the movement of this residue, the 2AXA structure (R = 1.80Å) [16] was used as the reference structure to remove these protein-ligand clashes. After that, the residues of Trp741, Met895, Met895 were energy minimized through Sybyl software v8.1 (www.tripos.com) using the Powell algorithm with a coverage gradient of ≤0.5 kcal (mol Å)−1 and a maximum of 1500 cycles. The main purpose of this slight energy minimization was to remove other potential clashes.

3) Dimer AR^{W741L} – R-bicalutamide: the structure at point 2 was used as starting structure. The mutation W741L was performed in each monomer using PyMOL (The PyMOLMolecular Graphics System, Version 2.5 Schrödinger, LLC).

9.2.2 Molecular dynamic simulation

Molecular dynamics simulations were performed in triplicate for a total simulation time of 500ns. Three different systems of the dimer androgen receptor were studied: a) dimer AR in complex with the natural ligand DHT in both the two monomers; b) dimer AR in

complex with the antagonist R-bicalutamide in both to the two chains; c) dimer AR with the point mutation W741L bound to the antagonist R-bicalutamide. The systems were simulated using GROMACS 2019.4 using the CHARMM36m forcefield [17–19]. The topology of the ligands was generated using CGenFF online server [20,21]. The systems were solvated using TIP3P water model in an octahedron box with a minimum distance of 1.5 nm from the edge of the protein. To maintain the overall neutrality of the system, 9 CL atoms were added as counterions. The minimization of the three systems was performed with the steepest descent minimization algorithm for 0.1ns with a maximum force of 1000kJmol^{-1}. The systems were gradually heated to reach the final temperature of 300K in three steps using the NVT mode (0-100K (0.1ns), 100-200K (0.1ns), 200-300K (0.1ns)), and were further equilibrated at a constant pressure (1.0 bar) for 0.5 ns (NPT) using the V-rescale temperature coupling method and Berendsen thermostat. For computing the long-range electrostatic interactions, the Particle Mesh Ewald method was used. Short-range forces were calculated using Verlet cutoff scheme with a minimum cutoff set to 1.2 nm. The systems were then simulated for a total production run of 500ns for each replicas. Once the MD was completed, translational and rotational movements were removed. The root mean square deviation (RMSD) and root mean square fluctuation (RMSF) were calculated using the *gmx rmsdist* and *gmx rmsf*, respectively. Moreover, to calculate the number of hydrogen bonds in the protein-ligand complex, and between the two monomers, the *gmx hbond* was used.

9.2.3 Essential molecular dynamic

The essential molecular dynamic was performed using the trajectories generated during the MD simulation, considering only backbone atoms. The MD starting structure has been used as a reference template for the trajectory superposition. Calculations have been performed using the *gmx covar* command of GROMACS. The resulting eigenvalues and eigenvectors were used to study the overall flexibility of the systems.

9.4 Results and Discussion

In the present study, we have applied molecular dynamic simulation to analyse the conformational changes induced by the R-bicalutamide (antagonist) binding on the AR homodimer. It is well known that after a prolonged administration of this drug, patients frequently undergo drug insensitive due to the development of the W741L mutation located in the ligand-binding site of the androgen receptor. In the presence of that point

mutation, the antagonist R-bicalutamide acts as an agonist or a partial agonist inducing the propagation of the disease.

Three different systems have been simulated in triplicate for a total simulation time of 500ns: (*a*) the wild-type AR homodimer bound to DHT and (*b*) R-bicalutamide, and (*c*) the mutated W741L AR homodimer bound to R-bicalutamide. For simplicity, in the following analyses, we will refer to the AR homodimer as AR and to R-bicalutamide as Bicalutamide or Bic. Firstly, the heavy atom RMSDs as a function of simulation time has been calculated to monitor the dimer stability of the three systems. As shown in **Figure 1A**, the two systems AR-DHT and AR^{W741L}-Bic reached stability very quickly (~50ns) and earlier than the system AR-Bic which reached stability only after ~100ns.

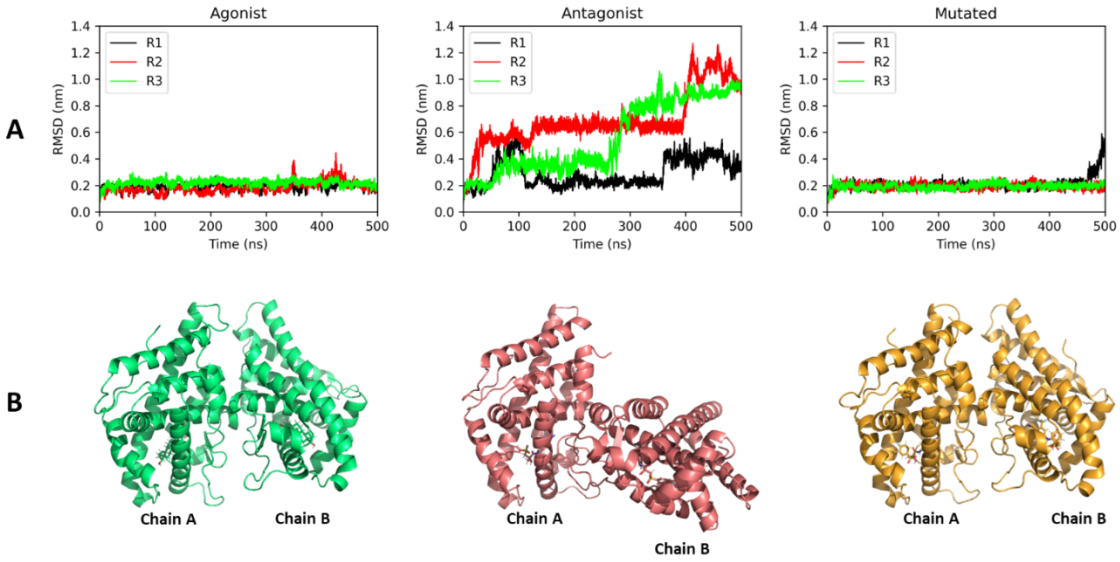


Figure 1: (**A**) RMSD calculations of the homodimer in complex with DHT (left), R-bicalutamide (center), and the mutated W741L homodimer in complex with R-bicalutamide. In black, red, and green is shown the RMSD trend for the replica 1, 2, and 3, respectively. (**B**) A snapshot of the three complexes at the end of the molecular dynamic simulation (from left to right: AR-DHT, AR – R-bicalutamide, and ARW741L – R-bicalutamide).

Moreover, while the first two systems are generally stable in the three replicas, the binding of R-bicalutamide to the wild-type homodimer determines a greater instability and the RMSD profile shows an evident soar after ~250-300ns. This behavior may indicate that the antagonist may introduce some conformational changes in the protein inducing a destabilization in the whole system, probably affecting the homodimer stability. According to our hypothesis, the presence of the point mutation W741L in the androgen receptor might reverse the activity of the antagonist since the system is very stable during the simulation time and the RMSD trend is very close to the agonist system.

From the point of view of ligands it can be seen that while dihydrotestosterone seems to be very stable during all the simulation time for the three replicas, the R-bicalutamide has a greater perturbation in both the systems, i.e., with and without W741L mutation (**Figure 2**).

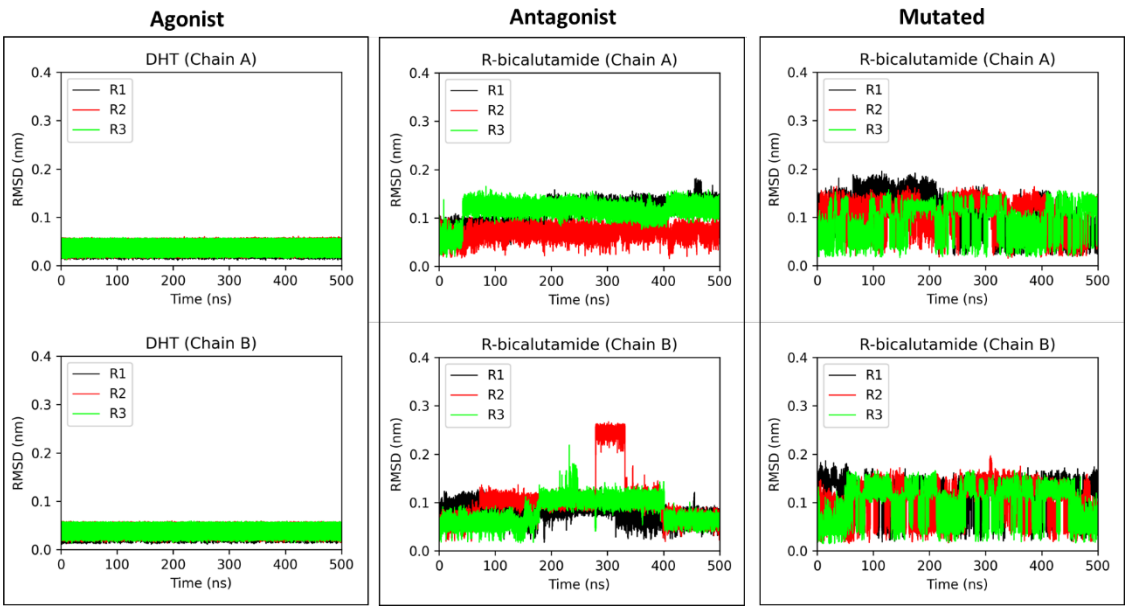


Figure 2: RMSD calculation of the ligands during the simulation time. On the top and bottom is shown the RMSD in chains A and B of the homodimer, respectively. From left to right, the complexes AR-DHT, AR – R-bicalutamide, and ARW741L – R-bicalutamide, respectively.

To have a deep insight into the reasons for this instability, the overall occupancies of H-bonds (**Supplementary Figure 1**) have been calculated. For the H-bonds analyses, we have considered the replicas with the higher stability, i.e., R2, R1, and R2 for the systems Agonist, Antagonist, and Mutated, respectively. The agonist ligand (DHT) forms mainly two hydrogen bonds with residues ASN^{705} and THR^{877}, which are maintained for ~57-64%- and ~30-50% of the simulation time for Chain A and Chain B, respectively. By contrast, the antagonist R-bicalutamide interacts with the receptor establishes two hydrogen bonds with the residues TRP^{741} and ASN^{705}. The point mutation W741L prevents the interaction of R-bicalutamide with residue 741 and it interacts mainly with the residue ASN^{705}. This might justify why R-bicalutamide in the system with the mutation has greater fluctuations in its RMSD compared to the agonist systems. In fact, while it is anchored from one side to the ASN^{705} residue, it, rather, is allowed to move from the other side.

Finally, to further evaluate the stability between the two monomers, the number of hydrogen bonds has been calculated during the simulation time. The two monomers establish ~7-8 hydrogen bonds in the agonist system in the three replicas. The binding of the antagonist drastically drops the number of H-bonds (~3-4) highlighting that the two monomers interact less tightly compared to the agonist system. The induced mutation W741L in the ligand-binding pocket enhances the stability between the two monomers when R-bicalutamide binds to AR. In fact, the number of hydrogen bonds rises to a value of ~6 (in-between agonist and antagonist systems). This may suggest that the binding of R-bicalutamide generates an overall instability (as is also evidenced by the RMSD trend) in the two monomers which might induce the homodimer dissociation.

To evaluate the flexibility of individual residues, the RMSF of the protein backbone atoms has been computed considering separately the two chains of the homodimer (**Supplementary Figure 2**). The three complexes share similar RMSF distributions, and their fluctuations have extremely similar trends in all three replicas. No differences can be evidenced for the residues of helix 12, showing that the R-bicalutamide does not affect H12 positions. To reveal monomer rearrangements from MD, the homodimer structures have been captured at the end of MD simulation and compared to the crystal structure. As shown in **Figure 1B**, it is very clear the effect of the antagonist binding whose induces a complete loss of the homodimer symmetry. The mutation W741L, on the contrary, retains the symmetry of the homodimer. This might explain the effect of the point mutation on drug function. In fact, the loss of symmetry evidenced in the antagonist system might suggest that the R-bicalutamide binding destroys the monomers' interaction, preventing the activation of gene transcription. AR^{W741L} in complex with R-bicalutamide may prevent homodimer dissociation. In this way, the drug might act as an agonist or a partial agonist instead of an antagonist. The following results might explain by a structural point of view the mechanism of action of the antiandrogen R-bicalutamide on the wild-type AR and the mutated AR^{W741L}.

To gain deeper insight into the conformational changes promoted upon the agonist and antagonist binding on the dimer stability, and to evaluate the effect of the induced point mutation W741L on the R-bicalutamide activity, essential molecular dynamics was used. PCA analyses were performed using the most stable molecular dynamic simulations among the three replicas considering the RMSD average value of the entire system (*replica 2, 1,* and *2* for AR-DHT, AR-Bic, and AR^{W741L}-Bic complexes, respectively).

The eigenvalue contributions given by PCA for the three systems are illustrated in **Supplementary Figure 3**, and the results show that the first four components account for most of the cooperative motions. The first four PCs cover around 92% of the total variance in the AR-Bicalutamide system, where the first eigenvector accounts for more than 70% of all protein fluctuations. On the contrary, they cover 41% and 60% of all protein motion for DHT-AR and R-bicalutamide-AR^{W741L} systems, respectively. However, although all the PCs are involved in the collective motion, the contribution to the total variance diminishes rapidly after the first few eigenvectors in the three systems. The result suggested that when the R-bicalutamide interacts with the mutated form of the androgen receptor, the system showed lower fluctuation than when it binds the wild-type form of AR and its behaviour is very close to the agonistic ligand. From the analysis of the conformational space visited by the three systems, it is clearly evident that the extent of sampling of the three systems is significantly different (**Figure 3**).

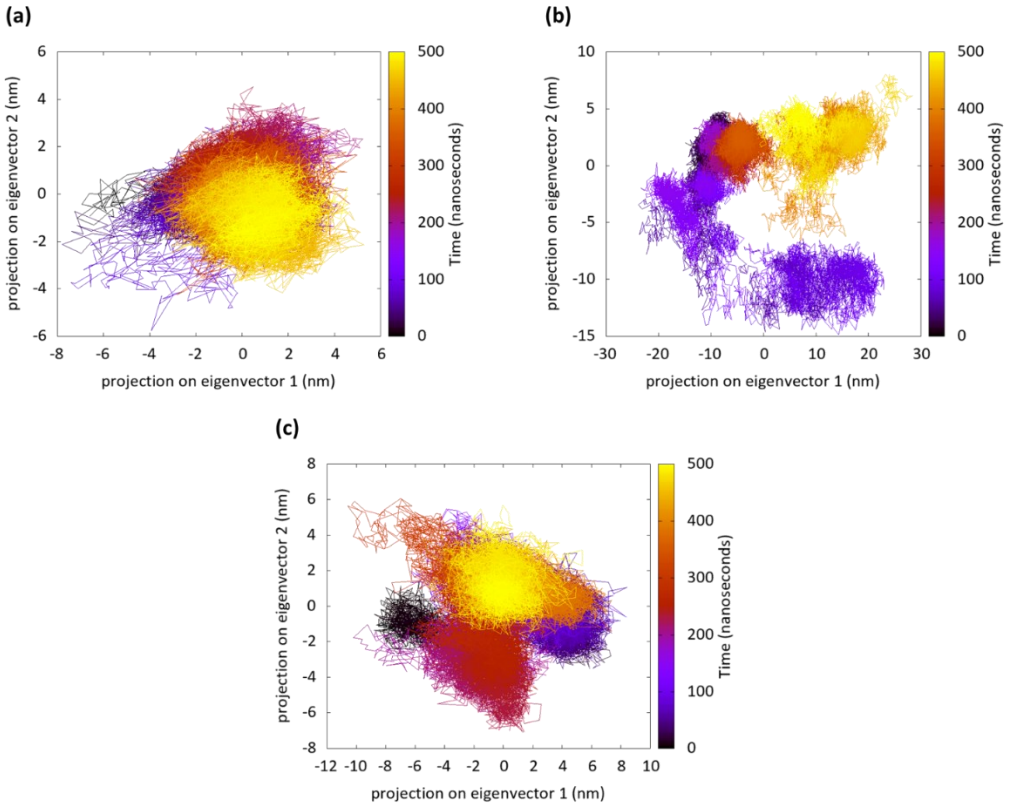


Figure 3: Cross plots of the first two principal components (PC1 and PC2) for the three systems: (a) AR-DHT, (b) AR-R-bicalutamide, (c) ARW741L – R-bicalutamide. The color code represents the time evolution.

The wild-type homodimer spans a small area when the natural ligand DHT is bound in each monomer, highlighting the great stability of the system during the simulation time. On the contrary, the wild-type homodimer spans a larger conformational space when it is bound with R-bicalutamide compared to the DHT system. This suggests the presence of high essential dynamics, that may be essential for the antagonistic action of the drug. Whereas, after the W741L mutation occurs in the receptor, the conformational landscape of the AR^{W741L}–Bicalutamide system is confined to a smaller area as compared to its wild-type counterparts (AR-Bicalutamide). However, it explores a greater space than the agonist system. These two evidences suggest that R-bicalutamide may act as a partial agonist after the mutation W741L has occurred.

The first principal component for the three systems is illustrated in **Supplementary Figure 4**. As it is evident the most prominent motions can be observed in the wild-type system in complex with the antagonist ligand, where the two monomers are directed opposite. The same behavior can be seen for the other two systems, but the extension of the motion between the two monomers is very lower than that observed in the antagonist system. Interestingly, the mutated homodimer in complex with R-bicalutamide shows a movement that is very close to the wild-type system in complex with the natural ligand, suggesting again an agonistic/partial agonistic mechanism of action. Moreover, while the movement in the AR-bicalutamide complex involves the entire monomer, in the AR-DHT the movement involves only the upper and bottom part of the receptor. In addition, the AR^{W741L}-Bicalutamide complex shows a trend similar to the agonistic system, and the motion involves only a little part of the receptor, although the essential motion explored by the receptor is greater than the agonistic system.

9.5 Conclusion

A recent structure of AR homodimer in complex with the natural ligand DHT has been recently solved [12]. The authors have highlighted that the majority of AIS-associated mutations are found on the core dimer interface, suggesting the important role of the homodimer formation for the androgen receptor function. Shizu et al. [11] have shown that AR exists as a homodimer in the cytoplasm and the binding of DHT allows the translocation of the receptor into the nucleus where it activates the gene transcription. Prostate cancer is the most common cancer in men all over the world and androgen plays an important role in its proliferation. For that reason, PCa is often treated with androgen

deprivation therapy. Non-steroidal antiandrogens are a class of drugs used to treat PCa and they specifically interact with the androgen receptor, inducing its de-activation. Among them, R-bicalutamide is one of the most used AR antagonists. However, after a protracted administration, patients become resistant to its treatment for the development of the W741L mutation into the ligand-binding pocket of AR. In this condition, R-bicalutamide acts as an agonist instead of having an antagonistic effect, promoting the progression of PCa. To study the effect of the point mutation on the homodimer, we have applied an *in silico* approach. Three different systems have been simulated for 500ns: (*a*) AR homodimer in complex with DHT, (*b*) AR homodimer in complex with R-bicalutamide, and (*c*) the mutated AR^{W741L} homodimer in complex with R-bicalutamide. Moreover, to evaluate large-scale motions in three systems during the explored simulation time, we performed a global principal component analysis (PCA), called also essential dynamics. We found that the homodimer complex is very stable during all the simulation time when the natural ligand is bound to the wild-type protein. On the contrary, the binding of R-bicalutamide induces an enormous instability in the complex, as evidenced by the trend of RMSD. Following the W741L mutation, the homodimer in complex with R-bicalutamide showed an increase stability and its behaviour was similar to the agonistic effect induced by DHT in the wild-type receptor. This result suggests that the mutation is involved in retaining the monomer-monomer stability, maybe allowing the translocation of the receptor into the nucleus. This hypothesis was also confirmed by the essential dynamics, where it was evident that the binding of the antagonist to the wild-type protein induced a high instability in the homodimer which probably causes the dissociation of the two monomers. In fact, the first principal component showed that the two monomer moves a lot in an opposite direction. This great movement was much lower evident in the mutated protein, suggesting that the mutation contributes to the change of the R-bicalutamide function, which becomes an agonist or a partial agonist. The agonist system has marked stability compared to the other two systems, suggesting that R-Bicalutamide may likely act as AR partial agonist rather than as an agonist. In fact, its behaviour was halfway between the agonistic and antagonistic effect.

To date, it has been thought that antagonist compounds affect the helix 12 (H12) orientation as it is well-known for the estrogen receptor. However, the presence of an additional β-sheet at the C-terminal of the H12 firmly anchors the helix to the body of the receptor, leading to significant doubt about the H12 reposition after the antagonist

binding. In this paper, we suggested a different mechanism of action. We found that the binding of R-bicalutamide influenced the homodimer stability, probably inducing its dissociation. The induced mutation W741L, which was at the basis of R-bicalutamide insensitivity, partly reverted the effect of the drug on the homodimer, which was more stable than the wild-type counterpart, and quite similar to the effect of the agonist.

Reference

[1] G. Aurilio, A. Cimadamore, R. Mazzucchelli, A. Lopez-Beltran, E. Verri, M. Scarpelli, F. Massari, L. Cheng, M. Santoni, R. Montironi, Androgen Receptor Signaling Pathway in Prostate Cancer: From Genetics to Clinical Applications., Cells. 9 (2020). https://doi.org/10.3390/cells9122653.

[2] R.A. Davey, M. Grossmann, Androgen Receptor Structure, Function and Biology: From Bench to Bedside, Clin. Biochem. Rev. 37 (2016) 3–15. https://pubmed.ncbi.nlm.nih.gov/27057074.

[3] T. Morova, D.R. McNeill, N. Lallous, M. Gönen, K. Dalal, D.M. Wilson, A. Gürsoy, Ö. Keskin, N.A. Lack, Androgen receptor-binding sites are highly mutated in prostate cancer, Nat. Commun. 11 (2020) 832. https://doi.org/10.1038/s41467-020-14644-y.

[4] M.H.E. Tan, J. Li, H.E. Xu, K. Melcher, E. Yong, Androgen receptor: structure, role in prostate cancer and drug discovery, Acta Pharmacol. Sin. 36 (2015) 3–23. https://doi.org/10.1038/aps.2014.18.

[5] H. Liu, X. An, S. Li, Y. Wang, J. Li, H. Liu, Interaction mechanism exploration of R-bicalutamide/S-1 with WT/W741L AR using molecular dynamics simulations, Mol. Biosyst. 11 (2015) 3347–3354. https://doi.org/10.1039/C5MB00499C.

[6] Y. Chen, N.J. Clegg, H.I. Scher, Anti-androgens and androgen-depleting therapies in prostate cancer: new agents for an established target, Lancet. Oncol. 10 (2009) 981–991. https://doi.org/10.1016/S1470-2045(09)70229-3.

[7] M.A. Rice, S. V Malhotra, T. Stoyanova, Second-Generation Antiandrogens: From Discovery to Standard of Care in Castration Resistant Prostate Cancer, Front. Oncol. 9 (2019) 801. https://doi.org/10.3389/fonc.2019.00801.

[8] N. Ammannagari, S. George, Anti-Androgen Therapies for Prostate Cancer : A Focused Review, Am J Hematol Oncol. 11 (2014) 15–19.

[9] S. Sakkiah, R. Kusko, B. Pan, W. Guo, W. Ge, W. Tong, H. Hong, Structural Changes Due to Antagonist Binding in Ligand Binding Pocket of Androgen Receptor Elucidated Through Molecular Dynamics Simulations , Front. Pharmacol. . 9 (2018) 492.

https://www.frontiersin.org/article/10.3389/fphar.2018.00492.

[10] M. Duan, N. Liu, W. Zhou, D. Li, M. Yang, T. Hou, Structural Diversity of Ligand-Binding Androgen Receptors Revealed by Microsecond Long Molecular Dynamics Simulations and Enhanced Sampling, J. Chem. Theory Comput. 12 (2016) 4611–4619. https://doi.org/10.1021/acs.jctc.6b00424.

[11] R. Shizu, K. Yokobori, L. Perera, L. Pedersen, M. Negishi, Ligand induced dissociation of the AR homodimer precedes AR monomer translocation to the nucleus, Sci. Rep. 9 (2019) 16734. https://doi.org/10.1038/s41598-019-53139-9.

[12] M. Nadal, S. Prekovic, N. Gallastegui, C. Helsen, M. Abella, K. Zielinska, M. Gay, M. Vilaseca, M. Taulès, A.B. Houtsmuller, M.E. van Royen, F. Claessens, P. Fuentes-Prior, E. Estébanez-Perpiñá, Structure of the homodimeric androgen receptor ligand-binding domain, Nat. Commun. 8 (2017) 14388. https://doi.org/10.1038/ncomms14388.

[13] P.W. Rose, A. Prlić, C. Bi, W.F. Bluhm, C.H. Christie, S. Dutta, R.K. Green, D.S. Goodsell, J.D. Westbrook, J. Woo, J. Young, C. Zardecki, H.M. Berman, P.E. Bourne, S.K. Burley, The RCSB Protein Data Bank: Views of structural biology for basic and applied research and education, Nucleic Acids Res. 43 (2015) D345–D356. https://doi.org/10.1093/nar/gku1214.

[14] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank., Nucleic Acids Res. 28 (2000) 235–242. https://doi.org/10.1093/nar/28.1.235.

[15] C.E. Bohl, W. Gao, D.D. Miller, C.E. Bell, J.T. Dalton, Structural basis for antagonism and resistance of bicalutamide in prostate cancer., Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 6201–6206. https://doi.org/10.1073/pnas.0500381102.

[16] C.E. Bohl, D.D. Miller, J. Chen, C.E. Bell, J.T. Dalton, Structural Basis for Accommodation of Nonsteroidal Ligands in the Androgen Receptor, J. Biol. Chem. 280 (2005) 37747–37754. https://doi.org/10.1074/jbc.M507464200.

[17] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B.L. de Groot, H. Grubmüller, A.D. MacKerell Jr, CHARMM36m: an improved force field for folded and intrinsically disordered proteins, Nat. Methods. 14 (2017) 71–73.

https://doi.org/10.1038/nmeth.4067.

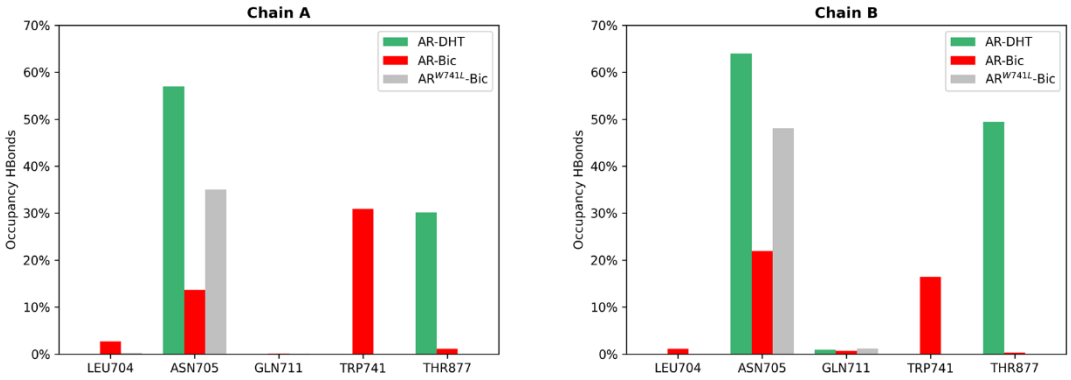
[18] P. Szilárd, M.J. Abraham, C. Kutzner, B. Hess, E. Lindahl, Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS, (2015). https://doi.org/10.1007/978-3-319-15976-8_1.

[19] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX. 1–2 (2015) 19–25. https://doi.org/https://doi.org/10.1016/j.softx.2015.06.001.

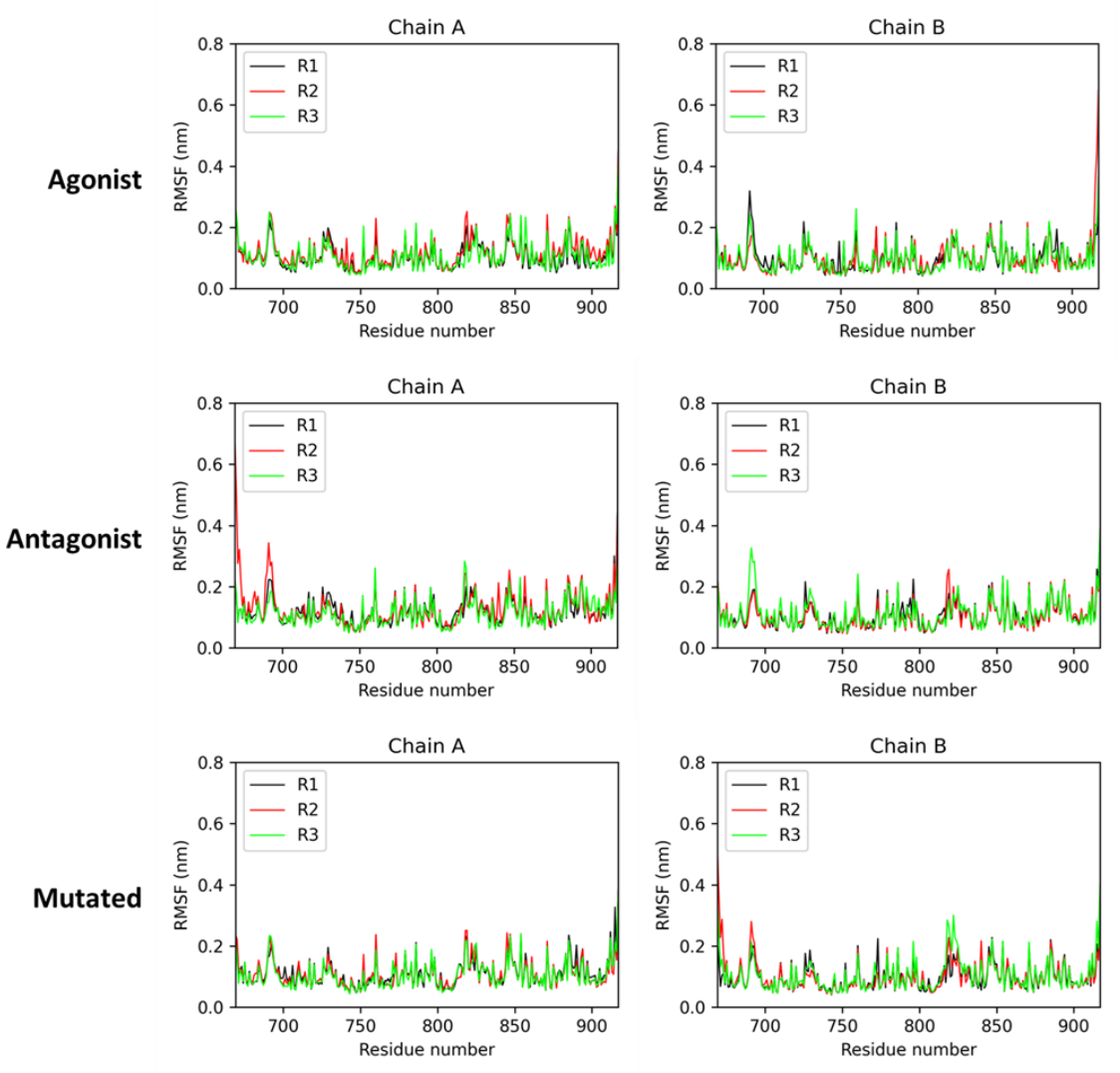
[20] K. Vanommeslaeghe, A.D. MacKerell, Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing, J. Chem. Inf. Model. 52 (2012) 3144–3154. https://doi.org/10.1021/ci300363c.

[21] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, A.D. Mackerell Jr, CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields, J. Comput. Chem. 31 (2010) 671–690. https://doi.org/10.1002/jcc.21367.

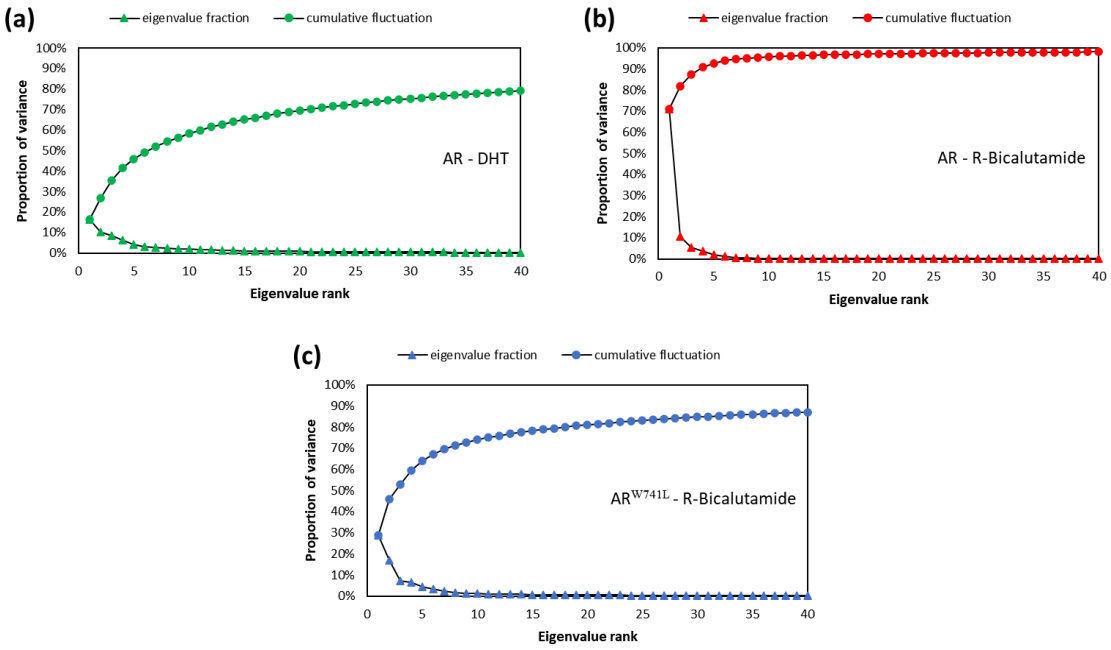
Supplementary Information



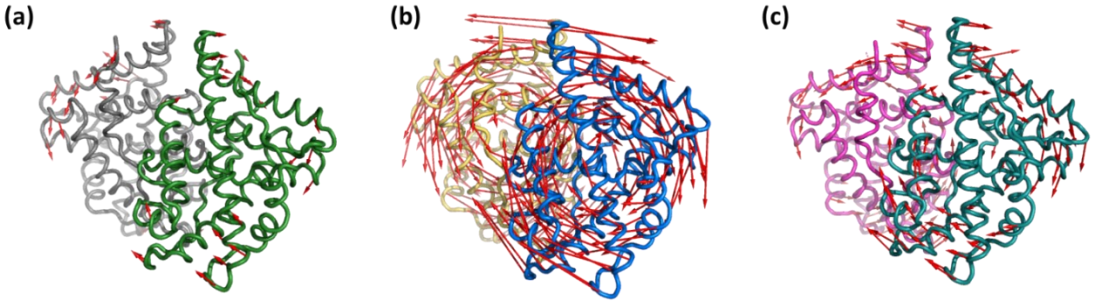
Supplementary Figure 1:The hydrogen bond occupancy (HBonds) for AR-DHT, AR – R-bicalutamide, and ARW741L – R-bicalutamide. The androgen receptor residues involved in the ligands binding are given on the X-axis and the occupancy (%) of H-Bonds is shown on the Y-axis. On the left and the right, it is illustrated the H-bonds in Chain A and Chain B, respectively.



Supplementary Figure 2: Residue-based RMSFs relative to the initial structure for each system considering the three MD trajectory replicas for the homodimer bound to the natural ligand DHT, the antagonist R-bicalutamide and for ARW741L bound to the antagonist R-bicalutamide.



Supplementary Figure 3: The proportion of the eigenvalue contribution of the eigenmodes (PCs) to the variance of the covariance matrix and the cumulative contribution for each eigenmode (descending order) for (a) AR-DHT, (b) AR-Bicalutamide, (c) ARW741L – Bicalutamide.



Supplementary Figure 4: The motion direction on the first principal component of the three systems: (a) AR-DHT, (b) AR-Bicalutamide, (c) ARW741L – Bicalutamide. The proteins are shown as cartoons and the two monomers are colored differently in each system. The arrows attached to each backbone atom indicate the direction of the eigenvector and the size of each arrow shows the magnitude of the corresponding eigenvalue.

CHAPTER 10

Future perspective studies

10.1 Study 1 - Protein-ligand interaction using Isothermal Titration Calorimetry (ITC)

In silico methods, such as molecular docking, consensus scoring, and molecular dynamics, are very useful to screen large datasets of molecules in a short time and a cost-effective manner. Although a lot of papers prove the reliability of these new techniques, some doubts still remain about their correlation with experimental results. One of the main limitations to correlate *in silico* methods with *in vitro/in vivo* experiments consists in the different endpoints considered by the two methods. In fact, computational methods study protein-ligand interaction, instead, experimental techniques evaluate the effects of a compound in the transcription levels, on the cell, on the whole organisms, etc. Thus, since the experiment is set out in an entire cell or organism, it is often difficult to correlate the results with a protein-ligand interaction.

Isothermal Titration Calorimetry (ITC) is a technique that directly measures the heat released or absorbed during a chemical reaction or along with a biomolecular reaction. Thus, it can easily measure the heat changes when two or more molecules in solution interact together under constant temperature. Based on the type of binding, the heat can be released for an exothermic reaction, or absorbed for an endothermic reaction.
The instrument mainly consists of two cells, one for the macromolecule and one used as the reference cell that contains the solvent. Both the cells are maintained at constant pressure and temperature. When the ligand is added and tritated in the main cell, the macromolecule and the ligand may interact. If there is an interaction between the protein and the ligand, there will be a heat discharge or consumption, which produces a change in the temperature. To maintain a constant temperature, the instrument provides energy that is proportional to the strength of the interaction. Integrating the energy power over time, it is possible to calculate the heat change and then the enthalpy of the reaction.
Thus, isothermal titration calorimetry might be a more suitable experimental technique to evaluate the direct protein-ligand interaction allowing to better correlate *in silico* prediction.

In chapter 6, it has been studied the interaction of several pesticides with the ligand-binding domain (LBD) of the RXR receptor of *Daphnia Magna*, where pyriproxyfen and its metabolite have shown a high affinity for the receptor. Since *D. magna* and *H. sapiens* show a high degree of amino acid sequence identity, we suggested that the pesticide can

also be a good binder of the human counterpart. Thus, the goal of the present study was to study the interaction between the pyriproxyfen with the LBD of the human RXRα using the ITC.

The study is in collaboration with Prof. Emilia Fisicaro and Prof. Angelo Bolchi teams of the University of Parma, which contributed to the ITC and the expression protein, respectively. However, since the analyses are still ongoing, the following discussion is very preliminary.

The expression vector of the human RXRα LBD was purchased and transfected into *Escherichia coli* cells. In the end, the protein was purified using the His-tag reaching a concentration of 120 μM. The buffer used to stock the protein contains 25mM Tris and 0.15M NaCl at pH=7.5. Since the pesticide pyriproxyfen was insoluble, stock ligand solutions were prepared in dimethyl sulfoxide (DMSO).

Though ITC can be carried out with polar ligands using organic solvents like DMSO to avoid ligand precipitation, many proteins will lose their native structures at the concentration needed to maintain ligand solubility. Since the main purpose of the present study was to evaluate the capability of this compound to interact with the protein RXRα, binding constant or thermodynamic calculations are not essential. For that reason, we tried to use the ligand diluted in the same protein buffer with 5% of DMSO during the ITC experiment. Thus, both the protein and the ligand were dissolved in 25mM Tris, 0.15M NaCl, and 5% DMSO at pH=7.5. When the ligand was tritated in the main cell, there was a high heat change, suggesting a strong interaction between the protein and ligand. However, the released heat did not decrease over the time. In that condition, it is difficult to understand if the heat change was due to the ligand interaction.

Different hypotheses have been suggested to explain this behavior:

- the protein acts as a surfactant, and the heat change was due to the solubilization of the pyriproxyfen during the titration;
- the ligand has a binding constant very high and binds to the protein very quickly. At the same time, it also has a very high dissociation constant. Thus, the protein saturation is never reached during the ITC experiment, making it impossible to observe a decrease in the heat change over time;
- a high or low ligand concentration was used.

From further perspectives, all the problems discussed above will be considered during future experiments.

10.2 Study 2 - Consensus scoring combined with machine learning to identify potential endocrine disruptors among food contact chemicals

Preface

During a collaboration with the Professor Martyn Smith team from the UC Berkeley Public Health Faculty and Jane Muncke and Ksenia J.Groh from the Food Packaging Forum (FPF) and Eawag (Swiss Federal Institute of Aquatic Science and Technology), two different computational methods (machine learning and robust consensus scoring) have been combined with the aim to screen all the intentionally used food contact chemicals against nine nuclear receptors (AR, ERα, ERβ, FXR, GR, PR, PPARγ, PPARδ, and RXR). This is a very innovative approach which combine two computational methods to prioritize the substances that have the characteristic to act as endocrine-disrupting compounds. First, the machine learning allowed to predict the activity of food contact chemicals against each nuclear receptor dividing them in active and inactive substances. In the second step, the active compounds have been screened using molecular docking and robust consensus scoring. This allowed to evaluate the capability of these molecules to interact with the specific nuclear receptor.

The results of the present study will be published in a manuscript that is still under preparation. For that reason, the following paragraphs are presented as scientific research paper.

Manuscript in preparation

10.2.1 Introduction

The nuclear receptors (NRs) family can be divided into seven subfamilies based on the molecules' structure and ligands. The first subfamily, the Subfamily 0 (NR0), is composed by two receptors characterized by only a ligand-binding domain. Subfamily 1 (NR1) is a large group composed of a large number of receptors that can bind a variety of different ligands. Subfamily 2 (NR2) is characterized by different receptors: the retinoid X receptors (RXRs), an important receptor because of the capability to form heterodimeric complexes with many other nuclear receptors, and the orphan receptors, that the putative ligand remains to be identified. Subfamily 3 (NR3) contains different receptors that are important in metabolic, reproductive, and development processes. The last three families, Subfamily 4 (NR4), 5 (NR5), and 6 (NR6), are composed of orphan nuclear receptors vital for the development and metabolism.

The retinoid X receptors (RXRs) family, composed by three members, RXRα (NR2B1), RXRβ (NR2B2), and RXRγ (NR2B3), are nuclear receptors members of Subfamily 2 important for the development, cell differentiation, and metabolism. RXRs can function as a dimer with either themselves or other nuclear receptors, such as farnesoid X receptor (FXR), peroxisome proliferator-activated receptors (PPARs), and liver X receptor (LXR), influencing their response [1]. Because of the implication of RXR heterodimers in the regulation of numerous nuclear signalling pathways, several ligands can cause harmful effects on human health (Egusquiza & Blumberg, 2020; Grün & Blumberg, 2009; Li *et al.*, 2016). From a structural point of view, the LBD structure of RXR is the same as the other members of the steroid/thyroid hormone NR superfamily. A particular characteristic of this receptor is the undefined and atypical conformation assumed by Helix12 (H12). In fact, when the receptor is bind to an agonist, the H12 position varied, but with the bind of CoA peptide, H12 adopts an agonist conformation [1]. 9-cis-retinoic acid (9-cis-RA) is proposed as the natural ligand of RXR but also some synthetic retinoids, such as SR11237 and SR11217, and some unsaturated fatty acids, such as linoleic acid and docosahexaenoic acids, are identified as RXR ligands (Heyman *et al.*, 1992; Jia-Hao *et al.*, 1995).

Farnesoid X receptor (FXR/NR1H4), the bile acid receptor, plays critical roles in several biological processes, such as bile acid regulation and glucose and lipid homeostasis (Wang *et al.*, 2018). The natural ligand of FXR is not yet identified, but some small

lipophilic and synthetic ligands have been determined as putative binders (Merk et al., 2019; Zheng et al., 2018).

Glucocorticoid receptor (GR/NR3C1) is involved in a wide variety of fundamental processes, such as metabolic homeostasis, inflammation, immune responses, development, and reproduction. GR is activated by cortisol, a natural glucocorticoid cholesterol-derived hormone secreted by the adrenal glands, but also by a synthetic glucocorticoids, drugs that resemble natural glucocorticoids [10]. Increasing evidence shows that the glucocorticoid receptor is potential targets for EDCs, in particular pesticides, causing several disorders, such as obesity cardiovascular and coronary diseases (Johansson *et al.*, 1998; Witorsch, 2016; Zhang et al., 2016; Zhang *et al.*, 2019).

PPARs belong to subfamily 1 of group C of nuclear receptors (NR1C). There are three different isoforms: PPARα (NR1C1), PPARβ/δ (NR1C2), and PPARγ (NR1C3). They regulate the genes transcription of lipids and glucose metabolism [15,16] and, thus, are involved in the beta-oxidation of lipids, fatty acids, and triglycerides, in anti-inflammatory processes and the glucose homeostasis [17,18]. From a physio-pathological point of view, PPARs are actively involved in the metabolic syndrome that comprises a series of risk factors such as hypertension, hypercholesterolemia, hypercholesterolemia, obesity, cardiovascular pathologies, and diabetes (Grundy *et al.*, 2004). Structurally, the PPARs LBD is common to all three isoforms and is close to the structure of other nuclear receptors. This region is composed of 13 helixes, H1-H12, and four β-sheets, S1-S4. However, differently from the other NRs, the LBD of PPARs contains an additional helix, called H2', between the first β-sheet and Helix 3. Moreover, H10 and H11 are unique helixes. The ligand-binding pocket is larger than those found in other nuclear receptors ranging from 1300 to 1400 A^3 [20]. The cavity has a peculiar Y-shape consisting of an entrance region with two pockets: Arm I (toward the helix 12) and Arm II located between H13 and β-sheet. The solvent-accessible part of the entrance and Arm I are mainly polar regions instead the interior region of the entrance and Arm II is substantially hydrophobic. Considering the nature of the ligand-binding pocket, ligands that bind PPARs, both natural and synthetic, are generally composed of a polar head, that occupies arm I, and of a hydrophobic tail that binds to arm II and the hydrophobic part of the entrance. Common ligands of PPARs are fatty acids, eicosanoids, and phospholipids.

The androgen receptor (AR/NR3C4) belongs to the steroid receptor subfamily of nuclear receptors, along with glucocorticoid receptor (GR/NR3C1), mineralocorticoid receptor

(MR/NR3C2), progesterone receptor (PR/NR3C3), and estrogen receptors α and β (ERα/NR3A1; ERβ/NR3A2). It plays a critical role in the normal development and homeostasis of male and female reproductive organs and their physiology [21]. Since AR has widely varied important roles, its dysregulation is involved in various diseases such as androgen insensitivity syndrome (AIS), prostate cancer (PCs), Kennedy's disease, and others [22]. Structurally, the ligand-binding domain of AR consists of eleven α-helices and four short β strands forming two anti-parallel β-sheets. Instead of having the helix H2, like the other nuclear receptors, the AR has a long flexible loop that connects H1 to H3 (Tan *et al.*, 2015). The internal ligand-binding pocket (LBP) is surrounded by helices H3, H5, and H11 where H12 acts as a closer of the LBP [23,24]. Endogenous ligands of AR include testosterone (TES) and its metabolite, 5α-DHT, but other synthetic ligands are well known, and they may be divided into steroidal and non-steroidal molecules. Synthetic ligands are normally used as antiandrogens to treat some AR-related diseases, such as PCs and AIS. Examples of drugs that bind AR are R-bicalutamide, nilutamide, and flutamide.

The estrogen receptor (ER) belongs to the steroid receptor transcriptional factor of nuclear receptors. ERα (NR3A1) and ERβ (NR3A2) are the two isoforms of the receptor. ERs are involved in the regulation of wide varied physiological processes. Thus, abnormal ER signaling leads to the development of different diseases, such as cancer, inflammation, osteoporosis, metabolic and cardiovascular disease, and neurodegeneration [25]. The LDB of the two isoforms shares a similar overall structure consisting of eleven helices arranged together in an antiparallel, three-layered sandwich topology. H12 has two main conformations. When an agonist ligand is bound to the receptor, the H12 is positioned over the LBP allowing the accommodation of the ligand. However, in the unliganded form and when an antagonist binds the LBD, the H12 flips outwards, preventing the coactivator recruitment and opening the entrance of the LBP [26,27]. Estradiol, estriol, and estrone are endogenous estrogens produced within the body that can recognize and bind the LBP of estrogen receptors. However, other estrogens are well-known binders of ERs, for example, phytoestrogens, a class of non-steroidal polyphenolic molecules produced by plants and the selective estrogen receptor modulators (SERMs) used as first-line drugs to treat breast cancer and other ER-related diseases comprising tamoxifen, raloxifene, clomiphene, etc. However, some xenoestrogens, coming from medicinal drugs, food additives, pesticides, cosmetics, and industrial chemicals, are also able to bind

ERs [27]. Thus, it is evident the important implication that the unintentional food contact chemicals can have on human health based on the ability of some of these compounds to bind the estrogen receptors and other NRs.

Miles of new chemicals are produced every year around the world, and a lot of these are possible endocrine disruptors compounds. Then, powerful tools are necessary in order to speed up the process of food safety and security. In this area, the use of computational approaches, such as virtual screening and molecular docking, are becoming essential to predict the interaction between a huge number of food contact chemicals with different nuclear receptors involved in human diseases and to decipher their mechanism of binding. Recently, machine learning methods have emerged as an important tool to the predict the effect of molecules on human health. For example, Hartung *et al.* have recently used a combination of a new type of QSAR and supervised ML approaches (RASAR) to predict nine health hazards based on data collected from dossiers of the European Chemical Agency (ECHA). They have addressed the interspecies variability of the *in vivo* animal assays, showing that guinea pigs and mice skin sensitisation assays were only 77% predictable of each other [28]. Instead, their computational method outperformed the reproducibility of interspecies assays, reaching an accuracy prediction of 92% [29].

10.2.2 Materials and Methods

10.2.2.1 Machine learning screening

The food contact chemicals list was obtained from recently published research (Groh *et al.*, 2021). SMILES were obtained from the CAS ID and salts and other small fragments of chemicals were removed from the downloaded dataset. The chemicals without SMILES were also removed from the food contact chemical dataset. The resultant dataset contained a total of 4847 chemicals. A recently developed predictive tool, NR-Toxpred (www.nr-toxpred.cchem.berkeley.edu) (Azhagiya Singam Ettayapuram Ramaprasad, Martyn T. Smith, David McCoy, Alan E. Hubbard, Michele A. La Merrill, Kathleen A. Durkin. Predicting the Binding of Small Molecules to Nuclear Receptors using Machine Learning (2021), in submission) was used to screen for the binding potential of the 4847 chemicals against nine different nuclear receptors, including AR, ERα, ERβ, FXR, GR, PR, PPARγ, PPARδ, and RXR. NR-Toxpred is a machine learning-based model developed with the NuRA dataset (Valsecchi *et al.*, 2020), which is the most exhaustive collection of small molecules annotated for their modulation of the nine different nuclear

receptors. NR-Toxpred uses machine learning to classify the given chemical as a binder or non-binder for a chosen receptor, and for seven of the receptors, it predicts whether the active binder is an agonist or antagonist (Azhagiya Singam Ettayapuram Ramaprasad, Martyn T. Smith, David McCoy, Alan E. Hubbard, Michele A. La Merrill, Kathleen A. Durkin. Predicting the Binding of Small Molecules to Nuclear Receptors using Machine Learning (2021), in submission).

10.2.2.2 Nuclear receptors structures preparation

The structures of the nine nuclear receptors of *H. sapiens* were downloaded from the Protein Data Bank (PDB) (https://www.rcsb.org/). The crystallographic structures were processed using Sybyl software v8.1 (www.tripos.com): water molecules and ligands were removed, hydrogen atoms were added, and energy minimized using the Powell algorithm with a coverage gradient of ≤ 0.5 kcal (mol Å)$^{-1}$ and a maximum of 1500 cycles. For the docking with AutoDock (*see below*), the receptors were processed using AutoDockTools software: polar hydrogen are added to the proteins and the Gasteiger charges were calculated for each atom to assign AD4 type to the atoms.

10.2.2.3 Preparation of ligands

The FCCs molecules resulted from the machine learning procedure and the structural coordinates of the endogenous and putative ligands, retrieved from the NCBI PubChem compound database (https://pubchem.ncbi.nlm.nih.gov/), were processed with FLAP (Fingerprint for Ligand and Protein) in order to assign the correct protonation state to them (pH=7.4).

10.2.2.4 Molecular docking procedure

The molecular docking procedure was carried out using two software, GOLD and AutoDock, to have the possibility of comparing multiple scoring functions, which have different protein-ligand interaction assessment methods.

10.2.2.5 Molecular docking with GOLD

The GOLD software v5.8.1 (CCDC; Cambridge, UK; http://www.ccd.cam.ac.uk) was applied to dock ligands into the binding site of the nine nuclear receptors. The centroid of the binding site of each receptor was defined using the coordinates of the crystallographic complexes and the side chain flexibility was allowed for each receptor amino acid (Table 1). For the genetic algorithm run, a maximum number of 100000

operations were performed on a population of 100 individuals with a selection pressure of 1.1. The number of islands and the niche size were set to 5 and 2, respectively. For each compound and receptor, 30 binding poses were generated. The default GOLD Score fitness function was applied for performing the energetic evaluations. The distance for hydrogen bonding and the cut-off value for the van der Waals calculation were set to 2.5 Å and 4.0 Å, respectively. Flip pyramidal N, flip amide bonds, and flip ring corners were allowed for ligand flexibility options. All the poses generated by GOLD were rescored using two scoring functions, Chem Score and Hint Score (HINT, Hydropathic INTeraction). The combination of these three scoring functions was chosen as follow: i) Gold Score allows taking into account different factors such as H-bonding energy, van der Waals energy, metal interaction, and ligand torsion strain; ii) Chem Score represents the total free energy charge and takes account of hydrophobic contact area, hydrogen bonding, ligand flexibility, and metal interaction; iii) Hint Score provides a quantitative evaluation of protein-ligand interaction (from logP experimental data) and takes into account both the enthalpic and entropic contributions to the ΔG of ligand-protein interaction (Eldridge *et al.*, 1997; Eugene Kellogg & Abraham, 2000).

Table 1. The nine nuclear receptors with the centroid of each binding site and side chain flexibility of the amino acids.

NUCLEAR RECEPTORS	PDB ID	SIDE CHAIN FLEXIBILITY OF THE AMINO ACIDS
RXRα	1FM6	Phe436, His435, Phe439, Leu436
RXRβ	1UHL	/
AR	2AM9	Leu701, Asn705, Gln711, Trp741, Met745, Met745, Met780, Thr877, Met895, Ile899
ERα	2YJA	Phe404, Met421, Ile424, Phe425, His524, Leu525
ERβ	2YJD	His475, Met340, Phe377
FXR	4QE6	Met265, Met290, His294, Phe336, Phe350, Tyr369, Met450, Trp454, Trp469
PPARδ	5U3Z	Ile327, His430, Trp228, Leu294, His287
GR	1M2Z	Ile747, Tyr735, Met646, Met604, Met560, Gln642, Arg611
PPARγ	3NOA	His449, Ser289, Phe282, Met364, Arg288

10.2.2.6 Molecular docking using AutoDock

Molecular docking experiments with AutoDock Vina 1.1.2 were performed using default setting [34]. The search space was included in a box of 24 × 24 × 24 Å, centred on the

binding site of the ligands as mentioned before. The side chain flexibility was allowed for the same residues defined in the GOLD docking. The ligand amide and backbone flexibility were allowed.

10.2.3 Results and Discussion

Among the NRs considered, ERα was the receptor with the highest number of active compounds, followed by AR, ERβ, PPARγ, FXR, PPARδ, GR, and RXR (**Table 2**).

Table 2. The number of food contact chemicals predicted as active and inactive from the machine learning approach for each nuclear receptor considered.

	Machine Learning	
	Active	Inactive
AR	216	4632
ERα	335	4513
ERβ	150	4698
FXR	118	4730
GR	50	4798
RXR	16	4832
PPARγ	176	4672
PPARδ	80	4768

The capability of the food contact chemicals (predicted as active by machine learning) to bind the corresponding nuclear receptor has been evaluated using a molecular docking approach. Specifically, two docking software and four scoring functions have been applied to reach a statistical consensus prediction. After that, based on the consensus scoring value (CS), the compounds have been classified as high (CS ≥ 0.8), medium (0.3 ≤ CS > 0.8), and low (CS < 0.3) interactors. As we can see in **Figure 1**, among the ML's active compounds, the FXR has the highest number of compounds in the high interactor class (76%). On the other side, the ERβ has more compounds predicted as low interactors (34%).

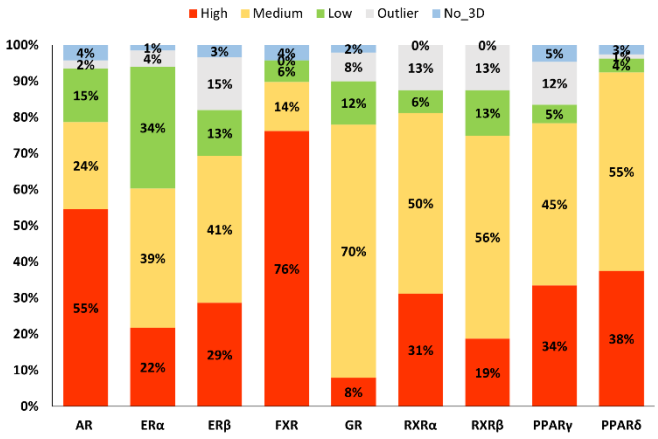


Figure 1. The percentage of the machine learning active molecules predicted as low, medium, and low interactor for each nuclear receptor. The figure also shows the number of molecules identified as outlier by the robust consensus scoring prediction and the percentage of molecules without a three-dimensional structure.

Considering the molecules that have been predicted as high interactors, after the machine learning method and the consensus molecular docking prediction, we have identified 118, 73, 43, 90, 4, 5, 3, 59, and 30 substances as more harmful compounds for AR, ERα, ERβ, FXR, GR, RXRα, RXRβ, PPARγ, and PPARδ, respectively (**Table 3**).

Table 3. The number of molecules identified as high interactor after the application of the machine learning and the consensus scoring prediction.

	High	Medium	Low	Outlier	No 3D
AR	118	52	32	5	9
ERα	73	129	113	15	5
ERβ	43	61	19	22	5
FXR	90	16	7	0	5
GR	4	35	6	4	1
RXRα	5	8	1	2	0
RXRβ	3	9	2	2	0
PPARγ	59	79	9	21	8
PPARδ	30	44	3	1	2

Since ML method has been trained using *in vitro/in vivo* (Martyn & Co) test (…). Thus, it is not an indicator of the direct binding between the NR and food contact chemicals. For this reason, it is not surprising that some of machine learning's active compounds have been judged as low interactors from the molecular docking analyses. In fact, the combination of ML and molecular docking approaches allows, in the first instance, to

screen a huge number of compounds in a time- and cost-effective way. Secondly, this approach allows also to identify the substances of very high concerns (SVHCs).
From a broader point of view, the method could also be used to understand the kind of *in vitro* tests that should be set up. In fact, if a compound has been predicted as positive by the machine learning method for the estrogenic pathway, but it has been judged as a low interactor of both ERα and ERβ, a test that considers the whole pathway should be advisable instead of using a methodology that considers the direct binding between the receptor and FCC (such as a colorimetric test).

References

[1] M.I. Dawson, Z. Xia, The retinoid X receptors and their ligands, Biochim. Biophys. Acta - Mol. Cell Biol. Lipids. 1821 (2012) 21–56. https://doi.org/10.1016/j.bbalip.2011.09.014.

[2] R.J. Egusquiza, B. Blumberg, Environmental obesogens and their impact on susceptibility to obesity: New mechanisms and chemicals, Endocrinol. (United States). 161 (2020) 1–14. https://doi.org/10.1210/endocr/bqaa024.

[3] F. Grün, B. Blumberg, Endocrine disrupters as obesogens, Mol. Cell. Endocrinol. 304 (2009) 19–29. https://doi.org/10.1016/j.mce.2009.02.018.

[4] N. Li, W. Jiang, M. Ma, D. Wang, Z. Wang, Chlorination by-products of bisphenol A enhanced retinoid X receptor disrupting effects, J. Hazard. Mater. 320 (2016) 289–295. https://doi.org/10.1016/j.jhazmat.2016.08.033.

[5] R.A. Heyman, D.J. Mangelsdorf, J.A. Dyck, R.B. Stein, G. Eichele, R.M. Evans, C. Thaller, 9-cis retinoic acid is a high affinity ligand for the retinoid X receptor, Cell. 68 (1992) 397–406. https://doi.org/10.1016/0092-8674(92)90479-V.

[6] X. Jia-Hao, D. Béatrice, P. Chambon, J. J.Voorhees, Endogenous Retinoic Acid Receptor (RAR)-Retinoid X Receptor (RXR) Heterodimers Are the Major Functional Forms Regulating Retinoid-responsive Elements in Adult Human Keratinocytes: BINDING OF LIGANDS TO RAR ONLY IS SUFFICIENT FOR RAR•RXR HETERODIMERS TO CO, J. Biol. Chem. 270 (1995) 3001–3011.

[7] N. Wang, Q. Zou, J. Xu, J. Zhang, J. Liu, Ligand binding and heterodimerization with retinoid X receptor (RXR) induce farnesoid X receptor (FXR) conformational changes affecting coactivator binding, J. Biol. Chem. 293 (2018) 18180–18191. https://doi.org/10.1074/jbc.RA118.004652.

[8] D. Merk, S. Sreeramulu, D. Kudlinzki, K. Saxena, V. Linhard, S.L. Gande, F. Hiller, C. Lamers, E. Nilsson, A. Aagaard, L. Wissler, N. Dekker, K. Bamberg, M. Schubert-Zsilavecz, H. Schwalbe, Molecular tuning of farnesoid X receptor partial agonism, Nat. Commun. 10 (2019) 1–14. https://doi.org/10.1038/s41467-019-10853-2.

[9] W. Zheng, Y. Lu, S. Tian, F. Ma, Y. Wei, S. Xu, Y. Li, Structural insights into the heterodimeric complex of the nuclear receptors FXR and RXR, J. Biol. Chem. 293 (2018) 12535–12541. https://doi.org/10.1074/jbc.RA118.004188.

[10] M. Kadmiel, J.A. Cidlowski, Glucocorticoid receptor signaling in health and

disease, Trends Pharmacol. Sci. 34 (2013) 518–530. https://doi.org/10.1016/j.tips.2013.07.003.

[11] M. Johansson, S. Nilsson, B.O. Lund, Interactions between methylsulfonyl PCBs and the glucocorticoid receptor, Environ. Health Perspect. 106 (1998) 769–772. https://doi.org/10.1289/ehp.98106769.

[12] R.J. Witorsch, Effects of elevated glucocorticoids on reproduction and development: Relevance to endocrine disruptor screening, Crit. Rev. Toxicol. 46 (2016) 420–436. https://doi.org/10.3109/10408444.2016.1140718.

[13] J. Zhang, J. Zhang, R. Liu, J. Gan, J. Liu, W. Liu, Endocrine-Disrupting Effects of Pesticides through Interference with Human Glucocorticoid Receptor, Environ. Sci. Technol. 50 (2016) 435–443. https://doi.org/10.1021/acs.est.5b03731.

[14] J. Zhang, Y. Yang, W. Liu, D. Schlenk, J. Liu, Glucocorticoid and mineralocorticoid receptors and corticosteroid homeostasis are potential targets for endocrine-disrupting chemicals, Environ. Int. 133 (2019) 105133. https://doi.org/10.1016/j.envint.2019.105133.

[15] V. Dubois, J. Eeckhoute, P. Lefebvre, B. Staels, Distinct but complementary contributions of PPAR isotypes to energy homeostasis., J. Clin. Invest. 127 (2017) 1202–1214. https://doi.org/10.1172/JCI88894.

[16] K. Schoonjans, B. Staels, J. Auwerx, The peroxisome proliferator activated receptors (PPARS) and their effects on lipid metabolism and adipocyte differentiation., Biochim. Biophys. Acta. 1302 (1996) 93–109. https://doi.org/10.1016/0005-2760(96)00066-5.

[17] M. Vázquez-Carrera, Unraveling the Effects of PPARβ/δ on Insulin Resistance and Cardiovascular Disease., Trends Endocrinol. Metab. 27 (2016) 319–334. https://doi.org/10.1016/j.tem.2016.02.008.

[18] E. Barroso, R. Rodríguez-Calvo, L. Serrano-Marco, A.M. Astudillo, J. Balsinde, X. Palomer, M. Vázquez-Carrera, The PPARβ/δ activator GW501516 prevents the down-regulation of AMPK caused by a high-fat diet in liver and amplifies the PGC-1α-Lipin 1-PPARα pathway leading to increased fatty acid oxidation., Endocrinology. 152 (2011) 1848–1859. https://doi.org/10.1210/en.2010-1468.

[19] S.M. Grundy, B. Hansen, S.C.J. Smith, J.I. Cleeman, R.A. Kahn, Clinical management of metabolic syndrome: report of the American Heart Association/National Heart, Lung, and Blood Institute/American Diabetes

Association conference on scientific issues related to management., Arterioscler. Thromb. Vasc. Biol. 24 (2004) e19-24. https://doi.org/10.1161/01.ATV.0000112379.88385.67.

[20] V. Zoete, A. Grosdidier, O. Michielin, Peroxisome proliferator-activated receptor structures: ligand specificity, molecular switch and interactions with regulators., Biochim. Biophys. Acta. 1771 (2007) 915–925. https://doi.org/10.1016/j.bbalip.2007.01.007.

[21] R.A. Davey, M. Grossmann, Androgen Receptor Structure, Function and Biology: From Bench to Bedside, Clin. Biochem. Rev. 37 (2016) 3–15. https://pubmed.ncbi.nlm.nih.gov/27057074.

[22] G.C. Shukla, A.R. Plaga, E. Shankar, S. Gupta, Androgen receptor-related diseases: what do we know?, Andrology. 4 (2016) 366–381. https://doi.org/https://doi.org/10.1111/andr.12167.

[23] M.H.E. Tan, J. Li, H.E. Xu, K. Melcher, E. Yong, Androgen receptor: structure, role in prostate cancer and drug discovery, Acta Pharmacol. Sin. 36 (2015) 3–23. https://doi.org/10.1038/aps.2014.18.

[24] M. Nadal, S. Prekovic, N. Gallastegui, C. Helsen, M. Abella, K. Zielinska, M. Gay, M. Vilaseca, M. Taulès, A.B. Houtsmuller, M.E. van Royen, F. Claessens, P. Fuentes-Prior, E. Estébanez-Perpiñá, Structure of the homodimeric androgen receptor ligand-binding domain, Nat. Commun. 8 (2017) 14388. https://doi.org/10.1038/ncomms14388.

[25] M. Jia, K. Dahlman-Wright, J.-Å. Gustafsson, Estrogen receptor alpha and beta in health and disease, Best Pract. Res. Clin. Endocrinol. Metab. 29 (2015) 557–568. https://doi.org/https://doi.org/10.1016/j.beem.2015.04.008.

[26] S. Lee, M.G. Barron, Structure-Based Understanding of Binding Affinity and Mode of Estrogen Receptor α Agonists and Antagonists, PLoS One. 12 (2017) e0169607–e0169607. https://doi.org/10.1371/journal.pone.0169607.

[27] A. Farooq, Structural and Functional Diversity of Estrogen Receptor Ligands, Curr. Top. Med. Chem. 15 (2015) 1372–1384. https://doi.org/10.2174/1568026615666150413154841.

[28] T. Luechtefeld, A. Maertens, D.P. Russo, C. Rovida, H. Zhu, T. Hartung, Analysis of public oral toxicity data from REACH registrations 2008-2014, ALTEX. 33 (2016) 111–122. https://doi.org/10.14573/altex.1510054.

[29] T. Hartung, Predicting toxicity of chemicals: software beats animal testing, EFSA J. 17 (2019) e170710. https://doi.org/10.2903/j.efsa.2019.e170710.

[30] K.J. Groh, B. Geueke, O. Martin, M. Maffini, J. Muncke, Overview of intentionally used food contact chemicals and their hazards, Environ. Int. 150 (2021) 106225. https://doi.org/https://doi.org/10.1016/j.envint.2020.106225.

[31] C. Valsecchi, F. Grisoni, S. Motta, L. Bonati, D. Ballabio, NURA: A curated dataset of nuclear receptor modulators, Toxicol. Appl. Pharmacol. 407 (2020) 115244. https://doi.org/10.1016/j.taap.2020.115244.

[32] M.D. Eldridge, C.W. Murray, T.R. Auton, G. V. Paolini, R.P. Mee, Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes, J. Comput. Aided. Mol. Des. 11 (1997) 425–445. https://doi.org/10.1023/A:1007996124545.

[33] G. Eugene Kellogg, D.J. Abraham, Hydrophobicity: Is LogP(o/w) more than the sum of its parts?, Eur. J. Med. Chem. 35 (2000) 651–661. https://doi.org/10.1016/S0223-5234(00)00167-7.

[34] O. Trott, A.J. Olson, Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading, J. Comput. Chem. 31 (2009) 455–461.

Conclusions

Food contact chemicals are widespread compounds that come into contact with our food as contaminants or as molecules intentionally added to food. Careful monitoring of all these molecules is necessary to guarantee food safety and food quality. Worldwide and European legislations are specifically set out to guarantee the safety and the quality of food products, from the raw material to the final product. Three different aspects might be taken into consideration to ensure food safety and food quality: (i) evaluate the potential risk that a molecule can cause harm for humans and the environment (hazard identification); (ii) evaluate the risk of these molecules considering the level at which they are found in the food chain (risk assessment); (iii) set out the maximum levels that might be considered safe, and restrict or abolish the use of some of these molecules (risk management). Globalization and the increasing demand for food quality products have forced industries to find new methods to protect food from external contaminations, enhance food properties, and avoid food spoiling and oxidation. Thus, this exigence has forested the use of new molecules in the entire food chain (pesticides, molecules of food packaging, preservatives, etc.). Due to the increasing number of new food contact chemicals, the standard protocols are often inefficient to evaluate and monitor food contact chemicals, at least in a limited time and at low costs, and a lot of molecules must be still tested.

In this context, this doctoral dissertation aims to highlight the application of computational methods for ensuring food safety. Specifically, two major aspects have been considered. First, the application of database and Big Data technologies has been addressed with the aim to collect food contact chemical information considering the issue from a computational and regulative point of view. Second, this PhD thesis shows the development of a robust *in silico* workflow to evaluate the potential endocrine-disrupting properties played by all food contact chemicals with a focus on the androgenic and estrogenic pathways.

The first chapters aim to illustrate the problem of endocrine-disrupting compounds (EDCs) and to evaluate the applicability of known computational techniques for testing

food contact chemicals. In this context, the molecular initiating event (MIE) plays an important role to understand the endpoint to be considered for EDCs evaluation. In fact, to determine an effect, a molecule often interacts with some specific macromolecules (proteins, nucleic acid, carbohydrates, etc.) triggering a cascade of events that lead to the harmful effect. Endocrine-disrupting effects are often determined by the interaction of an EDC with nuclear receptors, a transcriptional factor family. The dysregulation of their functions is involved in a series of diseases, like prostate and breast cancers (Chapter 1). From a computational point of view, the interaction between a food contact chemical and a nuclear receptor is a "classical" ligand-protein problem, that can be easily assessed using molecular docking and molecular dynamic techniques which are well-established methods derived from medicinal chemistry, and that can be directly applied in this context. However, the lack of high curated databases from which to retrieve food contact chemicals information and their three-dimensional structures often slow down the application of *in silico* methods for screening purposes. Thus, Chapter 2 of this dissertation lays an overview of the database system with the aim to illustrate the benefit of using such a system in the food context. Once the foundations of the database system have been provided, in Chapter 3, the use of computational methods (databases, big data, and molecular docking) is discussed focusing on their application for ensuring food safety. Although their use is a relatively recent challenge in food science, some case studies are provided to highlight how these techniques can be applied to a wide range of food safety problems. *In silico* tools are not just useful for speeding up the process of analysing food contact chemicals' safety, but they might be seen as alternative methods to animal tests in the first phase of the hazard identification step. In Chapter 4, alternative animal tests are discussed focusing on both *in vitro* bioassays and computational methods with a more emphases on the endocrine-disrupting problem. Thus, after giving a general dissertation about nuclear receptors and the associated diseases in which they are involved, the first part of the chapter discusses *in vitro* bioassays for evaluating food contaminants. However, the terrific amount of food contact chemicals highlights the need for methods that can predict endocrine-disrupting compounds in a fast way. Although *in vitro* studies have currently improved their performances, and several high-throughput testes exist, *in silico* methods are still faster and more affordable in terms of costs, and they can be considered as a useful and alternative support to experimental analyses. For this reason, in the second part of Chapter 4, the importance of ligand-based virtual

screening, molecular docking, consensus scoring, and molecular dynamics simulation is discussed. In the end, it has been also provided some real cases studies.

Once the general information about food contact chemicals, nuclear receptors, and computational methods has been provided, the following chapters illustrate the application of these concepts in food science.

In Chapter 5 it has been evaluated the applicability of computational methods to screen the endocrine-disrupting properties of food contact chemicals using bisphenols as a case study. Bisphenols are molecules mainly used to manufacture plastics for made food beverages, food containers, infant feeds, etc. Since 2017, one of the most used bisphenols, Bisphenol A (BPA), has been considered as a substance of very high concern by the European Chemical Agency (ECHA) for its toxic effects and its endocrine-disrupting properties. Thus, a lot of efforts are aimed at identifying alternative molecules with a lower toxicological profile to completely substitute or reduce the most harmful compounds. As a final result, the work identifies Bisphenol S as a safer compound. Thus, further *in vitro* analyses can correctly estimate the safety of this compound with the aim to substitute BPA. Generally, the results show that computational methods are in accordance with some previous experimental results, showing their applicability in hazard identification as preliminary techniques to screen large datasets of compounds.

In Chapter 6, the problem of food contact chemicals has been addressed from a different point of view that is in line with the "3R" concept. The 3R principle was introduced by Russel and Burch in 1959s to promote laboratory animal protection. The principle is based on the replacement, the reduction, and the refinement (reduce the animals suffering) of tests that involve animals. The work illustrates the application of *in silico* methods to screen the endocrine-disrupting properties of pesticides against the retinoid X receptor (RXR) of *Daphnia Magna*, a microcrustacean used as an experimental model, and that plays an important role in the aquatic environment. Using molecular docking and consensus scoring, the pyriproxyfen and its metabolite, the 4'-OH-pyriproxyfen have been identified as two molecules that interact tightly with the receptor. Their mechanism of action has been also evaluated using molecular dynamic simulations which have highlighted that the metabolite 4'-OH-pyriproxyfen is quite stable during the simulation time, suggesting its role in determining an endocrine-disrupting effect. Moreover, since

the RXR shares great sequence identity with the corresponding receptor of *Homo Sapiens*, it is very probable that it can also be harmful for human health.

Within the well-known context of food contact chemicals, information is widespread across the internet and scientific literature. This makes information retrieval very challenging. To this end, a food contact chemical database (foodchem DB) has been designed to store different types of information. The problem has been addressed from different points of view: (*i*) retrieve all food contact chemicals from well-established authority; (*ii*) provide a tool to accelerate computational method; (*iii*) consider the available legislative information. To this end, the database stores different types of information, such as identification names, chemical-structures properties, three-dimensional structures, and a link to European authorities. The increasing number of food contact chemicals, along with the heterogeneous data the database stores, points out the utility of Big Data approaches to collect all this information. In light of this, two different databases have been designed, written in SQL and NoSQL (big data) languages. In Chapter 7, it is shown the application of the "foodchem DB" to evaluate the endocrine-disrupting properties of all food contact chemicals against 31 nuclear receptors of *Homo Sapiens* using computational methods. The combination of molecular docking and robust consensus scoring techniques has highlighted that a lot of molecules can interact with more than 50% of the nuclear receptors, suggesting that they can determine a wide endocrine disruptor effect in humans.

A major attention on estrogenic and androgenic pathways has been covered in Chapter 8 since these two receptors are well-known targets involved in prostate and breast cancers. The results point out that the androgen receptor is the most affected nuclear receptor, since almost half of food contact chemicals are able to act as high interactors.

Molecular docking and consensus scoring predictions are undoubtedly important tool to evaluate the ligand-protein interaction. However, they do not give any information about the endocrine effect on humans, i.e., if a molecule determines an agonistic or antagonistic effect. Different strategies can be used to address this issue using computational methods. For example, it is well-known that estrogen receptor has two different conformations when it is bound to an agonist (close) or antagonist (open) compound. By using the close and open structures, *in silico* methods may be also give an idea of the EDCs effect. The same cannot be said for the androgen receptor, for which a close/open conformation is

not yet known. To shed light about the antagonistic effect, in Chapter 9, the problem has been discussed using a molecular dynamic approach. The results show that antagonistic compounds may affect the homodimer stability of the receptor instead of inducing a close/open conformation.

In Chapter 10, two preliminary studies have been reported showing future perspectives of the present PhD thesis. In study 1, it is illustrated the utility of Isothermal Titration Calorimetry (ITC) to study protein-ligand binding. This technique can provide experimental evidence of *in silico* protein-ligand interaction prediction. In study 2, a very recent computational workflow has been illustrated. Two different computational methods (machine learning and robust consensus scoring) joined forces to identify the substances of very high concern (SVHCs) that should be prioritized during experimental methods.

Overall, the results suggested that the combination of different computational methods can help to disclose the endocrine-disrupting effect and to screen large a dataset of molecules. This work shows the important role of *in silico* techniques in hazard identification since it is a cost- and time-effective choice during the first step of the risk evaluation. Moreover, it allows reducing animal tests according to the 3R' principle.

About the author

Francesca Cavaliere was born on November 15st, 1989 in Taranto (Italy) and she attended her academic studies at the University of Parma obtaining a Bachelor's Degree in Biology and a Master's Degree in Molecular Biology. After that, she spent two years as a Postgraduate Researcher at the Molecular Modelling Laboratory of Food Science, where she has performed computational methods for pharmacogenetics analyses and for studying protein-ligand interactions of some food contact chemicals. Three years ago, in 2018, she started the PhD in Food Science at the Department of Food and Drug of the University of Parma, under the supervision of Prof. Pietro Cozzini. Her doctoral research aimed to perform risk assessment analyses using computational methods to identify probable endocrine-disrupting compounds among food contact chemicals. Her PhD has been funded by Regione Emilia Romagna in the field of Big Data with the purpose to apply big data technologies in food science. To reach this goal, she has designed two different databases storing chemical-physical properties and links to European authorities of all food contact chemicals ensuring a high level of data quality.
She is the author of scientific articles in ranked journals and chapters book.

List of publications

P. Cozzini, **F. Cavaliere**, G. Spaggiari, G. Morelli, M. Riani, Computational Methods on Food Contact Chemicals: Big Data and In Silico Screening on Nuclear Receptors Family. Chemosphere. 292 (2022) 133422. https://doi.org/https://doi.org/10.1016/j.chemosphere.2021.133422.

Cavaliere, F. et al. (2021). "CHAPTER 12 - Computational Applications on Food Contact Chemicals as Nuclear Receptor Binders". Nuclear Receptors. Springer

Spaggiari G, **Cavaliere F**, Cozzini P. The Application of In Silico Methods on Umami Taste Receptor. Handb Exp Pharmacol. 2021 Jul 11. doi: 10.1007/164_2021_515

Beghi S, **Cavaliere, F.**, Manfredini, M., Ferrarese, S., Corazzari, C., Beghi, C., Buschini, A. - "Polymorphism rs7214723 in CAMKK1: a new genetic variant associated with cardiovascular diseases" - Biosci Rep. 2021 Jul 30; 41(7): BSR20210326.

Cavaliere, F., Spaggiari, G., Cozzini, P. (2021). "CHAPTER 23 – Molecular Docking: A contemporary story about food safety". Molecular Docking for Computer-Aided Drug Design. Elsevier

Vehagen, H., C. Alonso-Andicoberry, R. Assunção, **F. Cavaliere**, H. Eneroth, J. Hoekstra, S. Koulouris, A. Kouroumalis, S. Lorenzetti, A. Mantovani, D. Menozzi, M. Nauta, M. Poulsen, J. Rubert, A. Siani, V. Sirot, G. Spaggiari, S.T. Thomsen, M. Trevisan, P. Cozzini – Risk-benefit in food safety and nutrition – Outcome of the 2019 Parma Summer School – Food Research International 141 (2021) 110073

Beghi, S., **Cavaliere, F.**, Buschini, A. - Gene polymorphisms in calciumcalmodulin pathway: focus on cardiovascular disease – Mutation Research 786 (2020)

Cavaliere, F., Lorenzetti, S., Cozzini, P. - Molecular Modelling methods in food safety: bisphenols as case study – Food Chem Toxicol. 2020;137:111116

Cesari A., Uccello Barretta, G., Kirschner, K.N., Pappalardo, M., Basile, L., Guccione, S., Russotto, C., Lauro, M.R., **Cavaliere, F.**, Balzano, F. - Interaction of natural flavonoid eriocitrin with β-cyclodextrin and hydroxypropyl-β-cyclodextrin: an NMR and molecular dynamics investigation - New J. Chem., 2020, 44, 16431

Cavaliere, F. & Cozzini, P. – New in Silico Trends in Food Toxicology - Chem. Res. Toxicol. 2018, 31, 10, 992-993

Cavaliere, F., Montanari, E., Emerson, A., Buschini, A., Cozzini, P.- In silico pharmacogenetic approach: The natalizumab case study. - Toxicology and Applied Pharmacology, 330 (2017) 93–99

Submitted publications

F. Cavaliere & P. Cozzini, Understanding the mechanism of action of R-bicalutamide on the W741L androgen receptor through molecular dynamic, Toxicology and Applied Pharmacology, 2021

Fabio Fornari, Fabio Montisci, Federica Bianchi, Marina Cocchi, Claudia Carraro, **Francesca Cavaliere**, Pietro Cozzini, Francesca Peccati, Paolo Pio Mazzeo, Nicolò Riboni, Maria Careri, Alessia Bacchi, Chemometric-Assisted Cocrystallization: Supervised Pattern Recognition for Predicting the Formation of New Functional Cocrystals, Angewandte Chemie, 2021

Awards and distinctions

Winner of "Premio Nazionale Ricerca Big Data e AI 2021"
International Foundation Big Data and Artificial Intelligence for Human Development (IFAB), Rimini, Italy, July, 15, 2021

Courses, conferences and workshops

- Parma Summer School 2021 – "*Food Safety Aspects of Integrated Food Systems*", Online, 28-30 September 2021
- *First Virtual Workshop on the Developments in the Italian PhD Research on Food Science Technology and Biotechnology*, Palermo, 14-15 September 2021
- *English Language Course for PhD Students: Study skills - English for Academic Purposes*, From 07 April to 23 June, 2021.
- *High Performance Molecular Dynamics Course*, online, 07-09 April 2021
- *Data Science: Machine Learning and its Application in Genomics, Chemistry and Neuroscience*, online, 07-17 September 2020
- *Workshop: Application of differential scanning calorimetry (DSC) in food research*, online, 02 July 2020
- *Scientific writing*, online, 24-25 June 2020
- *Elastic Search Course*, online,11-19 June 2020

- *Course on Fondamenti di programmazione A*, Parma, From 10 October 2018 to 18 January 2019

Other activities

- Collaboration with Prof. Thomas Hartung from John Hopkins Schools for food contact chemicals assessment using a machine learning approach
- Collaboration with Prof. Alessia Bacchi from the University of Parma for predicting the formation of new functional cocrystals
- Teacher "Corso teorico-pratico sull'applicazione dei metodi computazionali nel Replacement", IPAM, 09 July 2021
- Teaching support for the course of "Modellistica molecolare"
- Exam support to the course of general chemistry