UNIVERSITÀ DI PARMA

UNIVERSITA' DEGLI STUDI DI PARMA

DOTTORATO DI RICERCA IN

SCIENZE DEGLI ALIMENTI

CICLO XXXIV

In silico methods for the detection of endocrine disruptors in food

Coordinatore:

Chiar.ma Prof.ssa Chiara Dall'Asta

Tutore:

Chiar.mo Prof. Pietro Cozzini

Dottorando:

Giulia Spaggiari

Anni Accademici 2018/2019 – 2020/2021

Abstract

During the twentieth century, the world experienced an enormous technological and industrial development which, however, had several negative effects, such as an increase in risks to human health due to the products and/or waste deriving from the industry. In particular, among them, we find a class of substances called "Endocrine Disrupting Chemicals" (EDCs), a heterogeneous group of chemicals characterized by the ability to interfere with the functioning of the endocrine system through several mechanisms. The main source of exposure to EDCs for humans is represented by food, but it is not the only one: in fact, it is possible to come into contact with these molecules also through the environment, such as air, water, and soil, or the use of different products, such as detergents, cosmetics, clothes, and toys. The spectrum of pathologies related to these compounds is very broad and includes tumors, birth effects, metabolic disorders, reproductive function problems in males and females, and many others. However, the consequences on human and animal health and the effects on the environment of these chemicals are not yet fully verified. Many points, including the biological mechanisms, the mechanisms of action, the risk factors, and the entire spectrum of pathologies potentially associated with exposure to EDCs, still need to be clarified.

The vastness of the problem requires the collaboration of experts, scientists, governments, and international agencies. A rationalization of efforts is also necessary, to fill those gaps in current knowledge that are of critical importance. It is necessary to obtain solid scientific knowledge regarding: i) the levels of environmental pollution; ii) the exposure extent of the population and in particular of certain risk groups; iii) the relationship between the absorbed dose and the occurrence of negative effects; iv) the mechanism of action of these chemicals; and v) the development of *in vitro* and *in vivo* experimental tests capable of both identifying with sufficient sensitivity and characterizing accurately the effects on endocrine balance. In this context, computational methods can be used to study the mechanisms and modes of action underlying the toxicity of endocrine disruptors chemicals. They are based on the premise that the chemical and physical properties and the bioavailability and toxicity of a chemical depend on its intrinsic nature (structure-activity relationship) and they can be directly predicted from its molecular structure and/or from similar structures with known functions and effects.

In this broad and complex context, this PhD thesis wants to highlight the criticality of the endocrine disruptors problem and the relative negative effects on human health, and the usefulness of the computational methods for detecting endocrine disruptors in food, for

understanding their mechanism of action, and for preventing their possible negative effects on human, animal, plant one health. In more detail, the aim is to detect the possible endocrine disruptors in food using *in silico* methods.

CHAPTER 1

General Introduction

Since the mid-twentieth century, the rapid and often uncontrolled development of industrial technologies has progressively caused an increase in the level and extent of risks for human health (Colborn et al., 1996). In particular, scientific and public concerns grow about a series of substances, called endocrine disruptors, capable of altering the endocrine system function with possible negative effects on human and animal health. These chemicals have a high environmental diffusion with effects that are still not fully known today. The possibility that some of these molecules interact negatively with the human and the animal endocrine system has received, especially in the last decade, considerable attention not only from the scientific community but also from public opinion. In fact, the related problems of endocrine disruptors are on the agendas of many groups of experts, commissions, international organizations, industries, and universities all over the world.

Endocrine system

Organs and various parts of our bodies must communicate with each other to ensure the maintenance of homeostasis, which allows them to function properly. Two systems help ensure this communication: the nervous system and the hormonal (neuroendocrine) system. The latter relies on the production and release of hormones from various glands (hypothalamus, pituitary, thyroid, adrenals, reproductive, and many others) and on their transportation via the bloodstream. Hormones are molecules that are produced in response to specific stimuli. When a hormone is released into the bloodstream, it interacts with certain docking molecules, called receptors, located either on the surface or inside of specific target cells (Alberts et al., 2002). This interaction triggers a cascade of biochemical reactions in the cell regulating the specific activity of hormone-responsive genes. More than 50 hormones have been identified in humans. They control and regulate many biological processes, such as blood sugar control (insulin), body growth, differentiation, and function of reproductive organs (estradiol and testosterone). Several conditions can cause issues in the endocrine system. Some of the most common disorders are underproduction or overproduction of a certain hormone, a malfunction in the production of a hormone or in its ability to function correctly. The causes of these disorders are various: wrong response of our body to hormones, stress, infections, and some chemicals called endocrine disruptors (Malcomson and Nagy 2015).

Endocrine disrupting compounds (EDCs)

Since the 1990s, endocrine disrupting compounds have begun to arouse growing interest in the European and international panorama of research and risk assessment in the fields of health, food safety, and the environment. The first definition of an endocrine disruptor was published in the concluding report of a workshop held in April 1995 in North Carolina organized by the U.S. Environmental Protection Agency (EPA). In Europe, the first definition of the phenomenon took place in December 1996, on the occasion of the "European Workshop on the Impact of Endocrine Disrupters on Human Health and Wildlife", held in Weybridge (United Kingdom), organized to address the problem of substances that alter the endocrine system. In this context, the following definition has been agreed by the international scientific community: "*An ED is an exogenous substance that causes adverse health effects in an intact organism, or its progeny, secondary to changes in endocrine function.*" (World Health Organisation/ European Centre for Environment and Health 1996). Starting from this definition, in 2002, the International Program on Chemical Safety (IPCS) developed the" official "definition adopted by the European Union:" (..) endocrine disruptors are defined in a generic sense as follows:

- *An endocrine disruptor is an exogenous substance or mixture that alters function (s) of the endocrine system and consequently causes adverse health effects in an intact organism, or its progeny, or (sub) populations.*
- *A potential endocrine disruptor is an exogenous substance or mixture that possesses properties that might be expected to lead to endocrine disruption in an intact organism, or its progeny, or (sub) populations.*" (WHO 2010).

This is an innovative definition. In fact, generally, the highlighted effects (endpoint) are directly observed in defining the toxicity of chemicals, while in this definition the new and additional element is the concept of “mode of action”, which is the impact mode of a chemical substance.

Also, the U.S. Environmental Protection Agency (EPA) gave an important definition of these chemicals. In fact, it 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” (Diamanti-Kandarakis et al., 2009).

The Scientific Committee of the European Food Safety Authority (EFSA), in approving this definition of the IPCS of 2002, concluded that, in order for a substance to be identified

as an Endocrine Disruptor, "*there must be a basis of reasonable evidence of a causal relationship, biologically plausible, between endocrine activity and the induced negative effect, observed in an intact organism or in a (sub) population* ". In conclusion, natural or synthetic endocrine disruptors can be identified based on the presence of three elements: i) endocrine activity; ii) negative effect in an intact organism or in a (sub) population; and iii) demonstrated or plausible causal relationship between the two (EFSA, 2013).

A wide range of substances, both natural and synthetic ones, cause endocrine disruption, including pharmaceuticals, mycotoxins, dioxins, polychlorinated biphenyls, pesticides (i.e., dichlorodiphenyltrichloroethane, commonly known as DDT, glyphosate, pyriproxyfen), and plasticizers (i.e., bisphenol A, phthalates). They can be found in many everyday products, including plastic bottles, metal food cans, air, detergents, water, food, toys, and cosmetics.

Endocrine disrupting chemicals can act through different mechanisms (Figure 1): i) mimicking the action of a naturally-produced hormone; ii) blocking hormone receptors in cells, thereby preventing the action of normal hormones; or iii) interacting indirectly by influencing the biosynthesis or availability of normal hormones (Schug et al., 2011). These disruptions can cause adverse effects such as Parkinson's and Alzheimer's diseases, metabolic disorders, diabetes, cardiovascular disease, obesity, early puberty, reproductive function problems in males and females, cancers, and several other disorders (Lorenzetti and Narciso, 2012).

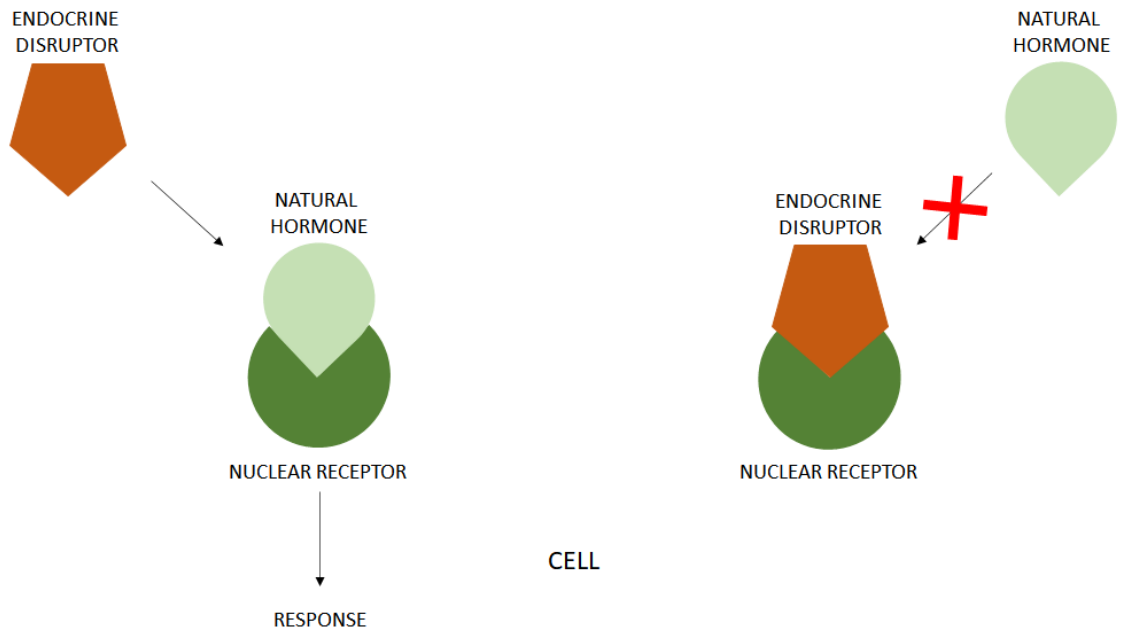


Figure 1. Common molecular mechanism of an endocrine disruptor. Endocrine disruptors act as receptors (especially nuclear receptors) binding inhibitors causing harmful effects.

These molecules are structurally and functionally similar to many hormones and for this reason, they are capable to mimic them in the modes of action, transport, and storage within tissues. Given the properties of these chemicals, they are particularly well suited for activating and antagonizing nuclear hormone receptors (i.e., androgen receptor, estrogen receptor, aryl hydrocarbon receptor, pregnane X receptor, constitutive androstane receptor, estrogen-related receptor, glucocorticoid receptor, thyroid hormone receptor, retinoid X receptor, etc.) (Schug et al., 2011) (Thomas Zoeller et al., 2012).

The EU legislation of EDCs

In the past decades, in order to limit human exposure to EDCs several regulatory and policy measures were taken. In December 1999, the European Commission adopted a Communication on a community strategy for endocrine disruptors with the objectives of identifying the endocrine disruption problem, its causes, and consequences and determining appropriate policy actions on the basis of the precautionary principle to respond quickly and effectively to the problem. EDCs are also dealt with under various pieces of EU legislation concerning different types of chemicals and with different regulatory purposes, such as the Regulation 1907/2006 on the Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH). In this regulation, ECDs are considered as substances of very high concern (SVHCs) similar to the regulatory concern posed to cancerogenic, mutagenic, and toxic molecules (Lorenzetti and Cozzini, 2017). In 2018, the EU reaffirmed its application of the precautionary principle and aim to minimize overall EDC exposures. The Member States have also launched several initiatives concerning EDCs.

Moreover, humans are exposed to multiple chemicals both simultaneously and in sequence in everyday life. In particular, chemical mixtures play a critical role in the development of adverse effects, and, in the majority of the cases, multiple EDCs may be more harmful even when single exposures are below the observable effect levels. Currently, human exposure to chemical mixtures is not considered when assessing the FCCs health impacts (Muncke et al., 2020).The problem is that these regulations and policies are insufficient to minimize exposure to the vast majority of EDCs. In fact, the

current approach to limiting exposure to EDCs in humans is dangerously slow and insufficient, and too few chemicals on the market have been thoroughly tested for endocrine-disrupting properties. Moreover, the list of chemicals requiring evaluation raises every year. Then, it is necessary that the EU, through its relevant bodies, gathers scientific evidence on EDCs, strengthens research and development efforts, improves the legislative framework, and aims at the development of an appropriate testing strategy based on expanded and alternative test methods to conclusively identify EDCs (Kassotis et al., 2020).

Food Contact Chemicals (FCCs)

Food contact chemicals (FCCs) are the chemical constituents of food contact materials (FCMs), that are the materials that come into contact with food, such as plastics, papers, glass, and finished food articles (FCAs), that are the final product used to store and/or to contain food, such as bottles and wraps. The food contact chemicals definition is supported by the Food Packaging Forum (www.foodpackagingforum.org), a charitable non-profit foundation based in Zurich, Switzerland, that provides scientific information of high-quality related to food packaging and the relative impact on health. Essentially, food contact chemicals can be defined as all chemicals which are not part of food but that come into contact with it. Because these chemicals are present in food contact materials, they can migrate into food (the migration depends on the nature of the FCMs, the temperature and the duration of the contact between the food and the FCMs, the nature of foodstuffs and their physical and chemical proprieties) and with a high probability, they could be ingested by most of the human population (Grob et al. 2006).

Food contact materials, and, consequently, food contact articles, can be divided into two groups: intentionally added substances (IASs) and non-intentionally added substances (NIASs). IASs are all chemical components that are deliberately used to manufacture FCMs and FCAs. Instead, NIASs refer to chemical components present in FCMs but that have not been added intentionally during the production process of a product and, thus, they do not have any specific function (Geueke, 2018). Several studies estimated that approximately 12,000 IASs and 30,000 to 100,000 NIASs can migrate into food from various food contact materials and that these are the most relevant source of human exposure to plasticizers (Groh et al. 2021) (Muncke et al. 2020). In addition, food contact material is not the only source of unintentional molecules present in the food. This broader class includes very heterogeneous chemicals which accidentally contaminate the food

product, like environmental pollutants, chemical residues due to human activities such as farming (i.e., pesticides), industry (i.e., dioxins, polychlorinated biphenyls (PCBs)), or as a result of human cooking and processing. But not only substances that accidentally contaminate food are considered food contact chemicals. Food additives and flavourings, intentionally added substances, can be included in the classes of food contact chemicals. 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. Man-made compounds are not the only molecules to be concerned about in the food contamination context. Several molecules (i.e., mycotoxins) are naturally occurring in the food supply due to their release in food products by plants, animals, or microorganisms.

Migration can impact food quality (some substances can alter the organoleptic aspects of food) and food safety (some substances may be harmful to human and animal health). Ensuring and complying with food safety is not a simple task, unfortunately. Different aspects should be considered, such as the good manufacturing practice (GMP) that must be followed during the food contact materials manufacturing chain, or different procedures that should be adopted to evaluate the safety of food contact materials constituents.

In Europe, different types of legislation regulate food contact materials and food contact articles. One of the most important is the Regulation (EC) No 1935/2004 on materials and articles intended to come into contact with food. This regulation requires that all FCMs and FCAs intended to come into contact with food, comply with the Framework Regulation. The principles included in this regulation establish that FCMs must not release their constituents into food at levels harmful to human health and they must not change food composition, taste, and odour in an unacceptable way.

Bisphenols

Bisphenols are a group of chemicals used to manufacture plastics and epoxy resins and are found in many products, such as food and drink packaging, store receipts, and medical devices. Among bisphenols, bisphenol A (BPA) has been shown to be an endocrine disruptor due to its ability to interfere with hormone homeostasis and, in particular, with estrogen receptor (ER), while bisphenol S (BPS) is recognized as a novel environmental pollutant and suspected to have similar endocrine disruptor (ED)-like concern than BPA for animal and human health (Duan et al., 2018) (Wu et al., 2018). Several studies have

demonstrated that BPA causes adverse health effects, such as skin reactions and respiratory irritation, reproductive, metabolic, and cardiovascular disorders, immunological and central nervous system diseases, and triggering and development of hormone-dependent cancers (Chen et al., 2001) (Patisaul and Carolina, 2019) (Pjanic, 2017) (Prins et al., 2018) (Stillwater et al., 2020). Moreover, BPA is listed in the Candidate List of substances of very high concern (SVHCs) due to its toxicity for reproduction and endocrine-disrupting properties. Some directives and regulations have been issued in particular for BPA, such as 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) 10/2011.

Dioxins

Dioxins are persistent environmental pollutants (POPs) that are produced by industrial processes including incineration, chlorine bleaching of paper, and the manufacture of some pesticides and herbicides, but also from many natural processes, such as volcanic eruptions and forest fires (Schecter et al., 2006). Dioxins are extremely persistent, bioaccumulative, and of concern because of their highly toxic potential. In humans and animals dioxins have been shown to be a risk of factors for several disorders both in short- and in long-term: chloracne and patchy darkening of the skin, cancer, reproductive and development disorders, diabetes, thyroid disorders, and many others (Birnbaum and Carolina, 1995) (Fingerhut et al., 1991) (Longnecker et al., 2015) (Pavuk et al., 1997) (Schecter and Gasiewicz, 2003) (Steenland et al., 1999). The WHO's International Agency for Research on Cancer (IARC) evaluated 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) as a "known human carcinogen", often called also "the most toxic man-made chemical". Two types of legislation are enacted for dioxins: the Commission Regulation (EU) 2017/644 that laying down methods of sampling and analysis for the control of levels of dioxins, dioxin-like PCBs and non-dioxin-like PCBs in certain foodstuffs and repealing Regulation (EU) 589/2014, and the Commission Recommendation 2013/711/EU on the reduction of the presence of dioxins, furans and PCBs in feed and food as amended by Commission Recommendation 2014/663/EU.

Food additives and Flavourings

Food additives are substances added intentionally to foodstuffs to perform certain technological functions. They are mainly used as colorants (to add or restore colour in a

food), preservatives (to prolong the food shelf-life of foods by protecting them against micro-organisms), antioxidants (to protect the food against oxidation), and flour treatment agents (to improve baking quality). In the European Union, all food additives are identified by an E number. Flavourings are substances used to impart taste and/or smell to food. 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. For these reasons, food additives and flavourings are constantly under control by Organizations, such as the Food and Agriculture Organization of the United Nations (FAO) and the World Health Organization (WHO) (Joint FAO/WHO Expert Committee on Food Additives) and Authorities, such as EFSA. According to Annex I of the Special Report 2/2019, to date, the European Union approve 334 food additives and 2549 food 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, the use conditions, the labelling, and the procedures.

Furans

Furans are highly volatile compounds abundant in the environment produced by processed food (thermally processed foods), industrial processes, and smoke (cigarettes, wood, exhaust gas). The potential health risks of these substances are well known. Based on several studies, furans may be different effects depending on the exposure: short-time exposure may be irritating to the skin, eyes, and respiratory tract, while long-term exposure may have effects on the liver and kidneys causing cancers (Everett and Thompson, 2014) (Food and Jecfa, 2011) (Nielsen et al., 2017) (Turyk et al., 2007). The IARC concluded that the evidence in humans for the carcinogenicity of furan was inadequate. However, there was sufficient evidence in experimental animals to classify furans as possibly carcinogenic to humans (Group 2B). In March 2007 the Commission adopted a Recommendation on the monitoring of the presence of furans in foodstuffs.

Mycotoxins

Mycotoxins are toxic secondary metabolites produced by fungi belonging to different genera such as *Aspergillus*, *Penicillium*, *Alternaria*, and *Fusarium*, that can grow on a variety of different crops and occur before/after harvest, during storage, on/in the food

itself often under warm and damp conditions. The Food and Agriculture Organization (FAO) estimated that each year 25% of global agricultural products are contaminated by mycotoxins (Boutrif & Canet, 1998). The toxic effects of mycotoxins are well known, in particular of ochratoxin A (OTA), zearalenone (ZEN), aflatoxin B1 (AFB1), fumonisin B1 (FBB1), deoxynivalenol (DON), and patulin that cause acute and chronic diseases, such as cancer, carcinogenesis, genotoxicity, immunotoxicity, mutagenicity, kidney toxicity, nervous disorders, and many others (Ahmadi et al., 2019) (Altunay et al., 2019) (Bennett and Klich, 2003) (Do et al., 2020) (Kőszegi and Poór, 2016) (Liu et al., 2017) (Marasas et al., 2004) (Travis R Bui-Klimke, 2015). The IARC has performed the carcinogenic hazard assessment of some mycotoxins in humans, especially of aflatoxins defined as carcinogenic to humans (Group 1), fumonisins and ochratoxin A defined possible carcinogens to humans (Group 2B). Two types of legislation are enacted for mycotoxins: the Commission Regulation (EC) 401/2006 that laying down the methods of sampling and analysis for the official control of the levels of mycotoxins in foodstuffs, and the Commission Recommendation 2012/154/EU that monitoring of the presence of ergot alkaloids in feed and food.

Polychlorinated biphenyls (PCBs)

Polychlorinated biphenyls (PCBs) are a group of manmade chemicals used widely in electrical equipment, hydraulic fluids, heat transfer fluids, lubricants, and plasticizers. PCBs have been categorized by the IARC as “Probably carcinogenic to humans” (Group 2A), and by the National Toxicology Program 11th Report on Carcinogens as “Reasonably anticipated to be human carcinogens”. PCBs are found throughout the environment and they can enter the body by eating or drinking contaminated food, through inhalation, or by dermal contact. PCBs can cause short-term changes in the activity of the liver, and they can affect the immune, endocrine (thyroid), and reproductive systems (Faroon et al., 2000) (Faroon and Ruiz, 2016) (Robertson and Ludewig, 2011) (Silverstone et al., 2012). Developing fetuses and young children are the most vulnerable to PCBs, which cause low birth weight, development problems, and high lifetime risk for several diseases (Carpenter, 2006).

Pesticides

A pesticide is defined as any substance or mixture of substances used to prevent, destroy, or control any pest (vectors of human or animal disease, unwanted species of plants or

animals) or administered to animals for the control of insects, or other pests in/on their bodies. Each year, over 4 million tonnes of pesticides are used all over the world, and more than 25 million agricultural workers experience unintentional poison by pesticides (Brief, 2018). This is due to the fact that in many developing countries programs to control exposures are limited or non-existent and in many cases that the maximum limits allowed for pesticides are not respected. Pesticides are known to be extremely useful and beneficial agents, but, at the same time, they have an extremely high acute toxicity for humans and other non-invasive species caused a number of health effects. As Mostafalou and Abdollahi report in their review, pesticides can cause short-term adverse health effects, called acute effects, such as eyes and skin irritation, nausea, vomiting, respiratory tract irritation, as well as chronic adverse effects, such as cancers, diabetes, neurodegenerative disorders, and amyotrophic lateral sclerosis (ALS), birth defects, and reproductive disorders, that can occur months or years after exposure (Mostafalou and Abdollahi, 2013). Two different regulations regulate pesticides: the Commission Regulation (EU) 37/2010 that regulates the maximum residue limits of pharmacologically active substances in foodstuffs of animal origin, and the Commission Regulation (EC) 1107/2009 concerns the placing of plant protection products on the market. Concerning animals, the Commission Regulation 1831/2003 and 429/2008 set out the authorized additives for use in animal feed and provide the rules for the presentation of the application to authorize new feed additives.

Phthalates

Phthalates are a group of chemicals used in several products, such as toys, detergents, lubricating oils, food packaging, pharmaceuticals, medical devices, and personal care products, such as nail polish, hair sprays, shampoos, perfumes. Humans and animals are exposed to phthalates through ingestion, inhalation, and skin contact during their whole lifetime. This has created concern that several studies have linked phthalates to interference with endocrine systems, development and reproduction, adverse outcomes of pregnancy, male fertility, obesity, and diabetes (Frederiksen et al., 2007) (Hauser and Calafat, 2005) (Heudorf and Mersch-sundermann, 2007) (Mankidy et al., 2013) (Tranfo et al., 2012). In particular, experimental studies have reported biological consequences of phthalate exposure relevant to child and prenatal development (Engel et al., 2010) (Miodovnik et al., 2011). Several strategies have been adopted from nations and the European Parliament for restricting the use of phthalates. Four phthalates, benzyl butyl

phthalate (BBP), bis-(2-ethylhexyl) phthalate (DEHP), bis-(2-methoxyethyl) phthalate dibutyl phthalate (DBP), diisobutyl phthalate (DIBP), are identified as substances of very high concern (Commission Implementing Decision (EU) 2017/1210) and they are listed as reprotoxic category 1B substances under EU Regulation (EC) 1272/2008. Moreover, some directives have been issued: Directive 2008/98/EC for waste and Directive 2009/48/EC for toy safety.

Structure-activity relationship and stereoisomers in food safety

A central axiom of chemistry is the structure-activity relationship (SAR), the relationship between the chemical structure of a molecule and its activity. Given that the activity of a molecule is reflected in its structure, similar molecules have similar activities. Therefore, this concept assumes that the structure of a molecule, such as its geometric (e.g. stereoisomerism) and electronic properties, contains the characteristics responsible for its physical, chemical, and biological properties.

Chirality is a geometric property of some molecules that cannot be superimposed on their mirror image by any combination of changes. A chiral molecule exists in two stereoisomers (substances with the same molecular formula, connectivity and bond multiplicity, and different spatial arrangement of two or more atoms) that are mirror images of each other, called enantiomers. The chirality phenomenon is common in nature and plays an important role in the biological recognition between an active molecule and its target. In fact, enantiomers can interact in different ways with receptors, proteins or/and enzymes. A lot of stereoisomers with identical physical and chemical properties will have different behaviour and will frequently show different biological activities. Enantiomeric forms can originate different effects, such as dissimilar taste or aroma, and affect the nutritional values of foods (vitamin C). Last but not least, different enantiomeric forms may vary in their toxicity. For example, more than 30% of pesticides are chiral and many of them are present in the environment as racemates. Therefore, pesticides enantiomers can have different toxicity and degradation rate with different impacts on human health. In the scientific literature, several cases are reported of differences in toxicological and environmental properties of stereoisomers. For this reason, stereoisomers need to be treated as different chemical components for the risk assessment (Bura et al. 2019).

Nuclear Receptors (NRs)

Nuclear receptors (NRs) are a superfamily of eukaryotic ligand-modulated transcription factors. In 1988 the first cDNA clones encoding polypeptides with structural features suggestive of steroid hormone receptors were cloned. Today, this superfamily is constituted of 48 members expressed in the animal kingdom. Nuclear receptors control numerous processes involved in development, growth, procreation, cell differentiation, proliferation, and the maintenance of homeostasis. NRs share a common structure composed of four independents but interacting functional modules: the modulator domain, the DNA-binding domain (DBD), the hinge region, and the ligand-binding domain (LBD) (Figure 2) (Chawta et al., 2001).

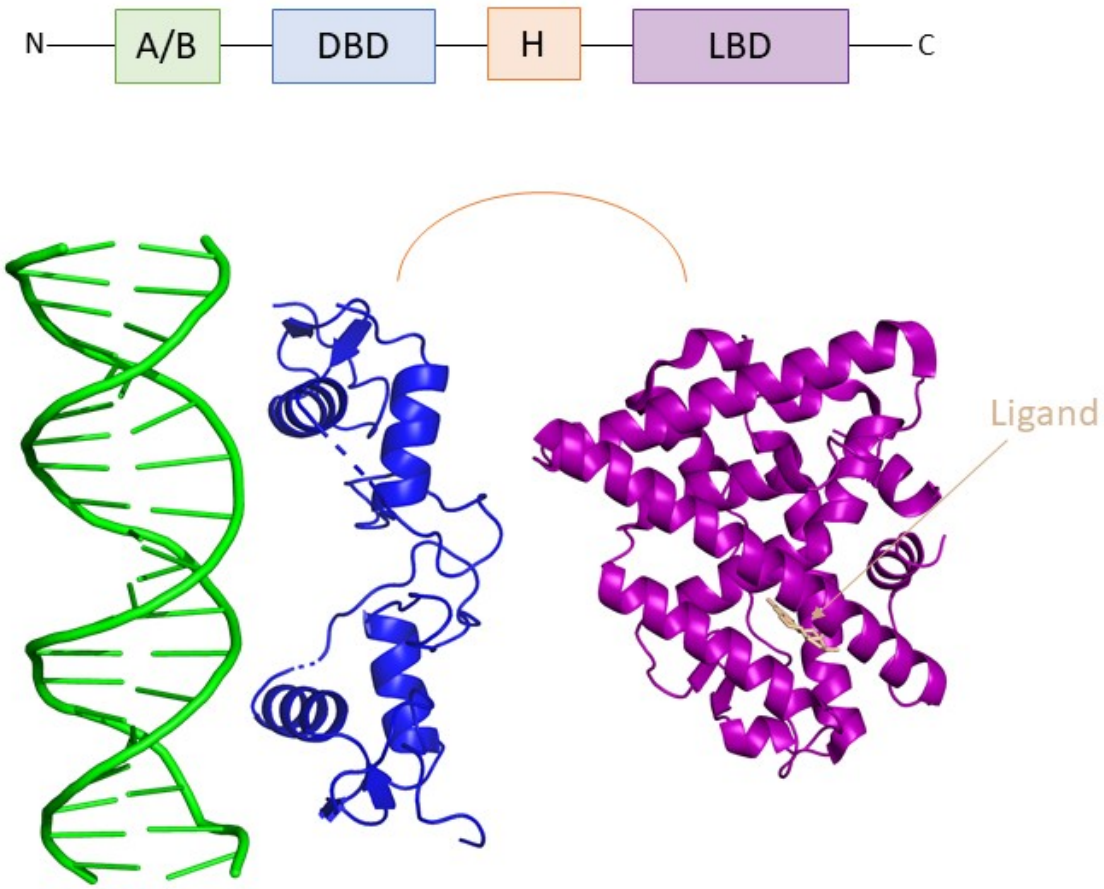


Figure 2. Structural organization of nuclear receptors.

The modulator domain, also called the A/B domain, is the most variable in length and sequence and contains the transcriptional activation function (AF-1). The DNA-binding domain (DBD) is the most conserved region. It contains two zinc finger modules encoded by approximately 70 amino acid residues and a carboxy-terminal extension (CTE). The

hinge region connected the DBD and the LBD. It is very flexible and highly variable in both length and primary sequence. The ligand-binding domain (LBD) is structured in α-helices and it is responsible for ligands binding. The LBD is contained in the E-F domain close to the carboxy terminus, together a region containing the AF-2 domain.

Nuclear receptors are activated by endogenous small lipophilic ligands that are able to cross the cell membrane and bind receptors. However, many nuclear receptors are "orphans" as their endogenous ligands are yet to be determined (Giguere, 1999). Once activated, nuclear receptors bind the promoter genes regulating and activating the gene transcription.

A peculiar characteristic of nuclear receptors is their ability to bind, in addition to endogenous ligands (e.g. estrogen, retinoic acid, fatty acids, and progesterone), very different types of molecules, also, unfortunately, "unintentional" binders, such as the endocrine disrupting compounds.

Computational methods in food science

Nowadays, computers and digital instruments support almost every activity of our life. In the last decades, the dissemination of technology has become the scientists' primary tool, thus allowing the achievement of more and more challenging tasks. Given the staggering amounts of data that scientists encounter in their day-to-day work, the development of more rapid and efficient methods is necessary. In this context computational science is placed a rapidly growing field that uses the power of computation to understand and solve complex problems. It is an area of science which contains many disciplines because it involves the development of models and simulations to understand natural systems. Problem domains for computational science include also computational biology, a very broad discipline that seeks to build models for different types of experimental data and biological systems using different mathematical and computational methods (e.g. algorithms, theories, software, etc.). The goal of this discipline is to gain understanding of natural systems, biology, applied mathematics, statistics, chemistry, molecular biology, genetics, genomics, and so on, mainly through the analysis of mathematical models implemented on computers. Computational biology gathers various expertise and techniques, most of which derive from bioinformatic, molecular biology, chemistry and medicinal chemistry. To the computational biology field belong also *in silico* toxicology and, more in general, the bioactivity assessment of foodborne compounds by using *in silico* methods.

The toxicity of a compound is the measure of any adverse effects it has on humans, animals, plants, or the environment. The studies of the effects of a chemical on human health are conducted by government agencies, such as the Scientific Committee of the European Food Safety Authority (EFSA) and the US Environmental Protection Agency (EPA), through risk assessments. Risks assessment of foreign chemicals to the body is still mainly based on *in vivo* (animal experimentation) tests, called toxicity tests, where animals are exposed to the test chemical. In Italy almost 700,000 animals are used in the laboratory, over 12 million in the European Union, to test drugs, chemicals, pesticides, detergents, and more (www.lav.it). During the 20th century, the agencies developed established animal test guidelines in order to reduce and replace *in vivo* experiments as much as possible. In fact, there are many advantages, including ethical and economic ones, to replace animal experimentations with other tests, called alternatives tests, such as *in vitro* or *in silico* methods. Moreover, some toxicity tests require hundreds to thousands of animals per substance examined. They can take months to years to conduct and can cost millions of dollars per substance examined.

The term *in silico* means literally silicium, a component of the computer chip. Therefore, *in silico* methods refer to experiments performed by computers (Hartung and Hoffmann 2009). The goal of *in silico* toxicity is to predict chemical toxicity through computational methods since they correlate the toxicity of a chemical with its structure. Nowadays, because of the ever-increasing availability and decreasing cost of computational power and algorithmic and software development, computational methods are used to study the molecular interactions of ligand-protein, structurally characterize binding sites of the proteins, develop targets compounds libraries, identify hits by virtual screening, estimate binding free energy, and optimize lead compounds. All of these elements can be used to rationalize and increase the efficiency, speed, and cost-effectiveness of evaluating the potential toxicological risk of chemicals, and not only. These new approaches could be implemented for foods and food ingredients with the purpose to evaluate novel foods from both nutritional and functional points of view (Cavaliere and Cozzini 2018).

Therefore it is important to stress that *in silico* methods should not be seen as an opposing method to *in vitro* and *in vivo* tests, but they should be considered as useful and preliminary methods to screen a huge number of molecules in a cost and time-effective manner. Given that people are exposed, both intentionally and not, to a large variety of different substances, a detailed characterization of the toxicological profile of all these substances is not feasible both from an economic and ethical point of view. In this way,

in silico methods can give considerable help to assign precedence to those substances for the which a safety evaluation (using *in vivo* and *in vitro* tests) is most urgent (Van Bossuyt et al., 2017).

Molecular Docking

Molecular Docking is a computational technique that attempts to predict and evaluate the structural chemical-physical interaction between two molecules (e.g. protein-protein, protein-ligand, protein-nucleic acid, ligand-nucleic acid) (Morris & Lim-Wilby, 2008). The molecular docking technique is based on the "lock and key" concept, developed by Emil Fischer in 1894. According to this model, the ligand (key) fits appropriately into the hole (binding pocket) of the protein (lock). Due to the formation of a series of weak bonds and favorable interactions, the ligand binds the receptor with high specificity and affinity. However, this model is too simplistic because both the protein and the ligand are not rigid bodies and protein/ligand flexibility should be considered (in this doctoral thesis the flexibility of both the protein and the ligand is considered).

Molecular docking is composed of two different steps: i) the prediction of the most favorable protein-ligand binding mode using molecular docking algorithms, ii) the ranking of a set of ligands using the values obtained from the scoring function implemented in the docking software, that is mathematical functions used to provide a value that predicts how tightly the two molecules interact. Various sampling algorithms have been developed in molecular docking software able to reproduce the experimental binding mode between two molecules.

Scoring functions can be divided into three groups: force-field-based methods, empirical scoring functions, and knowledge-based potentials. Although there has been some success in designing scoring functions that can describe protein-ligand interactions, some limitations have been pointed out. A solution to overcome these limitations and to have a more reliable docking result is consensus scoring, obtained using one package or more than one evaluation function to achieve a “convergence” to the best possible solution. In fact, it has been demonstrated that 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 & Fukunishi, 2007) (Wang, Lu, & Wang, 2003). Moreover, the use of three different scoring functions enhances the capability to reach hit rates from 10% up to 70% (Bissantz, Folkers, & Rognan, 2000).

Molecular Dynamics Simulation

Molecular dynamics (MD) simulation is a computational technique which, allows to study of the evolutionary dynamics of a physical and chemical system at the atomic and molecular levels. In the 1950s Alder & Wainwright studied the interaction of rigid spheres and in 1964 the first simulation of liquid argon was conducted by Raman (Alder and Wainwright, 1957). Molecular dynamics provides information on the temporal evolution of the molecular systems conformations, quantifies the properties of the system (e.g. the structure, the dynamics, the kinetics, and the thermodynamics), allows to explore the relationships between structures, dynamics, and function in biomolecules, and so on.

The molecular dynamic simulation is based on Newton's second law or the equation of motion:

$$F_i = m_i * a_i$$

where F_i is the force exerted on 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 as a function of time. From the trajectories, it is possible to determine the average value of these properties. In fact, by deriving the equation of motion it is 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 given by the sum of bonded energy (E_{bonded}), non-bonded energy ($E_{nonbonded}$), and other terms (E_{other}):

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

Moreover, the bonded energy is given by the stretching energy (E_{str}), the bending energy (E_{bend}), and the torsional energy (E_{tor}) terms:

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

while the non-bonded energy is given by the electric energy (E_{elec}) and Van der Waals energy (E_{vdw}) terms:

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

Each atom will be described by a new position characterized by new spatial coordinates. These coordinates, called conformations, will be very different from the starting ones (crystallographic or molecular modeling coordinates) depending on the movement it has made.

The resulting MD trajectory can be analysed to extract important information about the system (e.g. the root-mean-square-deviation (RMSD), which evaluates the general movements of the protein during the simulation time). Such as all computational methods, also molecular dynamics simulation benefits from the seemingly never-ending improvements in computer hardware. In particular with the advent of high-performance computers (HPCs), the molecular dynamics simulations that originally lasted less than 10 ps, today are often 1000 times as long (10 ns) but take a factor of about 50 less times for a system with the same dimension.

High Performance Computing

High Performance Computing (HPC) has become fundamental to scientific research. By performing millions and millions of calculations per second, high-performance computers help us to solve the most complex scientific challenges. The development of new technologies and techniques, combined with the availability of large computing resources, has allowed an important acceleration in: i) the study of new methods of machine learning, predictive analysis, and image processing; ii) making sense of the massive amounts of data generated by modern "omics" and genome sequencing technologies; iii) modeling increasingly large biomolecular systems using approaches such as quantum mechanics/molecular mechanics and molecular dynamics; iv) modeling biological networks and simulating how network perturbations lead to adverse outcomes and disease, and so on. HPC relies on the development of parallelized algorithms, that can spread the computational workload out among a number of computer cores that are conducting calculations simultaneously. HPC architectures have gone through rapid changes, such as multicore and manycore architectures, accelerator technologies, such as a graphics processing unit (GPU) designed to rapidly manipulate and alter memory to accelerate graphics rendering, persistent memory, and complex interconnection networks to connect compute nodes, processors, memory, and storage units, in order to meet the increasing computational demand of scientific applications. The HPC architecture can take different forms according to the owns needs: i) parallel computing, fundamental to deal with large and complex problems, allows the HPC clusters to perform calculations

simultaneously or in parallel; ii) cluster computing, where several computers or nodes are connected to each other through a local network to recreate an HPC cluster architecture; iii) grid computing, involving multiple networked computers sharing a common goal. In short, depending on the workload and processing goals, different HPC system architectures and support resources are available to help scientists get results in a timely manner and process huge amounts of data (Hager and Wellein 2010).

References

Ahmadi, A., Mohammad, N., Ramezani, M., Alibolandi, M., Lavaee, P., Sarreshtehdar, A., Abnous, K., Mohammad, S., 2019. A rapid and simple ratiometric fluorescent sensor for patulin detection based on a stabilized DNA duplex probe containing less amount of aptamer- involved base pairs. Talanta 204, 641–646. https://doi.org/10.1016/j.talanta.2019.06.057

Alberts B, Johnson A, Lewis J, et al. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002. Available from: https://www.ncbi.nlm.nih.gov/books/NBK21054/

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

Altunay, N., Elik, A., Gürkan, R., 2019. A novel, green and safe ultrasound-assisted emulsification liquid phase microextraction based on alcohol-based deep eutectic solvent for determination of patulin in fruit juices by spectrophotometry. J. Food Compos. Anal. 82, 103256. https://doi.org/10.1016/j.jfca.2019.103256

Bennett, J.W., Klich, M., 2003. Mycotoxins. Clin. Microbiol. Rev. 16, 497–516. https://doi.org/10.1016/B978-0-12-801238-3.02409-0

Birnbaum, L. S., 1995. Developmental effects of dioxins. Environmental health perspectives, 103 Suppl 7(Suppl 7), 89–94. https://doi.org/10.1289/ehp.95103s789

Bissantz, C., Folkers, G., & Rognan, D. (2000). Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *Journal of Medicinal Chemistry*, *43*(25), 4759–4767. https://doi.org/10.1021/jm0010441

Boutrif, E. & Canet, C., 1998. Mycotoxin prevention and control: FAO programmes. Rev. Médicine Vétérinaire 149: 681-694

Brief, F.A., 2018. Pesticides use Global , regional and country trends.

Bura, L.; Friel, A.; Magrans, J. O.; Parra-Morte, J. M.; Szentes, C. *Guidance of EFSA on*

Risk Assessments for Active Substances of Plant Protection Products That Have Stereoisomers as Components or Impurities and for Transformation Products of Active Substances That May Have Stereoisomers; 2019; Vol. 17. https://doi.org/10.2903/j.efsa.2019.5804

Carpenter DO. Polychlorinated biphenyls (PCBs): routes of exposure and effects on human health. Rev Environ Health. 2006 Jan-Mar;21(1):1-23. doi: 10.1515/reveh.2006.21.1.1

Cavaliere, F.; Cozzini, P. New *in silico* Trends in Food Toxicology. *Chem. Res. Toxicol.* 2018, *31* (10), 992–993. https://doi.org/10.1021/acs.chemrestox.8b00133.

Chawta, A., Repa, J. J., Evans, R. M., & Mangelsdorf, D. J. (2001). Nuclear receptors and lipid physiology: Opening the x-files. *Science*, *294*(5548), 1866–1870. https://doi.org/10.1126/science.294.5548.1866

Chen, M., Ike, M., Fujita, M., 2001. Acute Toxicity , Mutagenicity , and Estrogenicity of Bisphenol-A and Other Bisphenols 80–86. https://doi.org/10.1002/tox.10035

Colborn, T., Dumanoski, D., Myers, J.P., 1996. Our Stolen Future: Are We Threatening Our Fertility, Intelligence, and Survival? A Scientific Detective Story.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

Do, T.H., Tran, S.C., Le, C.D., Nguyen, H.B.T., Le, P.T.T., Le, H.H.T., Le, T.D., Thai-Nguyen, H.T., 2020. Dietary exposure and health risk characterization of aflatoxin B1, ochratoxin A, fumonisin B1, and zearalenone in food from different provinces in Northern Vietnam. Food Control 112, 107108. https://doi.org/10.1016/j.foodcont.2020.107108

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

EFSA, Scientific Opinion on the Hazard Assessment of Endocrine Disruptors: Scientific Criteria for Identification of Endocrine Disruptors and Appropriateness of Existing Test Methods for Assessing Effects Mediated by These Substances on Human Health and the Env. *EFSA J.* 2013, *11* (3). https://doi.org/10.2903/j.efsa.2013.3132

Engel, S.M., Miodovnik, A., Canfield, R.L., Zhu, C., Silva, M.J., Calafat, A.M., S., W.M., 2010. Children's Health Prenatal Phthalate Exposure Is Associated with Childhood

Behavior and Executive Functioning. Environ. Health Perspect. 118, 565–571. https://doi.org/10.1289/ehp.0901470

Everett, C.J., Thompson, O.M., 2014. Dioxins , furans and dioxin-like PCBs in human blood : Causes or consequences of diabetic nephropathy ? Environ. Res. 132, 126–131. https://doi.org/10.1016/j.envres.2014.03.043

Faroon, O., Jones, D., de Rosa, C., Effects of polychlorinated biphenyls on the nervous system. Toxicol Ind Health. 2000 Sep;16(7-8):305-33. doi: 10.1177/074823370001600708

Faroon, O., Ruiz, P., 2016. Polychlorinated biphenyls : New evidence from the last decade 32, 1825–1847. https://doi.org/10.1177/0748233715587849

Fingerhut, M.A., Halperin, W.E., Marlow, D.A., Piacitelli, L.A., Honchar, P.A., Sweeney, M.H., Greife, A.L., Dill, P.A., Steenland, K., Suruda, A.J., 1991. Cancer Mortality in Workers Exposed to 2,3,7,8-Tetrachlorodibenzo-P-Dioxin. N. Engl. J. Med. 212–218.

Food, W.H.O., Jecfa, F.A.O., 2011. Safety evaluation of certain contaminants in food 487–603.

Frederiksen, H., Skakkebæk, N.E., Andersson, A., 2007. Review Metabolism of phthalates in humans 899–911. https://doi.org/10.1002/mnfr.200600243

Giguere, V. (1999). Orphan nuclear receptors: From gene to function. *Endocrine Reviews*, *20*(5), 689–725. https://doi.org/10.1210/er.20.5.689

Grob, K.; Biedermann, M.; Scherbaum, E.; Roth, M.; Rieger, K. Food Contamination with Organic Materials in Perspective: Packaging Materials as the Largest and Least Controlled Source? A View Focusing on the European Situation. *Crit. Rev. Food Sci. Nutr.* 2006, *46* (7), 529–535. https://doi.org/10.1080/10408390500295490

Groh, K. J.; Geueke, B.; Martin, O.; Maffini, M.; Muncke, J. Overview of Intentionally Used Food Contact Chemicals and Their Hazards. *Environ. Int.* 2021, *150* (July 2020), 106225. https://doi.org/10.1016/j.envint.2020.106225

Geueke, B., 2018. Dossier - Non-intentionally added substances (NIAS). Food Packag. Forum 7. https://doi.org/10.5281/zenodo.1265331

Hartung, T.; Hoffmann, S. Food for Thought... on *in silico* Methods in Toxicology. *ALTEX* 2009, *26* (3), 155–166. https://doi.org/10.14573/altex.2009.3.155

Hauser, R., Calafat, A.M., 2005. Phthalates and human health 806–818. https://doi.org/10.1136/oem.2004.017590

Heudorf, U., Mersch-sundermann, V., 2007. Phthalates : Toxicology and exposure 210, 623–634. https://doi.org/10.1016/j.ijheh.2007.07.011

Kassotis, C. D.; Vandenberg, L. N.; Demeneix, B. A.; Porta, M.; Slama, R.; Trasande, L. Endocrine-Disrupting Chemicals: Economic, Regulatory, and Policy Implications. *Lancet Diabetes Endocrinol.* 2020, *8* (8), 719–730. https://doi.org/10.1016/S2213-8587(20)30128-5

Kőszegi, T., Poór, M., 2016. Ochratoxin a: Molecular interactions, mechanisms of toxicity and prevention at the molecular level. Toxins (Basel). 8, 1–25. https://doi.org/10.3390/toxins8040111

Liu, X., Guan, X., Xing, F., Lv, C., Dai, X., Liu, Y., 2017. Effect of water activity and temperature on the growth of Aspergillus flavus, the expression of aflatoxin biosynthetic genes and aflatoxin production in shelled peanuts. Food Control 82, 325–332. https://doi.org/10.1016/j.foodcont.2017.07.012

Longnecker, M.P., Michalek, J.E., Epidemiology, S., Jan, N., 2015. Serum Dioxin Level in Relation to Diabetes Mellitus among Air Force Veterans with Background Levels of Exposure Linked references are available on JSTOR for this article : 11, 44–48. doi: 10.1097/00001648-200001000-00010

Lorenzetti, S., Cozzini, P., 2017. The Substitution Principle within the REACH Regulation: Nuclear Receptor-Bound Endocrine Disruptors. Nucl. Recept. Res. 4, 1–11. https://doi.org/10.11131/2017/101205

Lorenzetti, S., Narciso, L., "Chapter 1. Nuclear Receptors: Connecting Human Health to the Environment," in Computational Approaches to Nuclear Receptors, RSC Drug Discovery, pp. 1–22, Royal Society of Chemistry, Cambridge, 2012. doi: 10.11131/2017/101205

Malcomson, R. D. G.; Nagy, A. The Endocrine System. *Keeling's Fetal Neonatal Pathol.* 2015, 671–702. https://doi.org/10.1007/978-3-319-19207-9_25.

Mankidy, R., Wiseman, S., Ma, H., Giesy, J.P., 2013. Biological impact of phthalates. Toxicol. Lett. 217, 50–58. https://doi.org/10.1016/j.toxlet.2012.11.025

Marasas W. F. O., Riley R. T., Hendricks K. A., Stevens V. L., Sadler T. W., Gelineau-Van Waes J., Missmer S. A., Cabrera J., Torres O., Gelderblom W. C. A., Allegood J., Martínez C., Maddox J., Miller J. D., Starr L., Sullards M. C., Roman A. V., K. A. V, and A.H.M.J., 2004. Fumonisins Disrupt Sphingolipid Metabolism, Folate Transport, and Neural Tube Development in Embryo Culture and In Vivo: A Potential Risk Factor for Human Neural Tube Defects among Populations

Consuming Fumonisin-Contaminated Maize. J. Nutr. 4, 711–716. https://doi.org/10.1093/jn/134.4.711

Miodovnik, A., Engel, S.M., Zhu, C., Ye, X., Soorya, L. V, Silva, M.J., Calafat, A.M., Wolff, M.S., 2011. NeuroToxicology Endocrine disruptors and childhood social impairment. Neurotoxicology 32, 261–267. https://doi.org/10.1016/j.neuro.2010.12.009

Morris G., M., & Lim-Wilby, M. (2008). Molecular Docking. In Andreas Kukol (Ed.), *Methods in Molecular Biology* (Vol. 443, pp. 365–382). Totowa, NJ. https://doi.org/10.1007/3-540-29623-9_3820

Mostafalou, S., Abdollahi, M., 2013. Pesticides and Human Chronic Diseases: Evidences, Mechanisms, and Perspectives. Toxicol. Appl. Pharmacol. https://doi.org/10.1016/j.taap.2013.01.025

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. Impacts of Food Contact Chemicals on Human Health: A Consensus Statement. *Environ. Heal. A Glob. Access Sci. Source* 2020, *19* (1), 1–12. https://doi.org/10.1186/s12940-020-0572-5

Nielsen, E., Bredsdorff, L., Olesen, P.T., Sharma, A.K., Beltoft, V., 2017. Identifying and collecting relevant literature related to the oral toxicity of furan and its methyl analogues , 2-methylfuran and 3-methylfuran. https://doi.org/10.2903/sp.efsa.2017.EN-1066

Patisaul, H.B., Carolina, N., 2019. Achieving CLARITY on Bisphenol A (BPA), Brain and Behavior. J. Neuroendocrinol. 0–3. https://doi.org/10.1111/jne.12730

Pavuk, M., Michalek, J. & Ketchum, N. Prostate cancer in US Air Force veterans of the Vietnam war. J Expo Sci Environ Epidemiol 16, 184–190 (2006). https://doi.org/10.1038/sj.jea.7500448

Prins, G.S., Islands, F., Smith, R., 2018. CLARITY-BPA Academic Laboratory Studies Identify Consistent Low-Dose Bisphenol A Effects on Multiple Organ Systems. Basic Clin. Pharmacol. Toxicol. https://doi.org/10.1111/bcpt.13125

Robertson, L. W., & Ludewig, G., 2011. Polychlorinated Biphenyl (PCB) carcinogenicity

with special emphasis on airborne PCBs. Gefahrstoffe, Reinhaltung der Luft = Air quality control, 71(1-2), 25–32.

Schecter, A., Birnbaum, L., Ryan, J.J., Constable, J.D., 2006. Dioxins : An overview 101, 419–428. https://doi.org/10.1016/j.envres.2005.12.003

Schecter, A.J., Gasiewicz, T.A., 2003. Dioxins and Health, Second. ed.

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

Silverstone, A.E., Rosenbaum, P.F., Weinstock, R.S., Bartell, S.M., Foushee, H.R., 2012. Polychlorinated Biphenyl (PCB) Exposure and Diabetes : Results from the Anniston Community Health Survey 120, 727–732. doi: 10.1289/ehp.1104247

Steenland K, Piacitelli L, Deddens J, Fingerhut M, Chang LI. Cancer, heart disease, and diabetes in workers exposed to 2,3,7,8-tetrachlorodibenzo-p-dioxin. J Natl Cancer Inst. 1999 May 5;91(9):779-86. doi: 10.1093/jnci/91.9.779

Stillwater, B.J., Bull, A.C., Romagnolo, D.F., Neumayer, L.A., 2020. Bisphenols and Risk of Breast Cancer : A Narrative Review of the Impact of Diet and Bioactive Food Components 7, 1–14. https://doi.org/10.3389/fnut.2020.581388

Teramoto, R., & Fukunishi, H. (2007). Supervised consensus scoring for docking and virtual screening. *Journal of Chemical Information and Modeling*, *47*(2), 526–534. https://doi.org/10.1021/ci6004993

Thomas Zoeller, R., Brown, T.R., Doan, L.L., Gore, A.C., Skakkebaek, N.E., Soto, A.M., Woodruff, T.J., Vom Saal, F.S., 2012. Endocrine-disrupting chemicals and public health protection: A statement of principles from the Endocrine Society. Endocrinology 153, 4097–4110. https://doi.org/10.1210/en.2012-1422

Tranfo, G., Caporossi, L., Paci, E., Aragona, C., Romanzi, D., Carolis, C. De, Rosa, M. De, Capanna, S., Papaleo, B., Pera, A., 2012. Urinary phthalate monoesters concentration in couples with infertility problems. Toxicol. Lett. 213, 15–20. https://doi.org/10.1016/j.toxlet.2011.11.033

Travis R Bui-Klimke, F.W., 2015. Ochratoxin A and human health risk: A review of the evidence. Crit. Rev. Food Sci. Nutr. 1860–1869. doi: 10.1080/10408398.2012.724480

Turyk, M.E., Anderson, H.A., Persky, V.W., 2007. Relationships of Thyroid Hormones with Polychlorinated Biphenyls , Dioxins , Furans , and DDE in Adults 115, 1197–1203. https://doi.org/10.1289/ehp.10179

Van Bossuyt, M., Van Hoeck, E., Vanhaecke, T., Rogiers, V., Mertens, B., 2017. Safeguarding human health using in silico tools? Arch. Toxicol. 91, 2705–2706. https://doi.org/10.1007/s00204-017-1931-z

Wang, R., Lu, Y., & Wang, S. (2003). Comparative evaluation of 11 scoring functions for molecular docking. *Journal of Medicinal Chemistry*, *46*(12), 2287–2303. https://doi.org/10.1021/jm0203783

WHO. International Programme for Chemical Safety (IPCS) in Global Assessment of the State-of-the-Science of Endocrine Disruptors. . 2010, *Chapter 1:*, 1–3.

World Health Organisation/ European Centre for Environment and Health. European Workshop on the Impact of Endocrine Disruptors on Human Health and Wildlife (2-4 December 1996, Weybridge, UK). *Rep. Proceedings, Publ. EUR Paris Fr.* 1996, *EUR 17549*, 1–127.

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

CHAPTER 2

Aim and outline of the thesis

Both humans and animals are exposed to chemicals in everyday life. Many of these, called endocrine-disrupting chemicals, can interfere with the endocrine systems altering the production of human hormones mainly by acting through their interaction with nuclear receptors. According to Eurostat, the EU production of chemicals hazardous to health reached 211 million tonnes in 2019. Thus, the possibility that some of these chemicals interact negatively with the human and animal body is unavoidable making the endocrine disruptors problem are an emerging one over the world. Wherefore, the increasing number of molecules released every year along with the long different steps needed to evaluate every single substance before entering the marketing system makes the endocrine-disrupting evaluation very challenging and long-term. Moreover, in the past decades, efforts and policies of Authorities and of the single States have been proved inefficient to decrease and minimize human exposure to endocrine disruptors. In addition, the endocrine-disrupting properties of several chemicals, the detailed characterization of their toxicological profile, and their effects on the human body are still waiting for evaluation. Added to this is that some weaknesses exist in testing approaches used for evaluating endocrine-disrupting chemicals, such as the costs, the time, and the enormous amount of test animals use. The aim of this dissertation is to detect the possible endocrine disruptors in food and the possible binding of these food contact chemicals with nuclear receptors using the *in silico* methods. These methods are used to study the molecular interactions of ligand-protein, structurally characterize binding sites of the proteins, develop targets compounds libraries, screen thousands of chemicals, and study the evolutionary dynamics of a physical and chemical system at the atomic and molecular levels. All of these elements can be used to rationalize and increase the efficiency, speed, and cost-effectiveness of evaluating the potential toxicological risk of chemicals, and not only In this way, *in silico* methods can give considerable help to screen a huge number of molecules at a cost and time-effective manner and to assign precedence to those substances for the which a safety evaluation is most urgent using *in vivo* and *in vitro* tests.

Using literature data and repositories data, all the endocrine disruptors are classified in a Structured Query Language (SQL) Database in order to make the data extraction and analysis quicker and more efficient. Then, nuclear receptors structures are analysed, and the possible binding of these molecules with nuclear receptors is determined using a combination of *in silico* (molecular docking) and statistical approach in order to obtain a final global evaluation based on the natural ligand.

A description of computational methods application in food safety is provided in **Chapter 3**. In this book chapter, repository or database design, screening, molecular docking, and consensus scoring techniques are described in order to give an overview of the application of these methods in problems of food safety.

In **Chapter 4**, the attention is focused on the most common *in vitro* bioassays and *in silico* analysis as methods used to screen food contact chemicals against nuclear receptors to evaluate their endocrine disruptors' proprieties.

In **Chapter 5**, a database (foodchem) with a high level of data curation from which retrieve chemical, structure, and regulative information about all food contact chemicals was created. After that, the 8091 food contact chemicals contained in foodchem database were screened against 31 nuclear receptors with the purpose to identify the molecules that require major attention about their safety for the human body.

An application of what is described in chapter 5 is reported in **Chapter 6**. Given the large potential impact of mycotoxins in terms of human exposure and related health effects, in this work, the integrated *in silico* and statistical approach was used in order to discover the potential endocrine disruptor activity of these molecules.

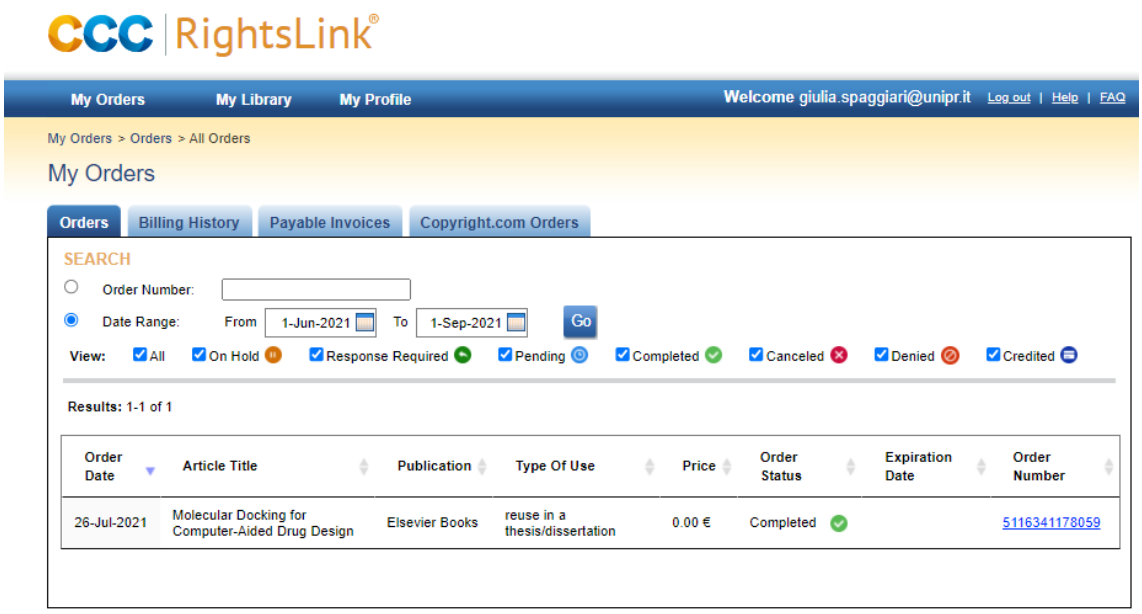
In **Chapter 7**, an application of *in silico* approaches to discover endocrine disruptors is presented. In particular, computational techniques are applied to investigate the possible negative effects of two pesticides, pyriproxyfen and its metabolite, the 4'-OH-pyriproxyfen, on human and bees health.

CHAPTER 3

Molecular Docking: A Contemporary Story About Food Safety

The content of this chapter has also been published in:

Cavaliere, F., Spaggiari, G., Cozzini, P., 2021. Molecular Docking: A Contemporary Story About Food Safety. Mol. Docking Comput. Drug Des. 479–492. https://doi.org/10.1016/b978-0-12-822312-3.00025-4



CCC RightsLink®

My Orders | My Library | My Profile | Welcome giulia.spaggiari@unipr.it Log out | Help | FAQ

My Orders > Orders > All Orders

My Orders

Orders | Billing History | Payable Invoices | Copyright.com Orders

SEARCH

Order Number:

Date Range: From 1-Jun-2021 To 1-Sep-2021 Go

View: All On Hold Response Required Pending Completed Canceled Denied Credited

Results: 1-1 of 1

Order Date	Article Title	Publication	Type Of Use	Price	Order Status	Expiration Date	Order Number
26-Jul-2021	Molecular Docking for Computer-Aided Drug Design	Elsevier Books	reuse in a thesis/dissertation	0.00 €	Completed		5116341178059

Introducion

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 women - all 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.

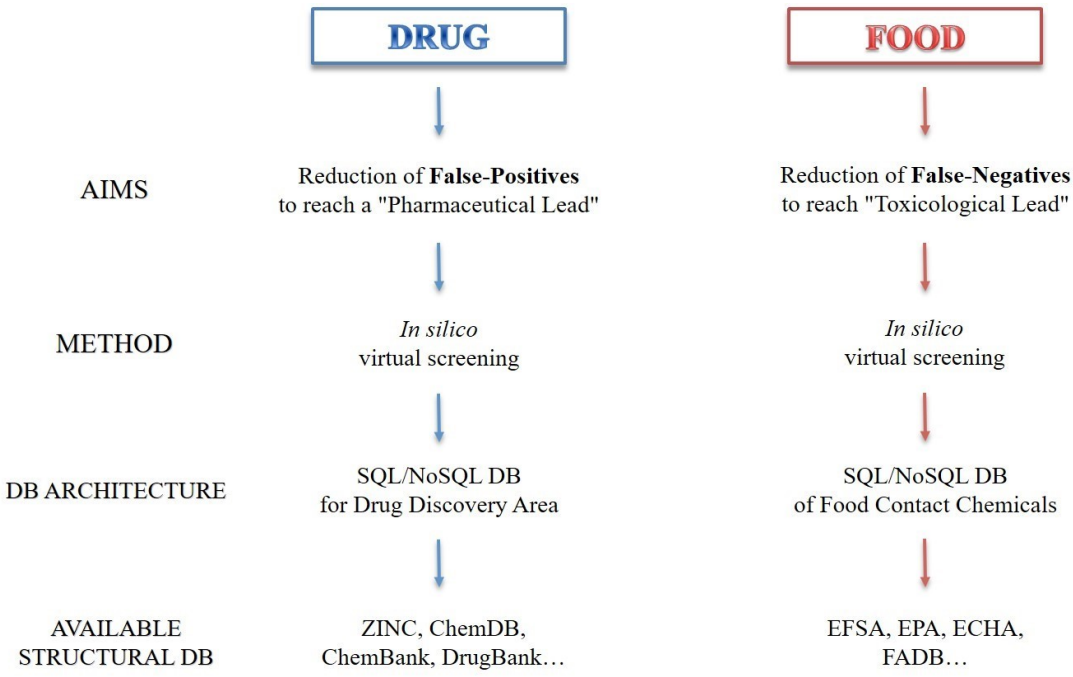


Figure 1. The different approaches to screen compounds in drug discovery and food safety' areas. The same *in silico* methods, widely used in drug discovery, can be applied in the food field. The unique difference between them is the aim of the screening process: in drug discovery, it is important to 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 to exclude true negatives.

Molecular docking is a well-know 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.

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 600 million (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-tobackground 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/).

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 term in 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% and 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.

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 Figure 2.

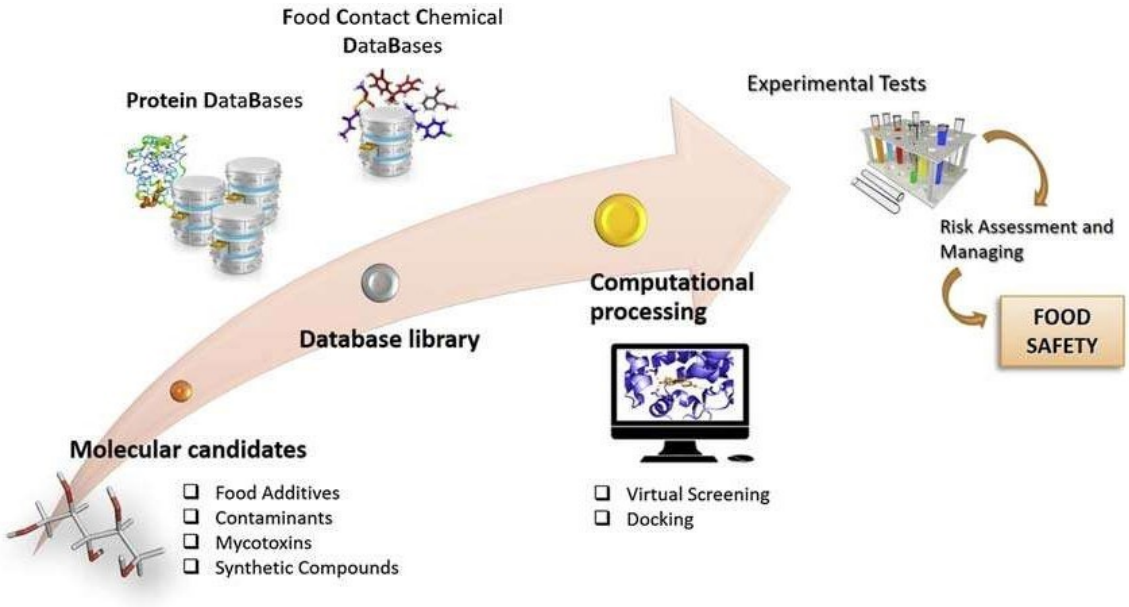


Figure 2. The *in silico* approach in food safety schema.

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. Structurebased 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 (10-40 kJ/mol); (2) hydrophobic interactions that constitute the "driving force" capable of promoting bond formation; (3) van der Waals forces (0.03-

0.1 kcal/mol); (4) electrostatic interactions (0.3-4 kcal/mol); (5) π-π 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):

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

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:

$$\Delta G^0 = RT\ In\ Kd$$

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 water-receptor and water-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, DG bind), we must use an equation like this:

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

where $\Delta G°_{int}$ is the interaction free energy of the complex, the solvation energy of the ligand ($\Delta G°^{lig}_{solv}$), the protein ($\Delta G°^{prot}_{solv}$), and the complex ($\Delta G°^{compl}_{solv}$), and the entropic ($T\Delta S°$) and conformational ($\Delta\lambda$) changes (Spyrakis, Cozzini, and Kellogg 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 differentiate based on their two main components: the sampling algorithm, that search the possible molecule position, and the scoring function, that evaluate the interaction energy of each position. The first is an "easy" geometrical aspect that "mix" 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 $(360^o/x)^n$.

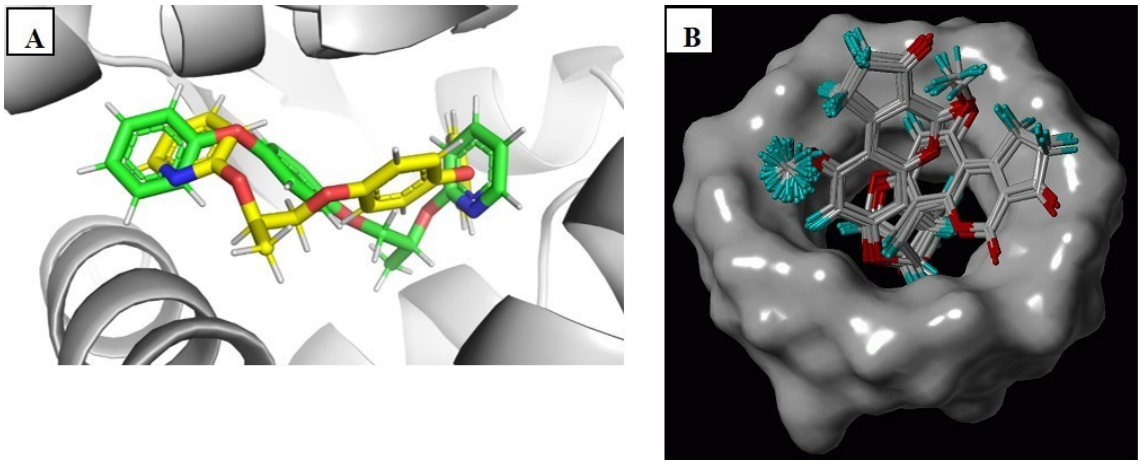


Figure 3. Different poses generated by a generic docking software (GOLD). (A) A

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. Semi-flexible 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 both the protein and the ligand.

We can summarize scoring functions in three classes: i) Force-field based, 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; ii) Empirical, where the binding affinity between the two molecules is estimated by the number of various types of interactions; iii) 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 and Rarey 2004), Glide (Friesner et al. 2004), Surflex (Jain 2003), distinguished by the algorithms, the evaluation methods, the docking types (rigid, semi-flexible 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 and 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 ionisable 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 protein-ligand 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 (RMSD) 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 protein-ligand 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%-67.8% for commercial programs and in the range of 47.4%-68.4% for the academic one's (Pagadala, Syed, and Tuszynski 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 and 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 and 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, Folkers, and Rognan 2000).

Case study

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. Since the majority of our key studies are focused on nuclear receptors and how food contact chemicals may act as endocrine disruptors, Figure 4. shows a schematic view of the perturbation induced by endocrine disruptor compounds (EDCs) activity.

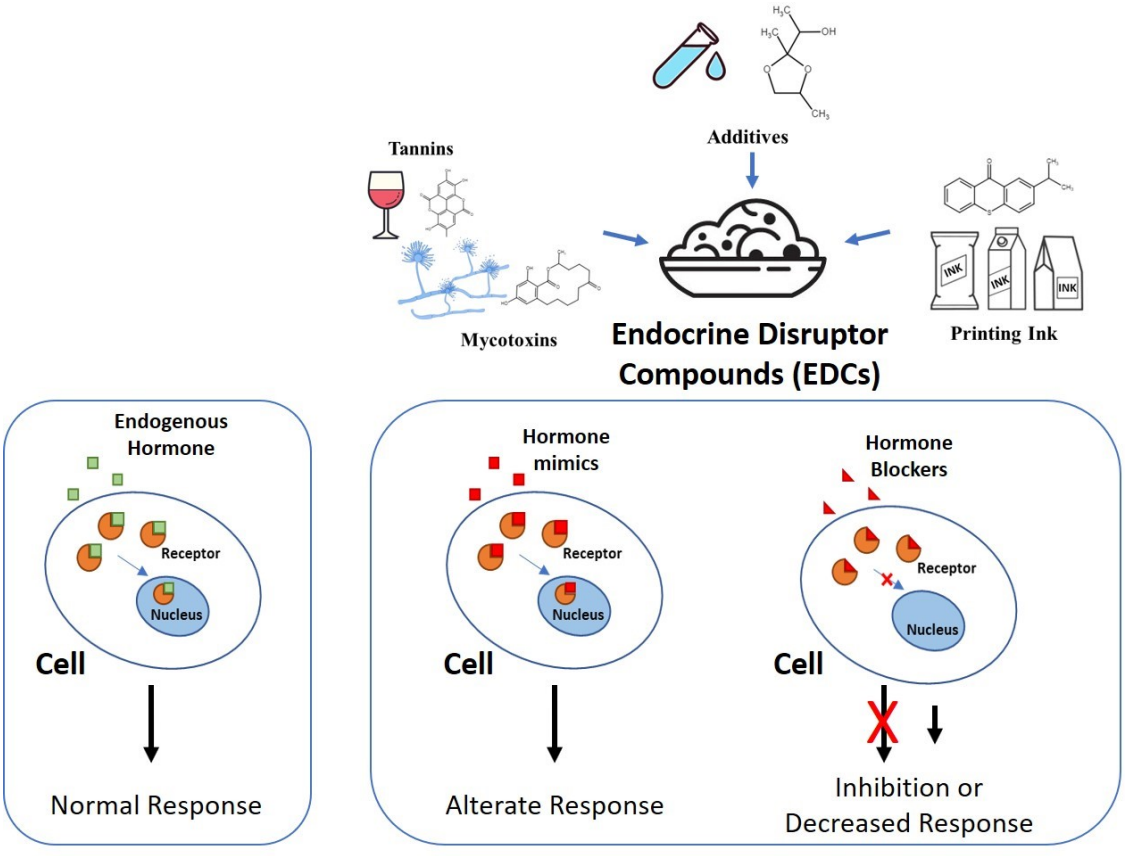


Figure 4. How food contact chemicals can affect the nuclear receptors pathway.

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.

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 nuclear receptors. Two well-known nuclear receptors are recognized as responsible for breast cancer in women and prostate cancer for men: Estrogen receptor and Androgen receptor, respectively.

Aflatoxins and Ochratoxins

Cyclodextrins are cheap and relatively easy to manage, 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 MOA understanding (Mechanism of Binding) allows designing specific cyclodextrins customized for different toxin structures (Fig. 5) (Amadasi et al. 2007). Aflatoxins and Ochratoxins (Fig. 7) affecting a large number of mays and grains production in Italy (more or less 5%).

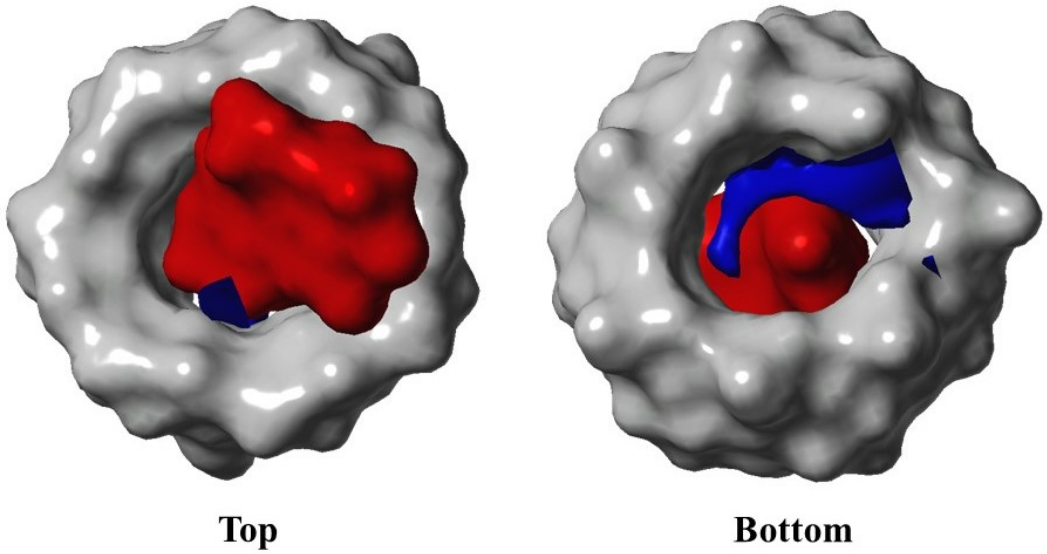


Figure 5. Results of a molecular docking simulation between Beta cyclodextrin and a mycotoxin. Molecules are depicted using occupancy volume. In red the volume

occupied by the mycotoxin, in grey the cyclodextrin volume, and in blue the empty volume that could be filled by water molecules (GRID analysis).

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.

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 and Dellafiora 2012) against estrogen receptor to understand if they can bind competitively with the endogenous ligand, estradiol. A molecular docking-based study demonstrates that it is possible to discriminate between *cis* and *trans* isomers for Zearalenone using the same docking approach: for the *cis* isomer, a stronger interaction has been predicted (Dellafiora 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 (MIE) and be predictive of adverse effects. The investigation of receptor-ligand 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-, 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.

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, Cruciani, Galaverna, & Cozzini, 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 potency may 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 of 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.

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, Mena, Cozzini, Brighenti, & Del Rio, 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 ER dependent diseases. Ones again the *in silico* approach to study the mechanism of action (MOA) suggested that hydroxylation can play an important role in the agonistic behavior of these derivates.

Printing Inks

As stated, another tumor marker is the androgen receptor, 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 (EHDAB) have 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 androgen receptor (AR). In fact, it is well known by *in vitro* analyses that this class of compounds are 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 computationaly 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, that are the three well known AR-

mediated endocrine disrupting compounds.

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.

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 (FCMs) (Cavaliere, Lorenzetti, & Cozzini, 2020). One of the most common bisphenols is bisphenol A (BPA) or 4,4′-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 (SVHC) 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 nuclear receptors interfering with the normal hormone response. Thus, a lot of efforts are 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 twenty-six different bisphenols (including seven 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.

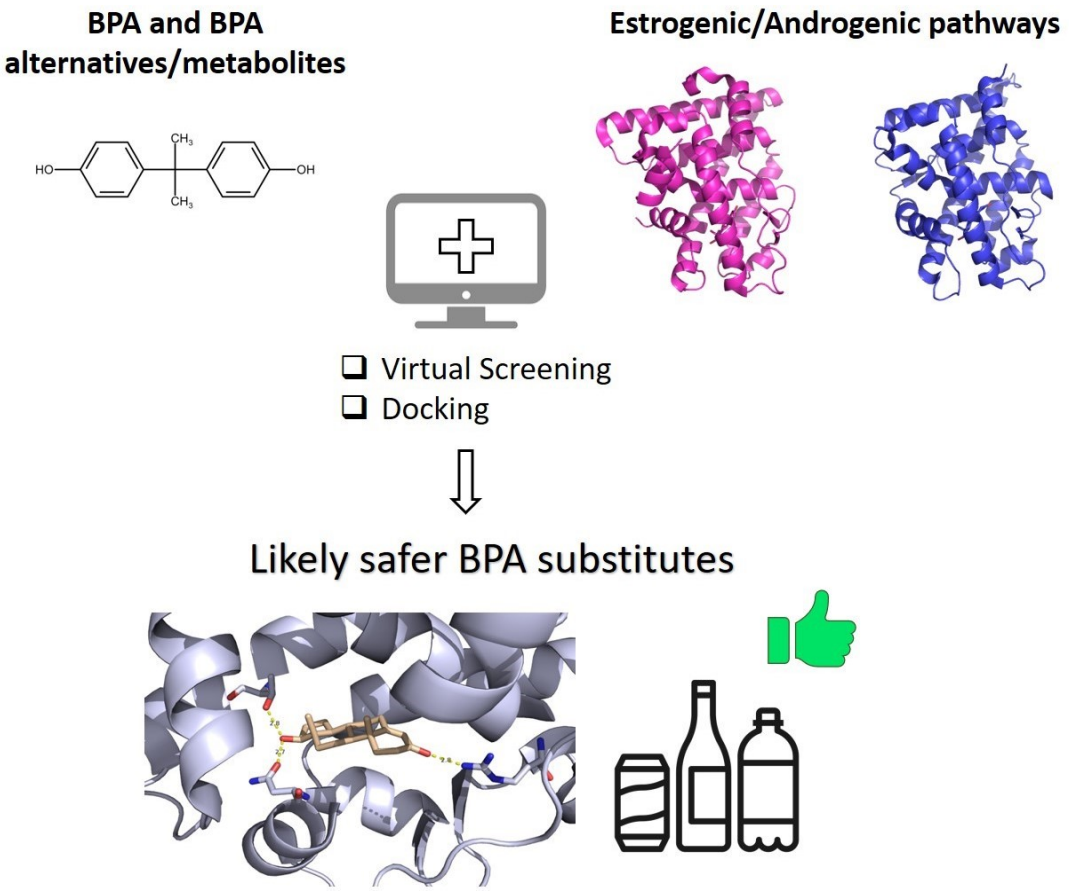


Figure 6. The case of docking/scoring application on food contact materials: Bisphenols.

Six different nuclear receptors (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 physico-chemical 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: i) some BPA metabolites could lower the harmful effects of BPA exposure; ii) Bisphenol S, a BPA' substitute, turned out a lower interactor for all NRs, except for androgen receptor (AR), for which its binding activity is found similar to a pharmacological anti-androgen; iii) only 2,2-Bis(4-hydroxyphenyl)propanol (BPAol), a BPA metabolite, was predicted as a lower interactor for all NRs considered.

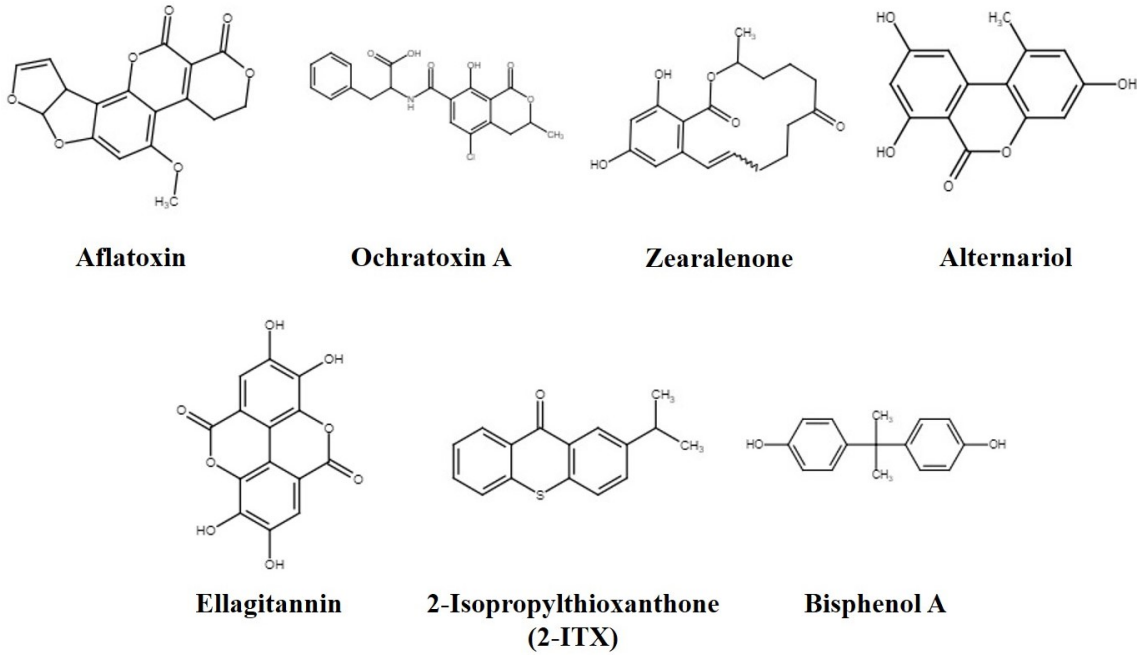


Figure 7. The chemical structres of case studies compounds.

Conclusions

The lesson learning from medicinal chemistry suggests we can use computational simulations in food safety, in particular molecular modelling 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: i) be careful with starting structural data (check the structural parameters); ii) be careful in choosing the software, there is not a general package able to solve all modeling problems; iii) a complete analysis should include a lot of factors: waters, protons, metals, cofactors…; iv) don't 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., … Rizzo, R. C. (2015). DOCK 6: Impact of new features and current docking performance. *Journal of Computational Chemistry*, *36*(15), 1132–1156. 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), 4585–4594. 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. *Chem. Res. Toxicol*, *12*, 52–63.

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), 1063–1067. 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*, 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 and Chemical Toxicology*, *137*. https://doi.org/10.1016/j.fct.2020.111116

Cozzini, P., & Dellafiora, L. (2012, October). *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*. Netherlands. 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), 3093–3118. 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), 2474–2494. 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*, 93–101. 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*, 65–71. 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 ellagitannin-derived metabolites. *in silico* tools to evaluate their potential xenoestrogenic behavior. *Food & Function*, *4*(10), 1442–1451. https://doi.org/10.1039/c3fo60117j

Dill, K. A. (1997). Additivity principles in biochemistry. *The Journal of Biological Chemistry*, *272*(2), 701–704. 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), 275–286.

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), 301–304. 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., … 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), 1739–1749. https://doi.org/10.1021/jm0306430

Ginex, T., Dall'Asta, C., & 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*(2), 279–289. 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), 792–798. 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), 499–511. 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), 1061–1066. 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). AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J Comput Chem*, *30*(16), 2785–2791. 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), 4953–4967. https://doi.org/10.1021/jm00026a001

Pagadala, N. S., Syed, K., & Tuszynski, J. (2017). Software for molecular docking: a review. *Biophysical Reviews*, *9*(2), 91–102. 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. 1–37). 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), 2844–2848. https://doi.org/10.1021/jz1010863

Schellhammer, I., & Rarey, M. (2004). FlexX-Scan: fast, structure-based virtual screening. *Proteins*, *57*(3), 504—517. 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), 723–732. 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. 601–684). https://doi.org/doi: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), 609–623. 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), 1422–1426. https://doi.org/10.1021/ci010025x

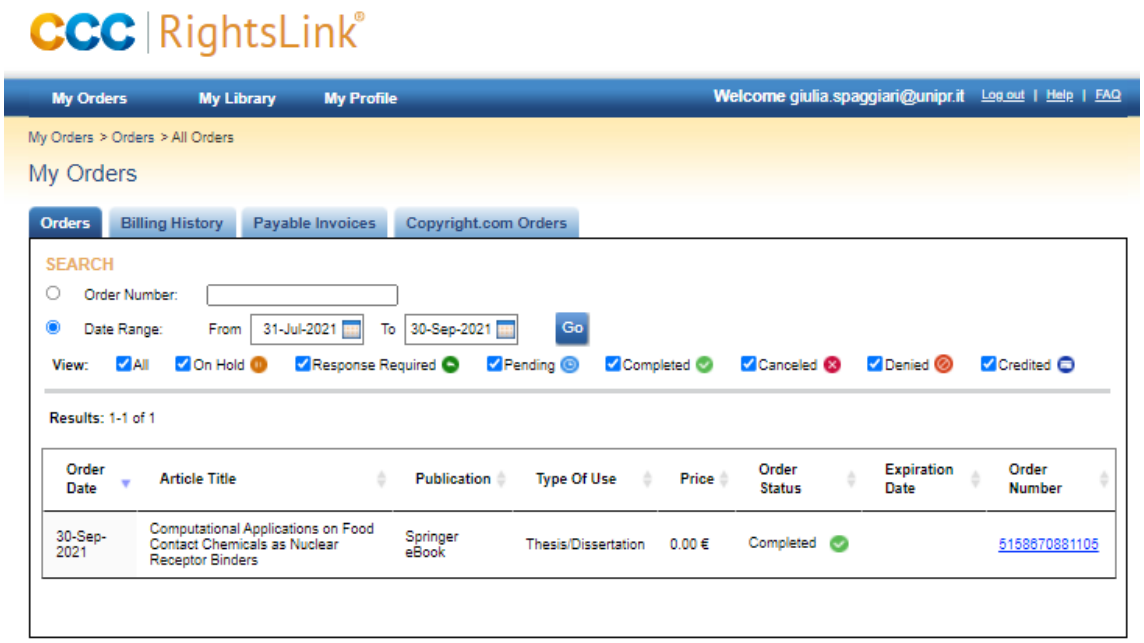
Williams, A. J., & Ekins, S. (2011). A quality alert and call for improved curation of public chemistry databases. *Drug Discovery Today*, *16*(17), 747–750. https://doi.org/https://doi.org/10.1016/j.drudis.2011.07.007

CHAPTER 4

Computational applications on Food Contact Chemicals as Nuclear Receptors binders

The content of this chapter has also been published in:

Cavaliere, F., Spaggiari, G., Cozzini, P., 2021. Computational Applications on Food Contact Chemicals as Nuclear Receptor Binders. In: Badr M.Z. (eds) Nuclear Receptors. Springer, Cham. https://doi.org/10.1007/978-3-030-78315-0_12



CCC | RightsLink®

My Orders | My Library | My Profile | Welcome giulia.spaggiari@unipr.it | Log out | Help | FAQ

My Orders > Orders > All Orders

My Orders

Orders | Billing History | Payable Invoices | Copyright.com Orders

SEARCH

Order Number:

Date Range: From 31-Jul-2021 To 30-Sep-2021 Go

View: All | On Hold | Response Required | Pending | Completed | Canceled | Denied | Credited

Results: 1-1 of 1

Order Date	Article Title	Publication	Type Of Use	Price	Order Status	Expiration Date	Order Number
30-Sep-2021	Computational Applications on Food Contact Chemicals as Nuclear Receptor Binders	Springer eBook	Thesis/Dissertation	0.00 €	Completed		5158670881105

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

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.

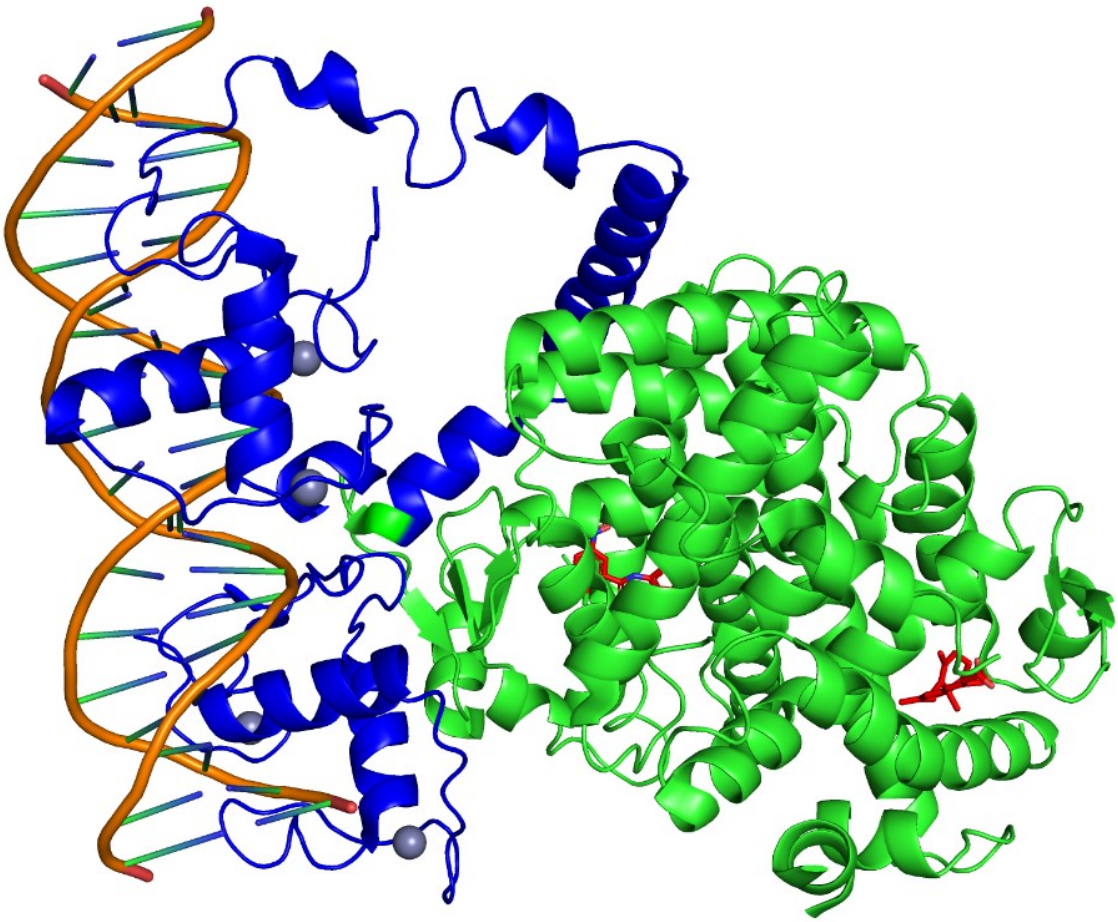


Figure 1. Representative nuclear receptors' structure (PDB ID: 3E00). The ligand-binding site is in green (in this case it is composed of two nuclear receptors, RXRα and PPARγ, and their, respectively, ligands, 9-cis-retinoic acid and 2-chloro-5-nitro-N phenylbenzamide, in red), while the DNA-binding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

The 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 β-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 α-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 (Cozzini et al., 2012).

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.

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 (Weikum et al., 2018).

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).

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.

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 (Hass et al., 2007). 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 (Hartung, 2008). 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.

Table 1. List of the 48 nuclear receptors with the respective diseases.

No.	Subfamily	Approved Name	Gene name	Crystallography PDB structure ([a])	NR diseases
1	NR0B1	Dosage-sensitive sex reversal, critical region on the X	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,

		chromosome			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, insulin resistance, cryptorchidism, 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

[a] In brackets are the number of the nuclear receptors' structures present in PDB until 2017

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).

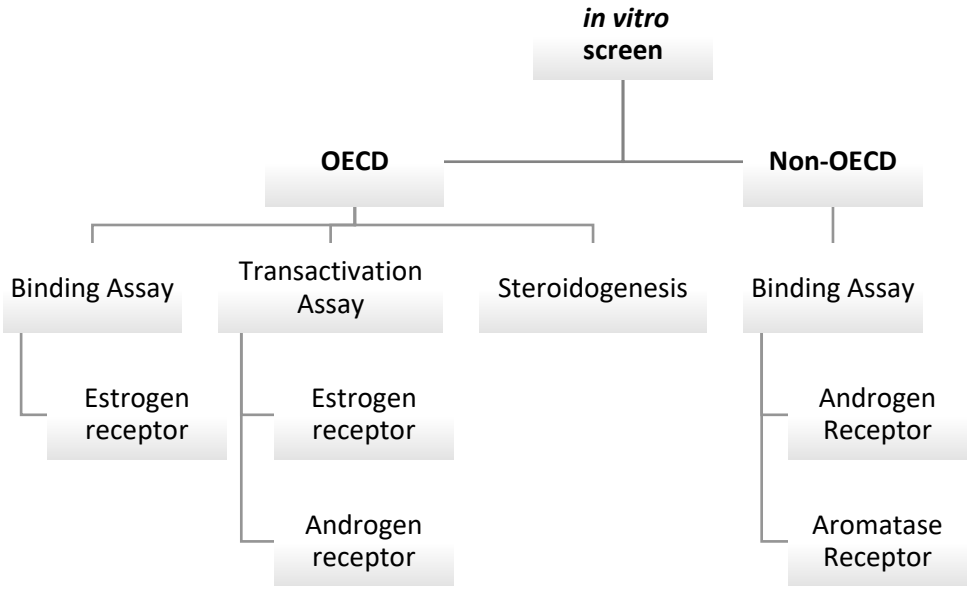


Figure 2. Flowchart of *in vitro* approaches.

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).

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		

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 nonbinders. 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 β1 (THRβ1) (Zhang et al., 2016). Of the 31 compounds predicted as potential (THRβ1) 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 (Balaguer et al., 2017). Thouennon and co-workers used the ITC to characterize the ability of some environmental chemical contaminants to bind oestrogen-related receptor γ (ERRγ) finding that bisphenol E was a more potent binder compared to bisphenol A (Thouennon et al., 2019). 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 Kd 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 (DeSantis et al., 2012).

Gene Reporter Assays

If the experiment's purpose is to distinguish against agonist and antagonist compounds, 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 β-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 a high 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 (Dellafiora et al., 2017). 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 oestrogenic activity showing that bisphenol A and bisphenol AF consistently can activate endogenous ER target genes (Li et al., 2013).

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 testosterone and estradiol. In brief, H295R 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 used to identify 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.

In silico Methods for Screening Endocrine Disruptor Compounds

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. 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).

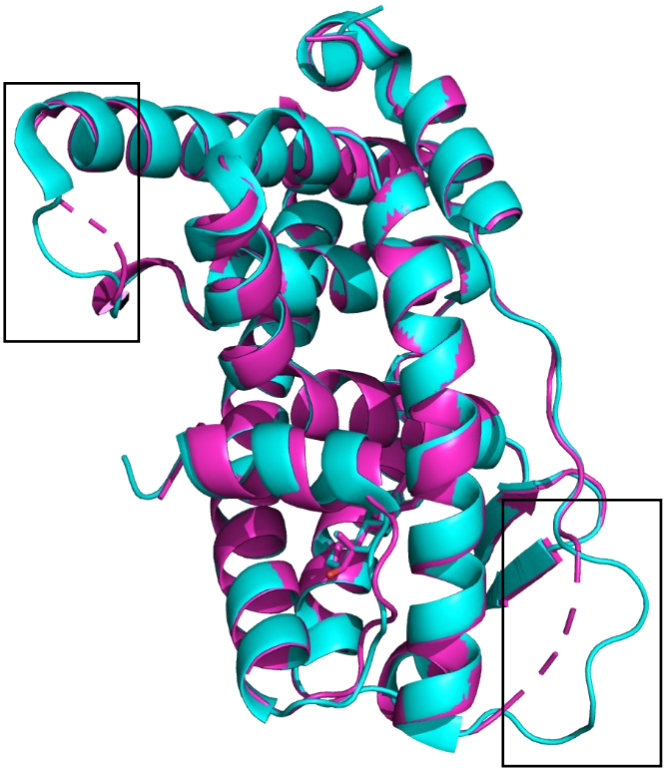


Figure 3. PDB structures of oestrogen receptor alpha with two different resolution values. *In blue* the protein (PDB ID: 2YJA) with a high-resolution value (1.82 Å resolution) and *in magenta* the protein (PDB ID: 1ERE) with a low-resolution value (3.10 Å resolution) are shown. The box highlights the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

The 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 (Cozzini et al., 2008). 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) (Vitkup et al., 2002). The latter provides both

the magnitudes and the directions of each atom shift, and, thus, it allows a dynamic description of the protein structure.

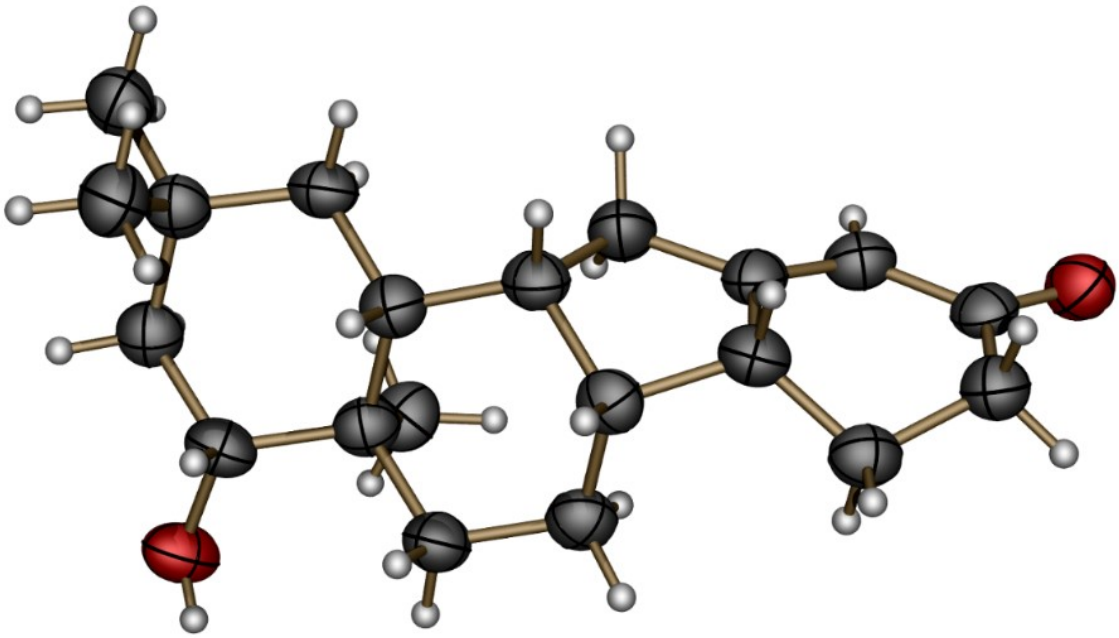


Figure 4. The epitestosterone shown above, made by the program ORTEP (Farrugia, 2012), illustrates the thermal ellipsoid.

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.

Ligand-Based Virtual Screening

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 (Davies et al., 2015) (Mendez et al., 2019) (Sterling and Irwin, 2015). 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 products 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. 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.

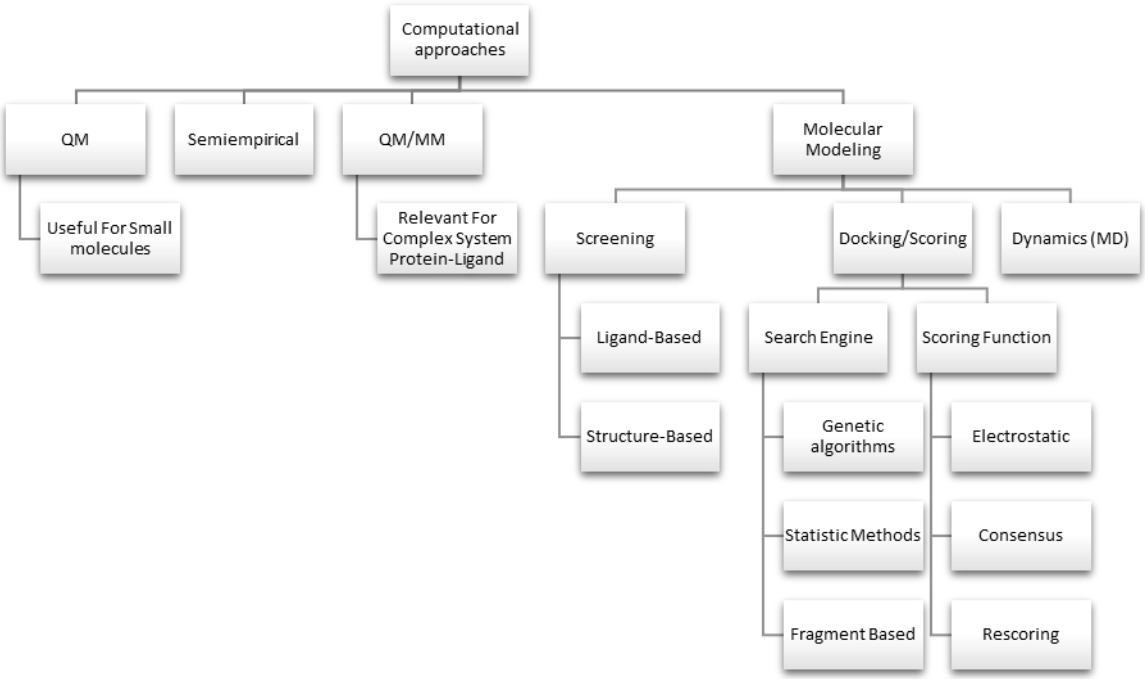


Figure 5. Flowchart of computational (or *in silico*) approaches.

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" (Wermuth et al., 1998). 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 (Rush et al., 2005) (Sastry et al., 2011). 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 (Barril, 2012).

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 (Cereto-Massagué et al., 2015) (Gimeno et al., 2019). 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 (Gimeno et al., 2019).

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. The MA based on molecular shape map 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.

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

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 (Wang and Wang, 2001). 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% (Bissantz et al., 2000).

Molecular Dynamics 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.

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 (Amadasi et al., 2009). 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 (Bampidis et al., 2020).

Kenda and co-workers conducted a screening of 1046 US-approved and marketed small-molecule drugs for estimating their endocrine-disrupting properties (Kenda and Dolenc, 2020). 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 (Cavaliere et al., 2020).

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

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. Chem. Res. Toxicol. 22, 52–63. https://doi.org/10.1021/tx800048m

Balaguer, P., Delfosse, V., Grimaldi, M., Bourguet, W., 2017. Structural and functional evidences for the interactions between nuclear hormone receptors and endocrine disruptors at low doses. Comptes Rendus - Biol. 340, 414–420. https://doi.org/10.1016/j.crvi.2017.08.002

Bampidis, V., Azimonti, G., Bastos, M. de L., Christensen, H., Dusemund, B., Durjava, M.K., Kouba, M., Lopez-Alonso, M., Puente, S.L., Marcon, F., Mayo, B., Pechova, A., Petkova, M., Ramos, F., Sanz, Y., Villa, R.E., Woutersen, R., Aquilina, G., Bories, G., Gropp, J., Nebbia, C., Innocenti, M.L., 2020. Safety and efficacy of propyl gallate for all animal species. EFSA J. 18, 1–16.

https://doi.org/10.2903/j.efsa.2020.6069

Barril, X., 2012. Physico-Chemical and Computational Approaches to Drug Discovery.

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, 111116. https://doi.org/10.1016/j.fct.2020.111116

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

Cozzini, P., Kellogg, G.E., Spyrakis, F., Abraham, D.J., Costantino, G., Emerson, A., Fanelli, F., Gohlke, H., Kuhn, L.A., Morris, G.M., others, 2008. Target Flexibility: J. Med. Chem. 51, 6237–6255. https://doi.org/10.1021/jm800562d.Target

Cozzini, P., Kellogg, G.E., Thurston, D.E., 2012. Computational Approaches to Nuclear Receptors, WorldCat.

Davies, M., Nowotka, M., Papadatos, G., Dedman, N., Gaulton, A., Atkinson, F., Bellis, L., Overington, J.P., 2015. ChEMBL web services: Streamlining access to drug discovery data and utilities. Nucleic Acids Res. 43, W612–W620. https://doi.org/10.1093/nar/gkv352

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. https://doi.org/10.1016/j.fct.2017.07.062

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

Gimeno, A., Ojeda-Montes, M.J., Tomás-Hernández, S., Cereto-Massagué, A., Beltrán-Debón, R., Mulero, M., Pujadas, G., Garcia-Vallvé, S., 2019. The light and dark sides of virtual screening: What is there to know? Int. J. Mol. Sci. 20. https://doi.org/10.3390/ijms20061375

Hartung, T., 2008. Toward a new toxicology - evolution or revolution? ATLA Altern. to Lab. Anim. 36, 635–639. https://doi.org/10.1177/026119290803600607

Hass, U., Scholze, M., Christiansen, S., Dalgaard, M., Vinggaard, A.M., Axelstad, M.,

Metzdorff, S.B., Kortenkamp, A., 2007. Combined exposure to anti-androgens exacerbates disruption of sexual differentiation in the rat. Environ. Health Perspect. 115, 122–128. https://doi.org/10.1289/ehp.9360

Kenda, M., Dolenc, M.S., 2020. Computational study of drugs targeting nuclear receptors. Molecules 25. https://doi.org/10.3390/molecules25071616

Li, Y., Luh, C.J., Burns, K.A., Arao, Y., Jiang, Z., Teng, C.T., Tice, R.R., Korach, K.S., 2013. Endocrine-disrupting chemicals (EDCs): In Vitro mechanism of estrogenic activation and differential effects on ER target genes. Environ. Health Perspect. 121, 459–466. https://doi.org/10.1289/ehp.1205951

Mendez, D., Gaulton, A., Bento, A.P., Chambers, J., De Veij, M., Félix, E., Magariños, M.P., Mosquera, J.F., Mutowo, P., Nowotka, M., Gordillo-Marañón, M., Hunter, F., Junco, L., Mugumbate, G., Rodriguez-Lopez, M., Atkinson, F., Bosc, N., Radoux, C.J., Segura-Cabrera, A., Hersey, A., Leach, A.R., 2019. ChEMBL: Towards direct deposition of bioassay data. Nucleic Acids Res. 47, D930–D940. https://doi.org/10.1093/nar/gky1075

Rush, T.S., Grant, J.A., Mosyak, L., Nicholls, A., 2005. A shape-based 3-D scaffold hopping method and its application to a bacterial protein-protein interaction. J. Med. Chem. 48, 1489–1495. https://doi.org/10.1021/jm040163o

Sastry, G.M., Dixon, S.L., Sherman, W., 2011. Rapid shape-based ligand alignment and virtual screening method based on atom/feature-pair similarities and volume overlap scoring. J. Chem. Inf. Model. 51, 2455–2466. https://doi.org/10.1021/ci2002704

Sterling, T., Irwin, J.J., 2015. ZINC 15 - Ligand Discovery for Everyone. J. Chem. Inf. Model. 55, 2324–2337. https://doi.org/10.1021/acs.jcim.5b00559

Thouennon, E., Delfosse, V., Bailly, R., Blanc, P., Boulahtouf, A., Grimaldi, M., Barducci, A., Bourguet, W., Balaguer, P., 2019. Insights into the activation mechanism of human estrogen-related receptor γ by environmental endocrine disruptors. Cell. Mol. Life Sci. 76, 4769–4781. https://doi.org/10.1007/s00018-019-03129-x

Vitkup, D., Ringe, D., Karplus, M., Petsko, G.A., 2002. Why protein R-factors are so large: A self-consistent analysis. Proteins Struct. Funct. Genet. 46, 345–354. https://doi.org/10.1002/prot.10035

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

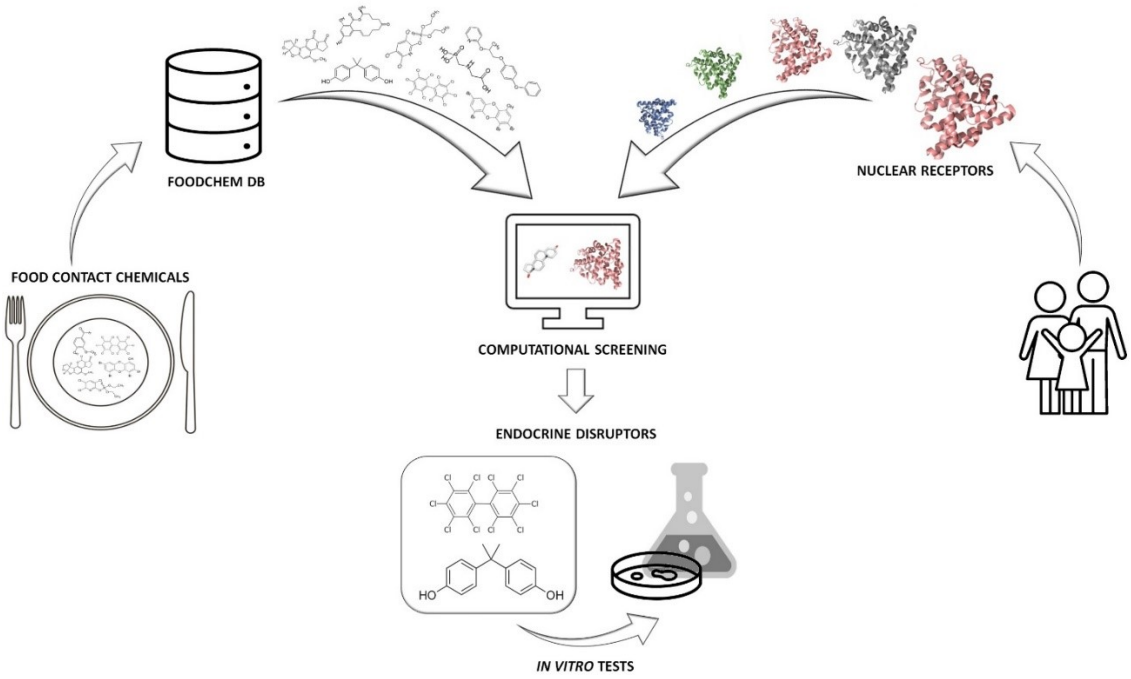
Wermuth, C.G., Ganellin, C.R., Lindberg, P., Mitscher, L. a., 1998. Glossary for Chemists of Terms Used in Medicinal Chemistry. Pure Appl. Chem. 70, 1129–1143. https://doi.org/10.1351/pac197951081725

Zhang, J., Li, Y., Gupta, A.A., Nam, K., Andersson, P.L., 2016. Identification and Molecular Interaction Studies of Thyroid Hormone Receptor Disruptors among Household Dust Contaminants. Chem. Res. Toxicol. 29, 1345–1354. https://doi.org/10.1021/acs.chemrestox.6b00171

CHAPTER 5

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

The content of this chapter has also been published in:
Cozzini, P., Cavaliere, F., Spaggiari, G., Morelli, G., Riani, M., 2021. Computational methods on food contact chemicals: Big data and in silico screening on nuclear receptors family. Chemosphere 292. https://doi.org/https://doi.org/10.1016/j.chemosphere.2021.133422


FOODCHEM DB
NUCLEAR RECEPTORS
FOOD CONTACT CHEMICALS
COMPUTATIONAL SCREENING
ENDOCRINE DISRUPTORS
HO
OH
IN VITRO TESTS

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.

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.

Keywords

Computational Chemistry · Consensus Prediction · Database · Nuclear Receptors · Toxicology

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 (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/6J 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. 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.

Material and Methods

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).

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").

Database descriptors

The foodchem DB stores 27 different fields that can be divided in three 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.

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

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.

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).

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).

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.

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 (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:

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

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.

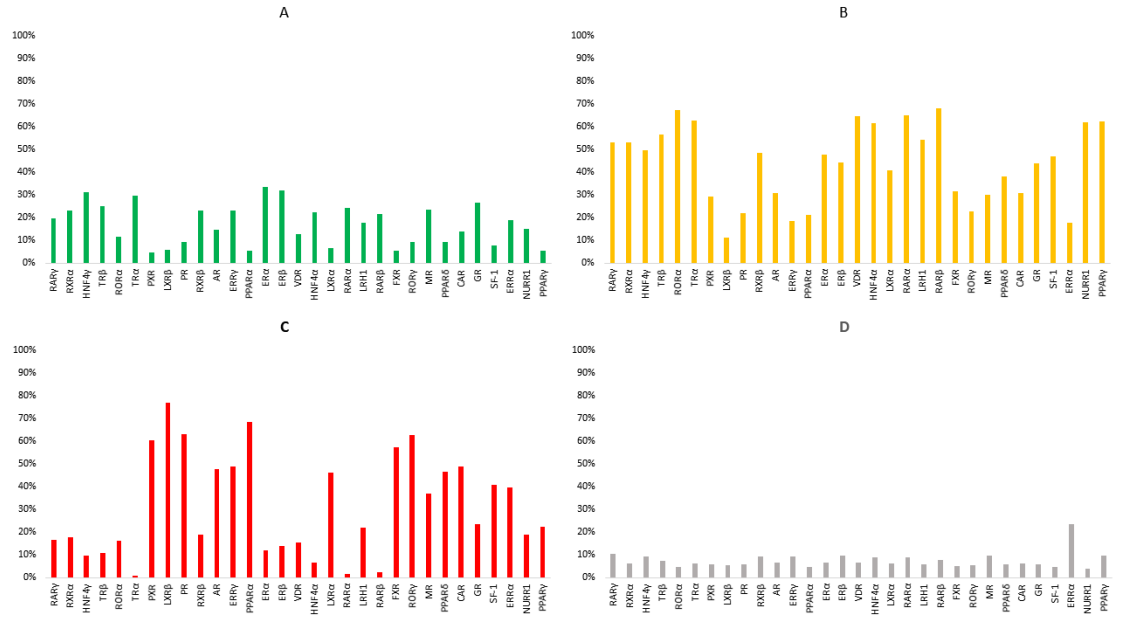


Figure 1. Results obtained from the robust multivariate statistical procedure. The 31 NRs are on the x-axis, while the number of the molecules (%) is on the ordinate. The molecules with a score smaller than 0.3 are highlighted in green (A), the molecules with a score between 0.3 and 0.8 are highlighted 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).

Figure 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.

Considering Figure 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 Figure 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.

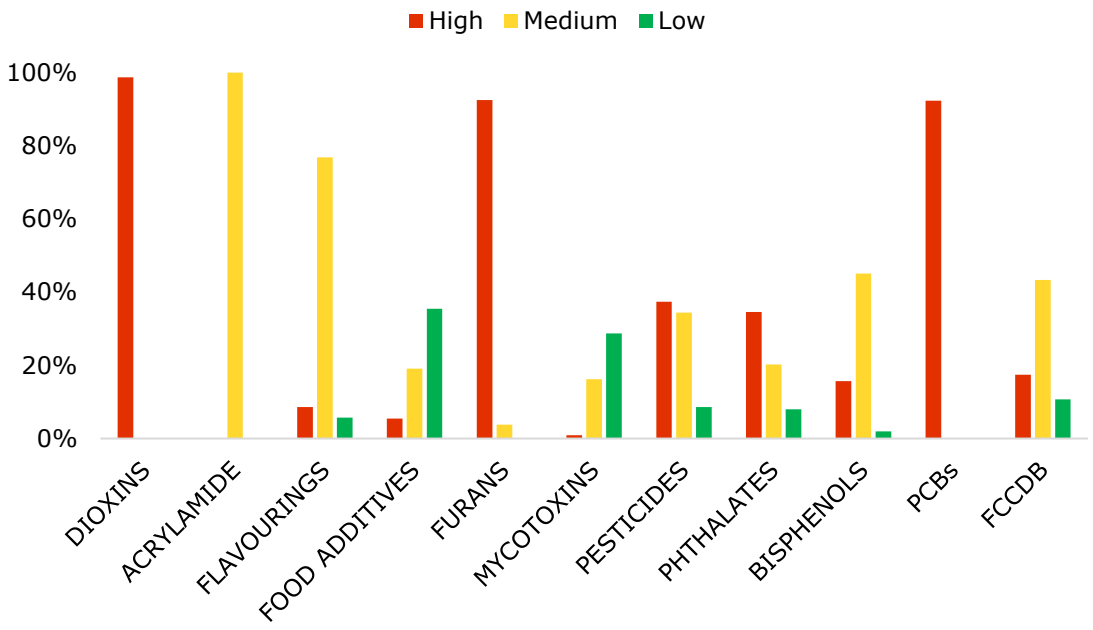


Figure 2. The percentage of molecules able to bind more than 15 nuclear receptors with high (≥0.8), medium (0.3-0.8), and low binding affinity (<0.3) considering each class of food contact chemicals.

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.

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.

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/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

Leist, M., Hartung, T., 2013. Inflammatory findings on species extrapolations: humans are definitely no 70-kg mice. Arch. Toxicol. 87, 563–567. https://doi.org/10.1007/s00204-013-1038-0

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

Luechtefeld, T., Maertens, A., Russo, D.P., Rovida, C., Zhu, H., Hartung, T., 2016. Analysis of public oral toxicity data from REACH registrations 2008-2014. ALTEX 33, 111–122. https://doi.org/10.14573/altex.1510054

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.

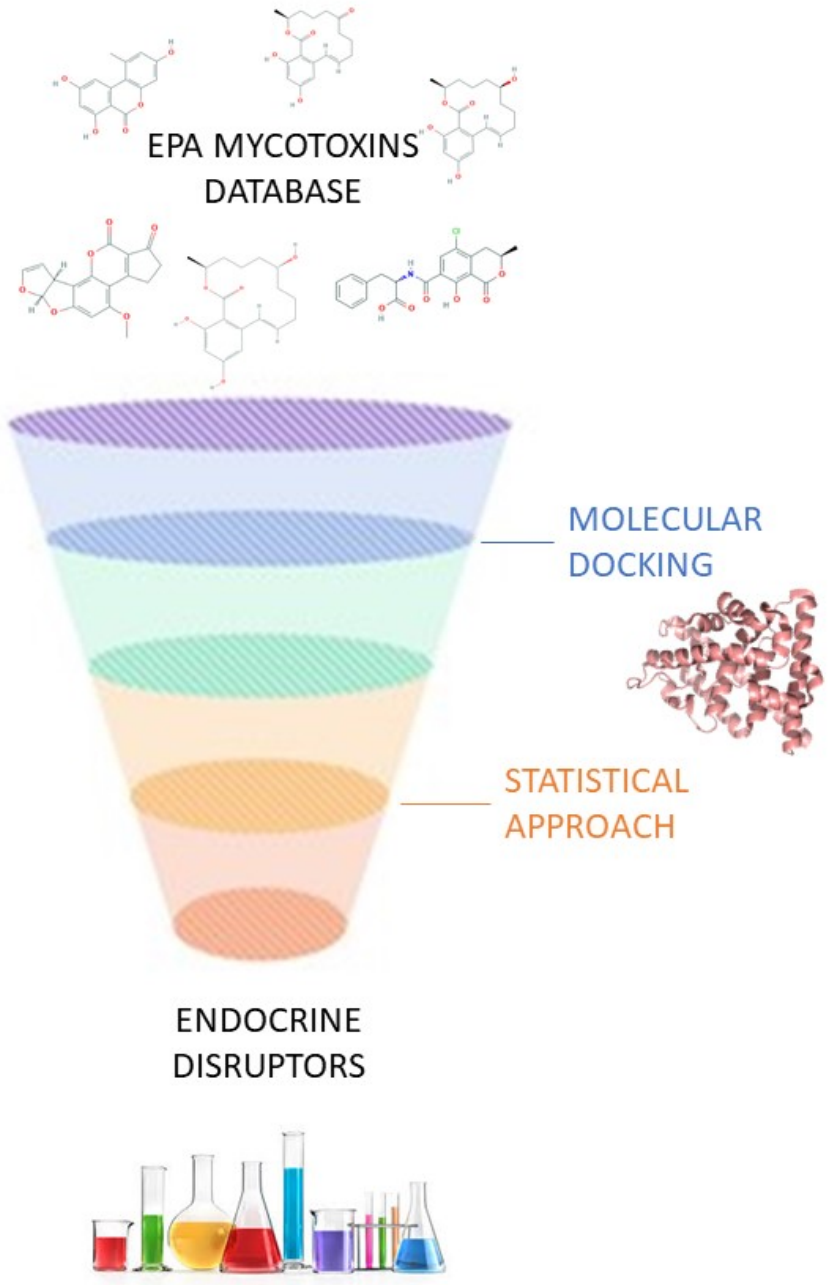
Wang, R., Lu, Y., Wang, S., 2003. Comparative evaluation of 11 scoring functions for molecular docking. J. Med. Chem. 46, 2287–2303. https://doi.org/10.1021/jm0203783

CHAPTER 6

A synergism of *in silico* and statistical approaches to discover new potential endocrine disruptor mycotoxins

The content of this chapter has also been published in:

Spaggiari, G., Morelli, G., Riani, M., Cozzini, P., 2022. A synergism of *in silico* and statistical approaches to discover new potential endocrine disruptor mycotoxins. Toxicol. Appl. Pharmacol. 435, 115832. https://doi.org/10.1016/j.taap.2021.115832


EPA MYCOTOXINS
DATABASE
MOLECULAR
DOCKING
STATISTICAL
APPROACH
ENDOCRINE
DISRUPTORS

Abstract

Mycotoxins are secondary metabolites produced by pathogenic fungi. They are found in a variety of different products, such as spices, cocoa, and cereals, and they can contaminate fields before and/or after harvest and during storage. Mycotoxins negatively impact human and animal health, causing a variety of adverse effects, ranging from acute poisoning to long-term effects. Given a large number of mycotoxins (currently more than 300 are known), it is impossible to use *in vitro*/*in vivo* methods to detect the potentially harmful effects to human health of all of these. To overcome this problem, this work aims to present a new robust computational approach, based on a combination of *in silico* and statistical methods, in order to screen a large number of molecules against the nuclear receptor family in a cost and time-effective manner and to discover the potential endocrine disruptor activity of mycotoxins. The results show that a high number of mycotoxins is predicted as a potential binder of nuclear receptors. In particular, ochratoxin A, zearalenone, α- and β-zearalenol, aflatoxin B1, and alternariol have been shown to be putative endocrine disruptors chemicals for nuclear receptors.

Keywords

Endocrine disruptors · Mycotoxins · Nuclear receptors · Molecular docking · Dimension reduction

Highlights

- A new integrated *in silico* and statistical approach useful to discover the potential endocrine disruptor activity of mycotoxins.
- Endocrine disruptor prediction using the molecular docking method saves time and cost.
- Molecular docking is useful to identify possible dangerous food contact chemicals.

Introduction

Mycotoxins are toxic secondary metabolites produced by fungi belonging, essentially, to *Aspergillus*, *Penicillium*, *Alternaria*, and *Fusarium* genera. They can grow on a variety of different crops, such as cereals, oilseeds, cocoa, nuts, and they can occur before or after harvest, during storage, on/in the food itself often under warm and damp conditions. The Food and Agriculture Organization (FAO) estimated that 25% of global agricultural products are contaminated by mycotoxins each year (Boutrif and Canet, 1998). Currently, more than 300 mycotoxins are known, but ochratoxin A (OTA), zearalenone (ZEN) and its two derivates, α-zearalenol (αZEL) and β-zearalenol (βZEL), aflatoxins (AFs), in particular aflatoxin B1 (AFB1), fumonisins (FBs), especially fumonisin B1 (FBB1), deoxynivalenol (DON), and patulin are the most studied and considered the most toxigenic to agriculture, animal and human health causing acute and chronic diseases, such as cancer induction, carcinogenesis, genotoxicity, mutagenicity, kidney toxicity, nervous disorders, and death (Bennett and Moore, 2019). In 1993, the WHO-International Agency for Research on Cancer (WHO-IARC) classified some mycotoxins into three different groups based on their carcinogenic potential: i) aflatoxins were classified as carcinogenic to humans (Group 1); ii) ochratoxins and fumonisins were classified as possible carcinogens (Group 2B); iii) Trichothecenes and zearalenone were not classified as human carcinogens (Group 3) (World Health Organization International Agency for Research on Cancer (WHO-IARC), 1993a) (World Health Organization International Agency for Research on Cancer (WHO-IARC), 1993). The literature shows that several mycotoxins can act as potential endocrine disruptors at the level of nuclear receptors (NRs) signaling. Among mycotoxins, aflatoxins and fumonisins exposure are of major concern in developing countries, such as Kenya, India, and Malaysia, where the level of food contamination is not sufficiently monitored and where in recent years repetitive aflatoxins outbreaks have occurred (Lewis et al., 2005). It is reported that Aflatoxin B1 is a primary cause of human liver cancer and, in particular, it is dangerous in populations with a high rate of hepatitis B virus (HBV) (Do et al., 2020) (Liu et al., 2017). It was demonstrated that aflatoxin B1 could activate several nuclear receptors, such as the pregnane X receptor (PXR), constitutive androstane receptor (CAR), and aryl hydrocarbon receptor (AhR) (Ates and Ortatatli, 2021) (Ayed-Boussema et al., 2012a, Ayed-Boussema et al., 2012b). Human aflatoxins exposure is often associated with fumonisins, which are possible carcinogen compounds in humans and that are implicated in the high incidence of neural tube defects (International Agency for Research on Cancer

(IARC), 2002) (Marasas et al., 2004). Among the mycotoxins, ochratoxin A induces several toxic health effects, such as nephrotoxicity, hepatotoxicity, genotoxicity, teratogenicity, immunotoxicity, and renal disease (Kőszegi and Poór, 2016) (Travis and Bui-Klimke, 2015). Patulin (PAT) causes various acute and chronic health effects due to the high liver, gastrointestinal tract, kidneys, nervous system, and immune system toxicity (Ahmadi et al., 2019) (Altunay et al., 2019) (Guo et al., 2017). Patulin also acts at the level of nuclear receptors. Especially, it increases the transcriptional activity of glucocorticoid receptor (GR), the estradiol levels, and the production of progesterone and decreases the production of testosterone (Frizzell et al., 2014). Furthermore, it appears that PAT can activate PXR and/or CAR and AhR (Ayed-Boussema et al., 2012a) (Ayed-Boussema et al., 2012b). Different studies demonstrated that zearalenone and its hydroxylate metabolites induced adverse effects on intestinal microflora and reproductive system (Dellafiora et al., 2020) (Tan et al., 2020) (Wang et al., 2018). It is reported that zearalenone and its metabolites have interaction with estrogen receptors (ERs) and probably with other nuclear receptors, such as liver X receptor (LXR), PXR, and progesterone receptor (PR) (Frizzell et al., 2011) (Molina-Molina et al., 2014) (Prouillac et al., 2012). Various *in vitro* and *in vivo* studies demonstrated the biological toxicity of deoxynivalenol, called vomitoxin, involved in digestive disorders, reproductive and endocrine disruptions (Akbari et al., 2014) (Bertero et al., 2018). Ndossi and co-workers suggested that trichothecenes deoxynivalenol (DON), T-2 and HT-2 toxins do not interact directly with steroid hormone receptors, but they may potentially act as endocrine disruptors causing effects on steroidogenesis and alterations in gene expression (Ndossi et al., 2012). Frizzell and colleagues established that ochratoxin A (OTA) can affect the endocrine system by modulating hormone production provoking adverse effects on development and reproduction (Frizzell et al., 2013a) (Frizzell et al., 2013b). According to several studies, alternariol (ALT) is related to DNA strand breaking activity and to oesophageal cancer, and it exhibited a weak oestrogenic activity but increased estradiol and progesterone production (Frizzell et al., 2013a) (Frizzell et al., 2013b) (Lehmann et al., 2006) (Pfeiffer et al., 2007). The information reported above illustrates that mycotoxins are capable of causing a variety of adverse effects on human and animal health, but, at the same time, only a limited set of mycotoxins is studied until now. Despite of intensive research over decades, there is still a lot to understand about effects and mechanism of action of mycotoxins, mainly in consideration of their possible endocrine disruption activity. Given the relevance of the problem and given a large number of

compounds potentially harmful to human health, a targeted strategy is necessary in order to deal with the rapid identification of the possible endocrine disruptor molecules. Then, the application of an integrated *in silico* and statistical approach allows us to have a consensus prediction and to increase the speed in the analysis of mycotoxins for the identification of potential endocrine disruptor compounds. In order to have more reliable results, the molecular docking procedure was carried out using two molecular docking software and four different scoring functions. The combination of more docking programs and scoring functions allows to reduce the number of false-positive and to obtain more reliable results (Bissantz et al., 2000) (Charifson et al., 1999) (Teramoto and Fukunishi, 2007). In this present work, 25 different nuclear receptors and a set of 328 mycotoxins were considered in order to discover new putative endocrine disruptor compounds and to decipher their mechanism of binding (Amadasi et al., 2009). In our thought, the approach presented in this work could be a useful tool to preliminarily evaluate the endocrine disruptor activity of a very broad spectrum of compounds

Materials and methods

Preparation of proteins

The structures of the 25 nuclear receptors of Homo sapiens were downloaded from the Protein Data Bank (PDB) (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. However, for the docking with AutoDock (*see below*), the receptors were further 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.

Preparation of ligands

The 328 mycotoxins structures (listed in Supplementary material, Table 1) were downloaded from the United States Environmental Protection Agency (EPA) database (https://comptox.epa.gov/dashboard/chemical_lists/MYCOTOX2). Structural coordinates of the endogenous and putative ligands were retrieved from the NCBI PubChem compound database (https://pubchem.ncbi.nlm.nih.gov/). Software FLAP (Fingerprint for Ligand and Protein) was used in order to assign the correct protonation state to the ligands (pH = 7.4).

Molecular docking with GOLD

The GOLD software v5.8.1 (CCDC; Cambridge, UK; www.ccd.cam.ac.uk) was applied to dock ligands into the binding site of the 25 nuclear receptors. For each compound and receptor, 30 binding poses were generated. 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. For the genetic algorithm run, a maximum number of 100,000 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 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. After that, all the poses generated by GOLD were rescored using the scoring functions Chem Score and Hint Score (HINT, Hydropathic INTeraction). The coupling of these three scoring functions was chosen as: i) Gold Score that allows taking into account different factors such as H-bonding energy, van der Waals energy, metal interaction, and ligand torsion strain; ii) Chem Score that represents the total free energy charge and takes account of hydrophobic contact area, hydrogen bonding, ligand flexibility, and metal interaction; iii) Hint Score that provides a quantitative evaluation of protein-ligand interaction and takes 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 (Eldridge et al., 1997) (Eugene Kellogg and Abraham, 2000).

Molecular docking with Autodock

Molecular docking experiments were performed with Autodock Vina 1.1.2 using default settings (Trott & 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.

Results and discussion

Molecular docking procedure

A total of 328 mycotoxins were analysed using a molecular docking procedure to identify the capability to bind the 25 nuclear receptors. For each molecule, four scoring values

(listed in Supplementary material, sheet name “Docking Results” in the file *.xlsx*) were obtained and their binding affinity has been scored in comparison to the respective endogenous or non-binder ligand, used as a reference compound. The nuclear receptors family is composed by 48 members. Many of these remain “orphans” because their endogenous ligands are yet to be determined. For this reason, if the endogenous ligand is known, we use its value for the following molecular docking procedure; on the other hand, we consider all the ligands present in PDB that are probably endogenous ligands (listed in Supplementary material, Table 2). For each endogenous or putative ligand, four scoring values were obtained and for each receptor, a threshold was selected on the basis of our previous considerations about the binding between natural ligand and nuclear receptor. 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.

Data pre-processing

To give an example of the analysed data, Fig. 1 shows the scatter plot matrix (with the univariate boxplots on the main diagonal) of the four scoring values obtained from the molecular docking of estrogen-related receptor gamma (ERRγ) (PDB ID 2E2R). Data distribution is non-normal and highly asymmetric.

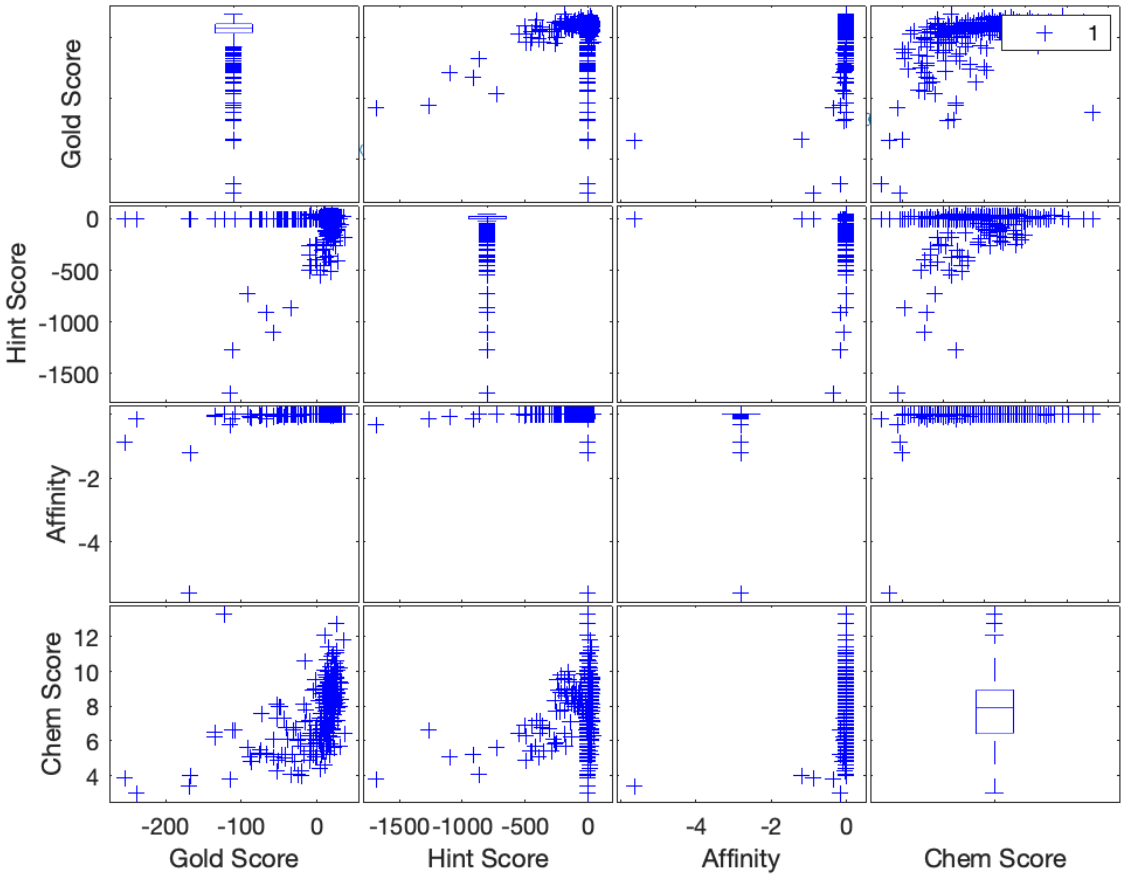


Figure 1. Scatter plot matrix of the four scoring values obtained from the molecular docking for the nuclear receptor ERRγ (PDB ID 2E2R).

Moreover, the figure shows the potential presence of several outliers.

It is now widely recognized in the statistical literature and many applications, that the presence of atypical values can affect the statistical analysis results. The problem is particularly acute when the number of such observations is so large that they "mask" each other, rendering traditional outlier detection techniques totally unreliable. For example, the distribution of the data shown in Fig. 1 clearly shows the need of identifying these observations. In this work, the robust method based on the forward search is used to detect these atypical observations given its dissemination and its validity internationally accepted (Riani et al., 2009). The univariate data distribution of the nuclear receptor ERRγ (PDB ID 2E2R) of the four scoring values before and after the outlier detection cleaning procedure is shown in the top and bottom row of Fig. 2 (note that Affinity score is shown using the absolute value).

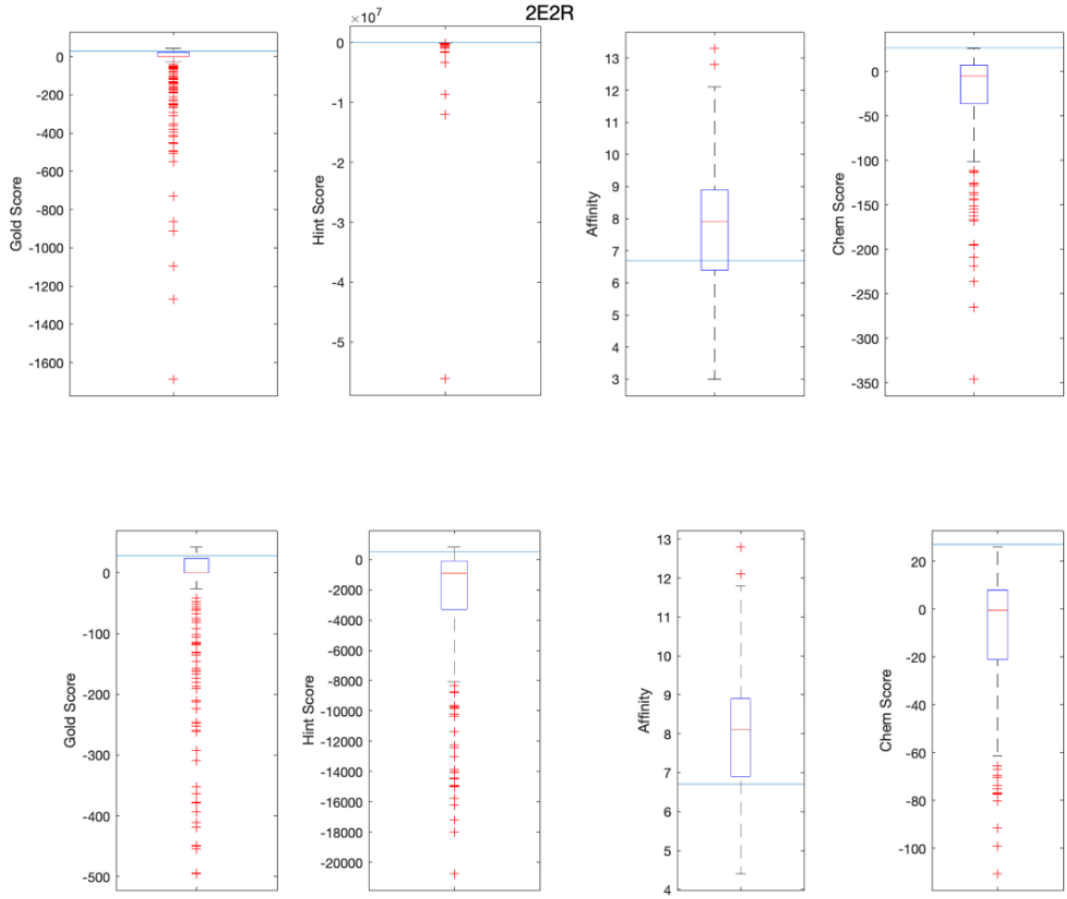


Figure 2. Univariate data distribution of the four scoring values before (top row) and after (bottom row) the outlier detection cleaning procedure. The blue horizontal line in each subplot is associated with the corresponding putative ligand for the nuclear receptor ERRγ (PDB ID 2E2R). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The blue horizontal line in each panel is associated with the corresponding putative ligand of each nuclear receptor.

It is clear that the data distribution after the multivariate outlier removal is much more symmetric and normally distributed. The confidence level used for outlier detection is 1% simultaneous to remove just the values that are very far from the bulk of the data. This multivariate outlier detection procedure is very different from the typical outlier detection procedure based on every variable independently. It is also interesting to see that while for Gold Score the line of the natural ligand lies beyond the upper whisker of the corresponding boxplot, for Affinity the line of the natural ligand is very close to the first quartile. This implies that while for Gold Score the strength of the empirical ligands is

much smaller than that of the natural ligand, for Affinity in more than 50% of the cases we found a strength of the empirical ligands greater than that of the corresponding natural ligand. It is necessary therefore to reach a unique combined score in an interpretable scale (that is to normalize molecular docking values in the 0–1 interval where the natural ligands are all equal to 1) which can give a final global evaluation based on the natural ligands (listed in Supplementary material, sheet name "Statistical Results" in the file *.xlsx*).

The normalization of molecular docking results using statistical analysis

After removing the outliers we rescaled the distribution of each mycotoxin in the domain [0 1] and we set the observation with a score larger than the natural ligand equal to 1. The purpose of the analysis was to combine the four scores into a single number which represents how closely the linear combination is to the natural ligand. The methodology which is used to achieve this goal is based on the technique of the principal components. In order to take into account the different degrees of dispersion of the four rescaled variables $X_1, \dots, X_4$ we have standardized them in order to obtain four new variables $Z_1, \dots, Z_4$ with 0 mean and variance equal to 1. At this point we have identified a weight w_j with $j = 1, \dots, 4$ such that, $w_j \geq 0$ and $\sum_{j=1}^{4} w_j^2 = 1$ in such a way that the explained variance of the original variables is as large as possible.

The final scores for the *i*-th mycotoxin 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 mycotoxins.

To simplify the results we have distinguished three cases based on the score: i) the mycotoxins 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; ii) the mycotoxins with a score between 0.3 and 0.8 are considered medium interactors compared to the respective natural ligands; iii) the mycotoxins with a score between 0.8 and 1.0 are judged as high interactors since they are able to bind the corresponding nuclear receptor with a binding affinity greater (more than 80%) than the respective natural ligand. Thus, all mycotoxins falling in this latter class

may be considered as substances of very high concern and should be the are illustrated in Fig. 3.

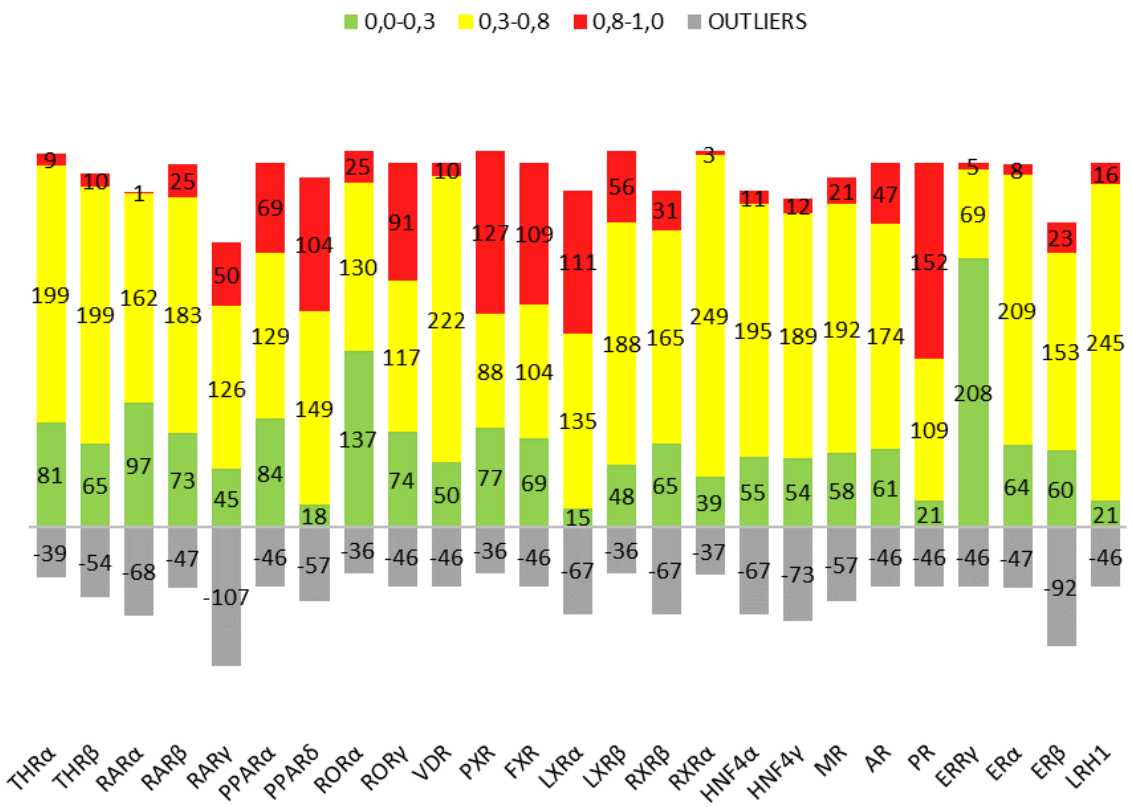


Figure 3. The results obtained from the application of the procedure. The 25 nuclear receptors are on the x-axis, while the mycotoxins number are on the ordinate. The mycotoxins with a score smaller than 0.3 are highlighted *in green*, the mycotoxins with a score between 0.3 and 0.8 are highlighted *in yellow*, the mycotoxins with a score greater than 0.8 are highlighted *in red*, while the outliers are highlighted *in grey*. In all ranges are showed the number of mycotoxins that bind each nuclear receptor.

As shown in Fig. 3, the majority of the mycotoxins are predicted as a medium or a good binder for the 25 nuclear receptors. In particular, we can see that more than 50% of the mycotoxins have been predicted as good interactors for all the 25 nuclear receptors excepted for ERRγ. If we focus on the single nuclear receptors, we can underline that more than one hundred mycotoxins are good interactors of PXR, FXR, PPARδ, LXRα, and PR. This statement is also supported by a few *in silico* and *in vitro* studies that suggest the endocrine disruptors' activity against these nuclear receptors (Ayed-Boussema et al., 2012a) (Ayed-Boussema et al., 2012b) (Prouillac et al., 2012).

In particular we focus our attention on the most studied and considered the most toxigenic mycotoxins: ochratoxin A, zearalenone, α- and β-zearalenol, aflatoxin B1, and alternariol (Fig. 4).

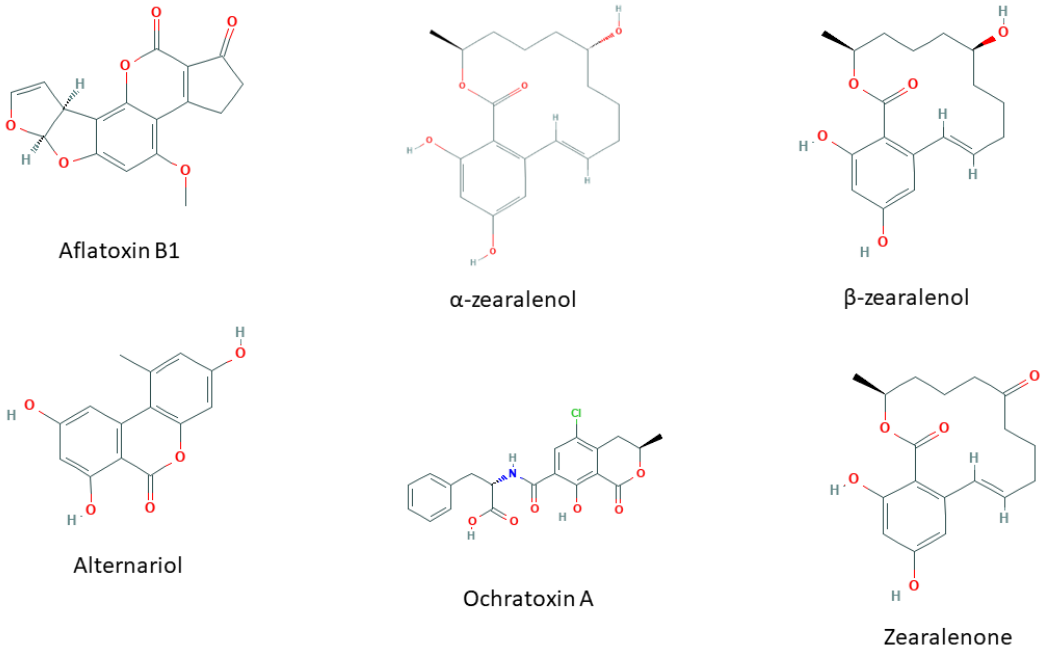


Figure 4. Six of the most toxigenic mycotoxins: ochratoxin A, zearalenone, α-zearalenol, β-zearalenol, aflatoxin B1, and alternariol.

We counted the number of nuclear receptors that each most toxigenic mycotoxin binds with high (≥0.8), medium (0.3–0.8), and low (≤0.3) binding affinity. As we can see in Fig. 5, ochratoxin A is the mycotoxin that can interact with eleven nuclear receptors with high binding affinity, like more than 0.8. On otherwise, α-zearalenol, β-zearalenol, and zearalenone bind only one nuclear receptor (liver X receptor β in the case of β-zearalenol and zearalenone, and estrogen receptor β in the case of α-zearalenol) with high binding affinity.

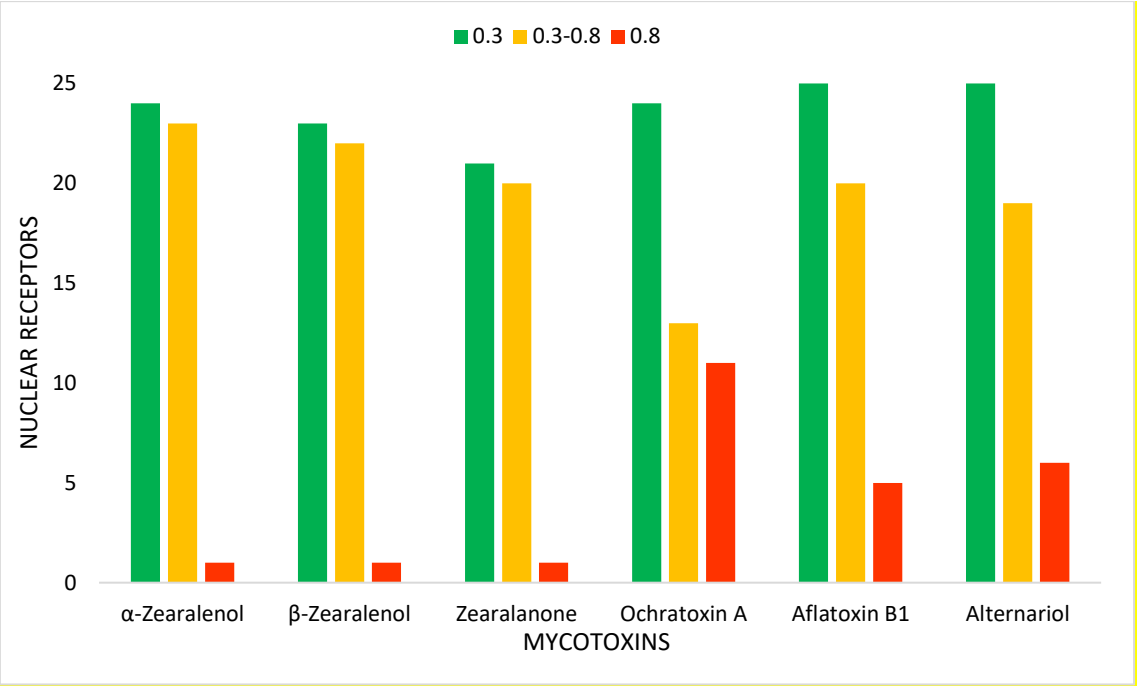


Figure 5. The number of nuclear receptors that each most toxigenic mycotoxin binds with high (≥0.8), medium (0.3–0.8), and low (≤0.3) binding affinity.

The impact of this results highlights the potential capability of these molecules to cause a very broad endocrine effects on the human health. Considering the medium interactors, all these most toxigenic mycotoxins are able to interact with more than 13 nuclear receptors.

Ochratoxin A

Ochratoxin A, a toxin produced by different Aspergillus and Penicillium species, is one of the most abundant in food and it has been shown to be nephrotoxic, immunosuppressive, and carcinogenic (Frizzell et al., 2013a) (Frizzell et al., 2013b). Our results show that OTA is a medium ligand for 21 nuclear receptors (its binding affinity is greater than 50%), but it is a good interactor for 11 nuclear receptors with a binding affinity greater than 80%. From the results, OTA shows a binding affinity greater than 90% for PXR (100%), LXRβ (97%), PPARα (95%), VDR (90%), LRH1 (92%), and PPARδ (99%). From literature, Lee and co-workers suggest that PXR is involved in modulating kidney damage caused by this toxin, while Shen and co-workers highlight that OTA and its derivates downregulate the expression of PXR (Lee et al., 2018) (Shen et al., 2020). Moreover, OTA probably interferes with VDR, in addition to PXR (Doricakova and Vrzal, 2015).

Zearalenone

Zearalenone, also known as F-2 mycotoxin, is a potent estrogenic metabolite produced by various Fusarium species. Zearalenone is known as estrogenic disrupters (ERs) causing several adverse effects on the reproductive system (Dellafiora et al., 2020) (Frizzell et al., 2011). Due to its recognized role as EDs, we can consider ZEN as a reference compound to detect other possible mycotoxins disruptors (67% of binding affinity for ERα and 73% for ERβ). Based on the results, ZEN was predicted also as a high interactor for RARγ (79%), LXRβ (90%), confirmed by Prouillac and co-workers too, FXR (85%), and RORγ (80%) (Prouillac et al., 2012).

α-zearalenol and β-zearalenol

α- and β-zearalenol, the major metabolites of zearalenone, reduce via intestinal and hepatic metabolism after intake showing similar estrogenic properties of ZEN. Our results are in agreement with *in silico* and *in vitro* studies confirming their estrogenic activity and the highest affinity of α-zearalenol (80% of binding affinity for ERα and 68% for ERβ) for ERs (Cozzini and Dellafiora, 2012) (Frizzell et al., 2011). Moreover, they were predicted as a good binder for RARγ (71%), PXR (72% only α-zearalenol), and LXRβ (90% only β-zearalenol).

Aflatoxin B1

Aflatoxin B1 is considered the most toxic aflatoxin and, besides its cancerogenic and genotoxicity activity, it has been reported to relate to several nuclear receptors, such as PXR, CAR, and AhR (Ayed-Boussema et al., 2012a) (Ayed-Boussema et al., 2012b).
Our results confirm the high affinity of this aflatoxin for PXR (70%), confirmed by Ayed-Boussema and colleagues too, but also for PR (83%), RXRβ (90%), AR (82%), and RORγ (86%).

Alternariol

Alternariol is a mycotoxin commonly produced by *Alternaria alternata* on a wide range of foods. A few studies have been performed to suggest that AOH may act as an endocrine disruptor in various ways, in particular acting with ERs, AR, and PR, how Frizzell and co-workers have been demonstrated in their study (Frizzell et al., 2013a) (Frizzell et al., 2013b). Our results show a binding affinity greater than 50% for all the 25 nuclear receptors, especially for LXRβ (95%), PPARα (97%), and FXR (93%).

Conclusions

Mycotoxins are toxic compounds that are naturally produced by different types of fungi. They are the most common contaminants of food and feed worldwide and they are considered an important risk factor for human and animal health. This study presented a preliminary evaluation of the endocrine disruptor activity of 328 mycotoxins. The study illustrated the reliability of using *in silico* structural approaches with the support of statistical analysis to assess the possible endocrine disruptor activity of these compounds against 25 nuclear receptors. The results show that the majority of the mycotoxins are predicted as a medium or a good binder for all the 25 nuclear receptors. In particular, ochratoxin A, zearalenone, α- and β-zearalenol, aflatoxin B1, and alternariol have shown to be putative endocrine disruptors chemicals for nuclear receptors, such as the pregnane X receptor, farnesoid X receptor, and liver X receptor, as highlighted from different studies. The combination of these two analyses, *in silico* and statistical, are relatively rapid and inexpensive and can be a powerful tool for a preliminary endocrine disruption evaluation of a large number of compounds.

In our opinion, the present study should not be seen as an opposing method to *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 328 mycotoxins against 25 nuclear receptor family in a relatively short time when compared to the time needed for *in vitro*/*in vivo* experiments.

Acknowledgment

The molecular docking was performed using High Performance Computing Facilities of Centro di Calcolo di Ateneo at the University of Parma.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.taap.2021.115832.

References

Ahmadi, A., Mohammad, N., Ramezani, M., Alibolandi, M., Lavaee, P., Sarreshtehdar, A., Abnous, K., Mohammad, S., 2019. A rapid and simple ratiometric fluorescent sensor for patulin detection based on a stabilized DNA duplex probe containing less

amount of aptamer- involved base pairs. Talanta 204, 641–646. https://doi.org/10.1016/j.talanta.2019.06.057

Akbari, P., Braber, S., Gremmels, H., Koelink, P.J., Verheijden, K.A.T., Garssen, J., Fink-gremmels, J., 2014. Deoxynivalenol : a trigger for intestinal integrity breakdown. FASEB J. 2414–2429. https://doi.org/10.1096/fj.13-238717

Altunay, N., Elik, A., Gürkan, R., 2019. A novel, green and safe ultrasound-assisted emulsification liquid phase microextraction based on alcohol-based deep eutectic solvent for determination of patulin in fruit juices by spectrophotometry. J. Food Compos. Anal. 82, 103256. https://doi.org/10.1016/j.jfca.2019.103256

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. Chem. Res. Toxicol. 22, 52–63. https://doi.org/10.1021/tx800048m

Ates, M.B., Ortatatli, M., 2021. Phase-1 bioactivation mechanisms of aflatoxin through AhR, CAR and PXR nuclear receptors and the interactions with Nigella sativa seeds and thymoquinone in broilers. Ecotoxicol. Environ. Saf. 208, 111774. https://doi.org/10.1016/j.ecoenv.2020.111774

Ayed-Boussema, I., Pascussi, J.M., Maurel, P., Bacha, H., Hassen, W., 2012a. Effect of aflatoxin B1 on nuclear receptors PXR, CAR, and AhR and their target cytochromes P450 mRNA expression in primary cultures of human hepatocytes. Int. J. Toxicol. 31, 86–93. https://doi.org/10.1177/1091581811422453

Ayed-Boussema, I., Pascussi, J.M., Rjiba, K., Maurel, P., Bacha, H., Hassen, W., 2012b. The mycotoxin, patulin, increases the expression of PXR and AhR and their target cytochrome P450s in primary cultured human hepatocytes. Drug Chem. Toxicol. 35, 241–250. https://doi.org/10.3109/01480545.2011.592194

Bennett, J.W., Moore, G.G., 2019. Mycotoxins. Encycl. Microbiol. 16, 267–273. https://doi.org/10.1016/B978-0-12-801238-3.02409-0

Bertero, A., Moretti, A., Spicer, L.J., Caloni, F., 2018. Fusarium molds and mycotoxins: Potential species-specific effects. Toxins (Basel). 10, 1–27. https://doi.org/10.3390/toxins10060244

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

Boutrif, E. & Canet, C., 1998. Mycotoxin prevention and control: FAO programmes. Rev. Médicine Vétérinaire 149: 681-694.

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. https://doi.org/10.1021/jm990352k

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. https://doi.org/10.1016/j.toxlet.2012.07.023

Dellafiora, L., Oswald, I.P., Dorne, J. Lou, Galaverna, G., Battilani, P., Dall'Asta, C., 2020. An in silico structural approach to characterize human and rainbow trout estrogenicity of mycotoxins: Proof of concept study using zearalenone and alternariol. Food Chem. 312. https://doi.org/10.1016/j.foodchem.2019.126088

Do, T.H., Tran, S.C., Le, C.D., Nguyen, H.B.T., Le, P.T.T., Le, H.H.T., Le, T.D., Thai-Nguyen, H.T., 2020. Dietary exposure and health risk characterization of aflatoxin B1, ochratoxin A, fumonisin B1, and zearalenone in food from different provinces in Northern Vietnam. Food Control 112, 107108. https://doi.org/10.1016/j.foodcont.2020.107108

Doricakova, A., Vrzal, R., 2015. A food contaminant ochratoxin A suppresses pregnane X receptor (PXR)-mediated CYP3A4 induction in primary cultures of human hepatocytes. Toxicology 337, 72–78. https://doi.org/10.1016/j.tox.2015.08.012

Eldridge, M.D., Murray, C.W., Auton, T.R., Paolini, G. V., Mee, R.P., 1997. 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, 425–445. https://doi.org/10.1023/A:1007996124545

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

Frizzell, C., Elliott, C.T., Connolly, L., 2014. Effects of the mycotoxin patulin at the level of nuclear receptor transcriptional activity and steroidogenesis in vitro. Toxicol. Lett. 229, 366–373. https://doi.org/10.1016/j.toxlet.2014.06.847

Frizzell, Caroline, Ndossi, D., Kalayou, S., Eriksen, G.S., Verhaegen, S., Sørlie, M., Elliott, C.T., Ropstad, E., Connolly, L., 2013. An in vitro investigation of endocrine disrupting effects of the mycotoxin alternariol. Toxicol. Appl. Pharmacol. 271, 64–71. https://doi.org/10.1016/j.taap.2013.05.002

Frizzell, C., Ndossi, D., Verhaegen, S., Dahl, E., Eriksen, G., Sørlie, M., Ropstad, E., Muller, M., Elliott, C.T., Connolly, L., 2011. Endocrine disrupting effects of zearalenone, alpha- and beta-zearalenol at the level of nuclear receptor binding and steroidogenesis. Toxicol. Lett. 206, 210–217. https://doi.org/10.1016/j.toxlet.2011.07.015

Frizzell, C., Verhaegen, S., Ropstad, E., Elliott, C.T., Connolly, L., 2013. Endocrine disrupting effects of ochratoxin A at the level of nuclear receptor activation and steroidogenesis. Toxicol. Lett. 217, 243–250. https://doi.org/10.1016/j.toxlet.2012.12.018

Guo, W., Pi, F., Zhang, H., Sun, J., Zhang, Y., Sun, X., 2017. A novel molecularly imprinted electrochemical sensor modified with carbon dots, chitosan, gold nanoparticles for the determination of patulin. Biosens. Bioelectron. 98, 299–304. https://doi.org/10.1016/j.bios.2017.06.036

International Agency for Research on Cancer (IARC), 2002. Traditional Herbal Medicines, Some Mycotoxins, Napthalene, and Styrene. Monographs on the Evaluation of Carcinogenic Risks to Humans 82–171.

Kőszegi, T., Poór, M., 2016. Ochratoxin a: Molecular interactions, mechanisms of toxicity and prevention at the molecular level. Toxins (Basel). 8, 1–25. https://doi.org/10.3390/toxins8040111

L. Lewis, M. Onsongo, H. Njapau, H. Schurz-Rogers, G. Luber, S. Kieszak, J. Nyamongo, L. Backer, A. M. Dahiye, A. Misore, K. DeCock, C. Rubin, J. Nyikal, C. Njuguna, A. Langat, I. K. Kilei, C. Tetteh, S. Likimani, J. Oduor, D. Nzioki, B. W. Kamau, J. Onso, and N.G., 2005. Aflatoxin contamination of commercial maize products during an outbreak of acute aflatoxicosis in eastern and central Kenya. Environ. Health Perspect. 12, 1763–1767.

Lee, H.J., Pyo, M.C., Shin, H.S., Ryu, D., Lee, K.W., 2018. Renal toxicity through AhR, PXR, and Nrf2 signaling pathway activation of ochratoxin A-induced oxidative stress in kidney cells. Food Chem. Toxicol. 122, 59–68. https://doi.org/10.1016/j.fct.2018.10.004

Lehmann, L., Wagner, J., Metzler, M., 2006. Estrogenic and clastogenic potential of the mycotoxin alternariol in cultured mammalian cells. Food Chem. Toxicol. 44, 398–408. https://doi.org/10.1016/j.fct.2005.08.013

Liu, X., Guan, X., Xing, F., Lv, C., Dai, X., Liu, Y., 2017. Effect of water activity and temperature on the growth of Aspergillus flavus, the expression of aflatoxin

biosynthetic genes and aflatoxin production in shelled peanuts. Food Control 82, 325–332. https://doi.org/10.1016/j.foodcont.2017.07.012

Molina-Molina, J.M., Real, M., Jimenez-Diaz, I., Belhassen, H., Hedhili, A., Torné, P., Fernández, M.F., Olea, N., 2014. Assessment of estrogenic and anti-androgenic activities of the mycotoxin zearalenone and its metabolites using in vitro receptor-specific bioassays. Food Chem. Toxicol. 74, 233–239. https://doi.org/10.1016/j.fct.2014.10.008

Ndossi, D.G., Frizzell, C., Tremoen, N.H., Fæste, C.K., Verhaegen, S., Dahl, E., Eriksen, G.S., Sørlie, M., Connolly, L., Ropstad, E., 2012. An in vitro investigation of endocrine disrupting effects of trichothecenes deoxynivalenol (DON), T-2 and HT-2 toxins. Toxicol. Lett. 214, 268–278. https://doi.org/10.1016/j.toxlet.2012.09.005

Pfeiffer, E., Eschbach, S., Metzler, M., 2007. Alternaria toxins: DNA strand-breaking activity in mammalian cells in vitro. Mycotoxin Res. 23, 152–157. https://doi.org/10.1007/BF02951512

Prouillac, C., Koraichi, F., Videmann, B., Mazallon, M., Rodriguez, F., Baltas, M., Lecoeur, S., 2012. In vitro toxicological effects of estrogenic mycotoxins on human placental cells: Structure activity relationships. Toxicol. Appl. Pharmacol. 259, 366–375. https://doi.org/10.1016/j.taap.2012.01.016

Riani, M., Atkinson, A.C., Cerioli, A., 2009. Finding an unknown number of multivariate outliers. J. R. Stat. Soc. Ser. B Stat. Methodol. 71, 447–466. https://doi.org/10.1111/j.1467-9868.2008.00692.x

Shen, Y., Shi, Z., Fan, J.T., Yan, B., 2020. Dechlorination and demethylation of ochratoxin A enhance blocking activity of PXR activation, suppress PXR expression and reduce cytotoxicity. Toxicol. Lett. 332, 171–180. https://doi.org/10.1016/j.toxlet.2020.07.012

Tan, S., Ge, W., Wang, J., Liu, W., Zhao, Y., Shen, W., Li, L., 2020. Zearalenone-induced aberration in the composition of the gut microbiome and function impacts the ovary reserve. Chemosphere 244, 125493. https://doi.org/10.1016/j.chemosphere.2019.125493

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

Travis R Bui-Klimke, F.W., 2015. Ochratoxin A and human health risk: A review of the evidence. Crit. Rev. Food Sci. Nutr. 1860–1869.

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.

W. F. O. Marasas, R. T. Riley, K. A. Hendricks, V. L. Stevens, T. W. Sadler, J. Gelineau-Van Waes, S. A. Missmer, J. Cabrera, O. Torres, W. C. A. Gelderblom, J. Allegood, C. Martínez, J. Maddox, J. D. Miller, L. Starr, M. C. Sullards, A. V. Roman, K. A. V, and A.H.M.J., 2004. Fumonisins Disrupt Sphingolipid Metabolism, Folate Transport, and Neural Tube Development in Embryo Culture and In Vivo: A Potential Risk Factor for Human Neural Tube Defects among Populations Consuming Fumonisin-Contaminated Maize. J. Nutr. 4, 711–716.

Wang, X., Yu, H., Shan, A., Jin, Y., Fang, H., Zhao, Y., Shen, J., Zhou, C., Zhou, Y., Fu, Y., Wang, J., Zhang, J., 2018. Toxic effects of zearalenone on intestinal microflora and intestinal mucosal immunity in mice. Food Agric. Immunol. 29, 1002–1011. https://doi.org/10.1080/09540105.2018.1503233

World Health Organization International Agency for Research on Cancer (WHO-IARC), 1993a. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans.

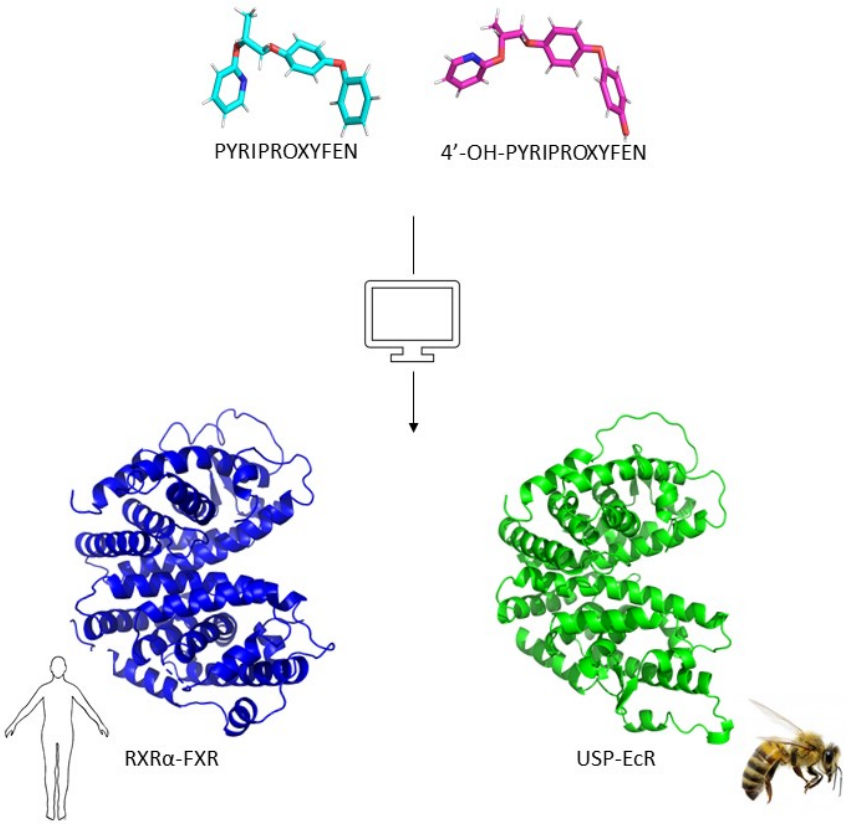
World Health Organization International Agency for Research on Cancer (WHO-IARC), 1993b. Toxins derived from Fusarium moniliforme: fumonisins B1 and B2 and fusarin C. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans 56, 445–462.

CHAPTER 7

In silico Prediction of the Mechanism of Action of Pyriproxyfen and 4′-OH-Pyriproxyfen against *A. mellifera* and *H. sapiens* Receptors

The content of this chapter has also been published in:

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


PYRIPROXYFEN
4'-OH-PYRIPROXYFEN
RXRα-FXR
USP-EcR

Abstract

Background. Poisoning from pesticides can be extremely hazardous for non-invasive species, such as bees, and humans causing nearly 300,000 deaths worldwide every year. Several pesticides are recognized as endocrine disruptors compounds that alter the production of the normal hormones mainly by acting through their interaction with nuclear receptors (NRs). Among the insecticides, one of the most used is pyriproxyfen. As analogous to the juvenile hormone, the pyriproxyfen acts in the bee's larval growth and creates malformations at the adult organism level.

Methods. This work aims to investigate the possible negative effects of pyriproxyfen and its metabolite, the 4'-OH-pyriproxyfen, on human and bee health. We particularly investigated the mechanism of binding of pyriproxyfen and its metabolite with ultraspiracle protein/ecdysone receptor (USP-EcR) dimer of *A. mellifera* and the relative heterodimer farnesoid X receptor/retinoid X receptor alpha (FXR-RXRα) of *H. sapiens* using molecular dynamic simulations.

Results. The results revealed that pyriproxyfen and its metabolite, the 4'-OH-pyriproxyfen, stabilize each dimer and resulted in stronger binders than the natural ligands.

Conclusion. We demonstrated the endocrine interference of two pesticides and explained their possible mechanism of action. Furthermore, *in vitro* studies should be carried out to evaluate the biological effects of pyriproxyfen and its metabolite.

Keywords: molecular dynamic simulations; computational methods; nuclear receptors; bees; endocrine disruptors compounds; pesticides

Introduction

Apis mellifera is the most widespread species in Europe among the *Apis genus*. Being pollinator insects, bees play a fundamental role in the environment by promoting pollination that makes them important, if not necessary, for many crops and for the maintenance of biodiversity (Patel et al., 2021). The Food and Agriculture Organization of the United Nations (FAO) estimates that of the 100 crop species that provide 90% of food worldwide, 71 are pollinated by bees (EFSA, The European Food Safety Authority). Since 1962, bees are used as bioindicators for environmental pollution in a two-fold way: (i) monitoring the mortality; (ii) monitoring the presence of pollution residues in honey, pollen, and bee larvae (Celli and Maccagnani, 2003). In the last 10–15 years, bees' mortality and colony losses have increased (Kielmanowicz et al., 2015) (Hristov et al., 2020) (Neov et al., 2019). The cause for this increase is a combination of factors that affect bees vitality: virus, pathogen, invasive species, and the increasing use of pesticides. The high use of pesticides, first introduced in 1960, causes their persistence in air, soil, and water (Carvalho, 2017). Pesticides are substances or mixtures of substances that are mainly used in agriculture to protect plants from weeds (herbicides), fungus (fungicides), insects (insecticides), and rodents. In fact, some compounds are degraded by light, soil bacteria, or chemical processes, while other compounds persist in air, soil, and water (Bilal et al., 2019). This causes the constant exposure of living beings to many substances that can have harmful effects. Agriculture is the largest consumer but pesticides are also used in public health activities to control vector-borne diseases, unwanted plants and to suppress the proliferation of insects, bacteria, and others (Bilal et al., 2019) (Carvalho, 2017). However, exposure to pesticides can be extremely hazardous to humans and other non-invasive species, such as bees, causing 300,000 deaths worldwide every year (Kim et al., 2017) (Md Meftaul et al., 2020). Pesticides can cause acute health effects (such as stinging eyes, rashes, blisters, blindness, nausea) or chronic adverse effects (such as cancers, birth defects, reproductive harm, neurological and developmental toxicity, immunotoxicity, and disruption of the endocrine system) that can occur months or years after exposure (Carvalho, 2017) (Faber, 2020) (Sabarwal et al., 2018). Some people, such as infants and young children, are more vulnerable than others to pesticide impacts (Rosas and Eskenazi, 2008). In 2002, the World Health Organization (WHO) recognized several pesticides as endocrine disruptors compounds (EDCs), which act on the endocrine system, causing adverse health effects in different organisms and their offspring (Kaur et al., 2019) (Rosas and Eskenazi, 2008). Endocrine disruptors can act, mimic, or partially

mimic the natural hormones in the body altering their metabolism (Monneret, 2017). Many of the insects' endocrine systems are used as targets for the synthesis of pesticides (46% are insecticides, 21% herbicides, and 31% fungicides) that act as endocrine disruptors (Mnif et al., 2011).

Pyriproxyfen, defined as an insect growth regulator (IGR), is the active ingredient used since 1995 in several insecticides, both as a single compound and in combination with other compounds (Chen et al., 2016) (Diamanti-Kandarakis et al., 2009) (Monneret, 2017) . EFSA declared that pyriproxyfen cannot be considered an endocrine disruptor for mammals because there are not sufficiently toxicological studies where adverse effects were observed; while in the case of bees, it stated that the proofs indicated a high risk for the larvae (Chen et al., 2016) (Truman, 2019). Pyriproxyfen acting as a juvenile hormone analogue (JHA) blocks the development of larvae and thus increases mortality, while sublethal doses affect the behavior of bees and create malformations at the adult organism level (Abdourahime et al., 2019) (Fiaz et al., 2019) (Wilson, 2004). These malformations cause problems in the behavior and recognition of bees by the colony (Devillers and Devillers, 2020). This failure to recognize both larvae and adult has the final effect of an increase in mortality, as it affects the stability and growth of the colony (Fiaz et al., 2019) (Wilson, 2004). Pesticides can undergo chemical change after contact with light, heat, soil, plant and after ingestion by an animal with lower or higher toxicity than the pesticide. More than ten metabolites of pyriproxyfen have been characterized in soil, water, plants, mammals, and insects. One of the main metabolites is 4'-OH-pyriproxyfen (4'-OH-PPF) that is generated by the degradation of pyriproxyfen in soil, but also rats and mice (Fiaz et al., 2019) (Fisher et al., 2018) (Liu et al., 2019) (Sullivan and Goh, 2008) (Yoshino et al., 1996).

As a juvenile hormone analogue, pyriproxyfen can affect the function of the ecdysone receptor interacting with the ultraspiracle protein (USP). Ultraspiracle protein/ecdysone receptor (USP-EcR) dimer is an arthropod receptor and is composed of two monomers: EcR (NR1H1) and USP (NR2B4), the latter is an ortholog of RXR (retinoid X receptor, NR2B1), the receptor for the vitamin A metabolite 9-cis-retinoic acid (9-cis-RA) (Clayton et al., 2001) (Oro et al., 1990) (Sasorith et al., 2002). 20-hydroxyecdysone (20E) has been identified as the natural ligand of EcR, on the contrary, the natural ligand of USP has not yet been identified even if several studies have highlighted the possible binding of USP with juvenile hormones (JHs) (Jones and Sharp, 1997) (Nakagawa and Henrich, 2009a) (Sasorith et al., 2002). Henrich and co-workers studied the possible similarity of EcR to

the human FXR (farnesoid X receptor, NR1H4) (Henrich et al., 2003). Farnesoid X receptor is a member of nuclear receptor family that is highly expressed in the liver, intestine, kidney, and adrenal glands, and is involved in maintaining many metabolic pathways, such as bile acid regulation, cholesterol metabolism, glucose and lipid homeostasis (Zheng et al., 2018). To activate the expression of its target genes, FXR heterodimerizes with another nuclear receptor, the retinoid X receptor α (RXRα). The alteration of expression and function of this heterodimer has been reported as a contributing factor in the development of many cancers and other diseases, such as insulin resistance, liver cirrhosis, cholestasis, coronary and crohn diseases, liver and cardiovascular diseases (Kemper, 2011).

Given the lack of information on the risks from pyriproxyfen and its metabolites in animals and in humans, we investigated their possible negative effects on the human and bees health applying molecular modeling techniques. We particularly investigated the mechanism of binding of pyriproxyfen and its metabolite, 4'-OH-pyriproxyfen, against USP-EcR bees dimer and the relative human heterodimer RXRα-FXR. Once the USP and the EcR models were built and the interactions in the USP-EcR and RXRα-FXR dimers interface were studied, molecular docking has been carried out, in order to predict and evaluate the structural physical interactions between the receptor and the pesticides. We thought that a computational study based on nanosecond time-scale molecular dynamic simulation constitutes an appropriate approach to analyze the dynamic behavior of receptors of the bees and humans, USP-EcR and RXRα-FXR, respectively. Moreover, we analysed the interactions with the natural ligand and with pyriproxyfen and its metabolite.

Materials and Methods

Molecular Model of USP and EcR

Since the structure of USP and EcR monomers of *A. mellifera* is not available in Protein Data Bank (PDB), they were modeled using homology modeling techniques. Homology research was carried out using BLAST (National Center for Biotechnology Information, 8600 Rockville PikeBethesda MD, 20894 USA) setting Refseq as a database, an Expected Threshold of 10-5, and a max target of 1000 (Altschul et al., 1990). The sequences of USP (UniProt: Q9NG48) and EcR (UniProt: A2PZF8) of *A. mellifera* used as query sequences for the homology research were found in UniProt (Bateman, 2019). The two monomers were modeled using the LBD sequences and two different templates: *H. sapiens* RXRα (PDB ID: 1FM6) for the modeling of USP and *T. castaneum* EcR (PDB

ID: 2NXX) for the modeling of EcR. In order to obtain different structures to be compared four software were used for the modeling: SWISS-MODEL (Protein Structure Bioinformatics Group c/o Prof. Torsten Schwede Swiss Institute of Bioinformatics Biozentrum, University of Basel Klingelbergstrasse 50/70 CH-4056 Basel/Switzerland), I-TASSER (Iterative Threading ASSEmbly Refinement, 100 Washtenaw Avenue, Ann Arbor, MI 48109-2218), Phyre2 (Protein Homology/analogY Recognition Engine V 2.0), and Chimera MODELLER (UCSF RBVI) (Benkert et al., 2011) (Bertoni et al., 2017) (Bienert et al., 2017) (Guex et al., 2009) (Kelley et al., 2015) (Pettersen et al., 2004) (Roy et al., 2010) (Yang et al., 2014) (Zhang, 2008). The reliability of the models was checked using ProSA-web (Protein Structure Analysis) and Procheck (EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK) which provide, respectively, z-score and G-factor values, in order to evaluate their stereochemistry and energy distribution (Laskowski et al., 1993) (Laskowski et al., 1996) (Waterhouse et al., 2018) (Wiederstein and Sippl, 2007).

Preparation of Proteins

The crystal structures of human RXRα (PDB ID: 1FM6) and FXR (PDB ID: 4QE6) monomers were downloaded from the Protein Data Bank (PDB). Both the crystallographic structure and the predicted structural models of USP and EcR monomers were processed using Sybyl software v8.1 (www.tripos.com. Archived on 5 November 2019): water molecules and ligands were removed, hydrogen atoms were added, and energy was 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 (see below), the receptors were further processed as follows: 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.

Preparation of Ligands

The structural coordinates of the ligands, such as juvenile hormone III, 9-cis-retinoic acid, 20-hydroxyecdysone, chenodeoxycholic acid, and pyriproxyfen were retrieved from the NCBI PubChem compound database (https://pubchem.ncbi.nlm.nih.gov/. Archived on 10 November 2019). In the case of 4'-OH-pyriproxyfen, the three-dimensional structure was built, and energy minimized with Sybyl software v8.1 using the Powell algorithm with a coverage gradient of ≤0.5 kcal/(mol·Å) 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.

GOLD Docking

The GOLD (Protein Ligand Docking Software) software v5.8.1 (CCDC; Cambridge, UK; http://www.ccd.cam.ac.uk. Archived on 28 November 2019) was applied to dock ligands into the binding site of the receptors. For each compound and receptor, 30 binding poses were generated without any constraints. In bees cases, the centroid of the binding site was defined using the coordinates of the crystallographic complexes, 2NXX in the case of EcR monomer (#C24 of P1A: x = 29.069, y = 6.239, z = 8.576) and 1FM6 in the case of USP monomer (#C10 of 9CR: x = 17.688, y = 14.021, z = 14.525), while in the human case was used 1FM6 for RXRα (#C10 of 9CR: x = 17.688, y = 14.021, z = 14.525) and 4QE6 in the case of FXR (#C13 of JN3: x = 10.872, y = 15.018, z = 11.917). Side chain flexibility was allowed for the amino acids: EcR: Glu17, Thr52, Lys93, Phe437, Tyr114; USP: Val30, Ile33, Thr37, Lys39, Leu91, Thr93, Ile110, Leu201, Phe202; RXRα: Phe436, His435, Phe439, Leu436; FXR: Met265, Met290, His294, Phe336, Phe350, Tyr369, Met450, Trp454, Trp469. For the genetic algorithm run, a maximum number of 100,000 operations were performed on a population of 100 individuals with a selection pressure of 1.1. 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 Chem Score and Hint Score (HINT, Hydropathic INTeraction) with the aim to obtain a consensus.

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

USP-EcR and RXRα-FXR Dimers and the Interfaces Key Interactions

To build the USP-EcR and RXRα-FXR dimers, the structural similarities of the structures present in PDB were analysed using the Pymol software: 2R40 of *H. virescens*, 4OZT of

B. ovis, 2NXX of *T. castaneum*, and 5Z12 of *H. sapiens*, in bee case, while 6A5Y (RXRα-FXR with the best resolution value) of *H. sapiens* in the human case. Discovery Studio (Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 2020, San Diego: Dassault Systèmes, 2020) was used to analyze the key interactions on the interface of the two dimers.

Molecular Dynamic Simulations

The best molecular docking pose for each ligand–protein complex was chosen as the starting point of the molecular dynamic simulations. The protein–ligand complex was prepared using the web-based graphical user interface CHARMM-GUI (Effective Simulation Input Generator and More, Lehigh University, Bethlehem) (http://www.charmm-gui.org/. Archived on 25 January 2020). Each complex was solvated in a rectangular 15 Å water box (TIP3S). Molecular dynamic simulations were performed using the NAMD 2.13 (NAMD was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at UrbanaChampaign) software package (Phillips et al., 2020). For each system, two rounds of energy minimization were performed, each comprising 0.1 ns of conjugate gradient minimization. First, a weak constraint (1.0 kcal/(mol·Å^2)) was assigned to all heavy atoms for both the protein and the ligand, allowing the minimization of hydrogen atoms. Second, no restraints were employed in order to allow the minimization of the entire system. Both systems were gradually heated from 50 to 300 K in NVT mode (number of atoms volume temperature) for 0.2 ns, while the heavy atoms of the protein were restrained with a force constant of 0.5 kcal/(mol·Å^2). The system was further equilibrated at constant pressure (1.0 bar) for 1 ns (NPT). Each molecular dynamic simulation was performed for 250 ns without any constraint, allowing the movement of the entire system.

Results and Discussion

Molecular Model of USP and EcR

The structure of EcR was obtained using the two monomers modeled with a homology model approach (Jones et al., 2006). Before monomers construction, the similarity between the sequences was verified using the sequences of the LBD present in UniProt (https://www.uniprot.org/. Archived on 10 October 2019): Q9NG48 for USP, and A2PZF8 for EcR. The analysis of the sequences showed that USP has an identity of

69.51% with the RXRα of *H. sapiens* (Velarde et al., 2006). Using the Clustal (https://www.ebi.ac.uk/Tools/msa/clustalo/. Archived on 22 October 2019) program, comparison of the two sequences was done to verify the preservation of residues important for the ligand binding and for the stabilization of the H12 in an agonist conformation. These residues were identified through literature research (Ala271, Ala272, Gln275, Trp305, Asn306, Phe313, Arg316, and Cys432) (Sasorith et al., 2002). As shown in the alignment (Figure 1) most of the residues are preserved. There are exceptions for some residues that are replaced by residues with the same chemical properties, Gln270 is replaced by Asn236, Leu315 is replaced by Val291. The only exception is His435 that is replaced in USP by Tyr401. This could affect the ligand binding. Moreover, the structures of USP and RXRα were compared in the binding pocket region to analyze the preservation of the residues and of the structure.

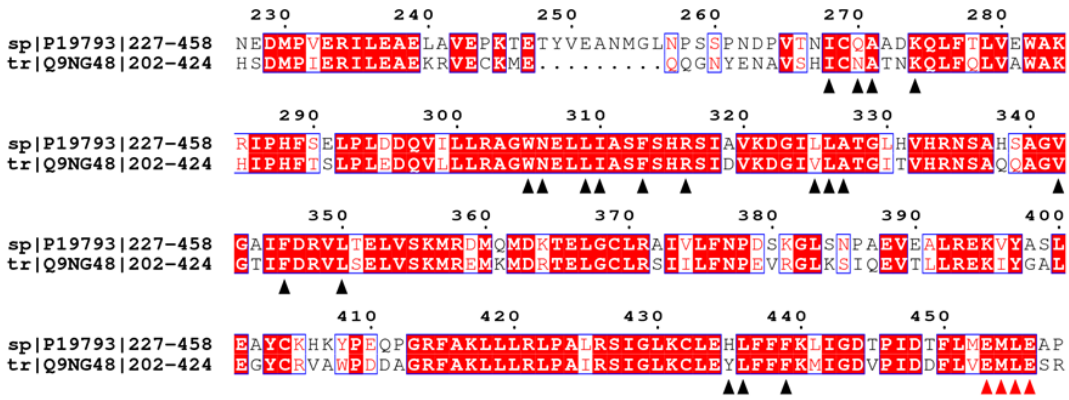


Figure 1. The proteins sequence alignment of *H. sapiens* RXRα (sp|P19793|) and *A. mellifera* USP (tr|Q9NG48|). The alignment was done using ClustalX and ESPript. Black arrows indicate the residues that in *H. sapiens* bind the natural ligand. Red arrows indicate the EMLE sequence important for the stabilization of H12 in *H. sapiens*. The numbering at the top refers to the sequence of *H. sapiens*.

The binding pocket residues found in the RXRα structure and those found in the USP model of *A. mellifera* are preserved in both the sequence and the structure (Figure 2).
In addition, through the overlapping of the structures, it was also possible to see how the width of the binding pocket does not change. The sequence of FXR has an identity of 36% with the EcR of *A. mellifera*. Because this value is pretty low, we searched for a better sequence for the modeling using BLAST. We used the bee's EcR LBD sequence as a query and setting an expected threshold of 10-5 in order to obtain sequences with a

high identity for the model construction (identity values greater than 40%). Among the results obtained with BLAST, only sequences with a PDB crystallographic structure were taken into account: USP of *Heliothis virescens* (PDB ID: 2R40), USP of *Brucella ovis* (PDB ID: 4OZT), and USP of *Tribolium castaneum* (PDB ID: 2NXX). These sequences were aligned to analyze the sequence identity and after comparing the different sequences with the bee's sequence,

given the greater similarity, the EcR of *T. castaneum*, which has an identity of 85.90% to EcR LBD of *A. mellifera*, was used as a template. The two monomers were modeled using the LBD sequences of Apis mellifera and two different templates, whose structure was taken from PDB: RXRα of *H. sapiens* (PDB ID: 1FM6) for modeling of USP and EcR of *T. castaneum* (PDB ID: 2NXX) for modeling of EcR. The homology modeling was carried out using four software in order to obtain four structures to compare using the z-score and the G-factor values to evaluate their stereochemistry and energy distribution. Using this method is possible to obtain a more reliable structure. The structure modelled with Chimera MODELLER was used for USP (z-score: −6.7; G-factor: −0.12) and the one modelled with Phyre2 was used for EcR (z-score: −8.27; G-factor: 0.34).

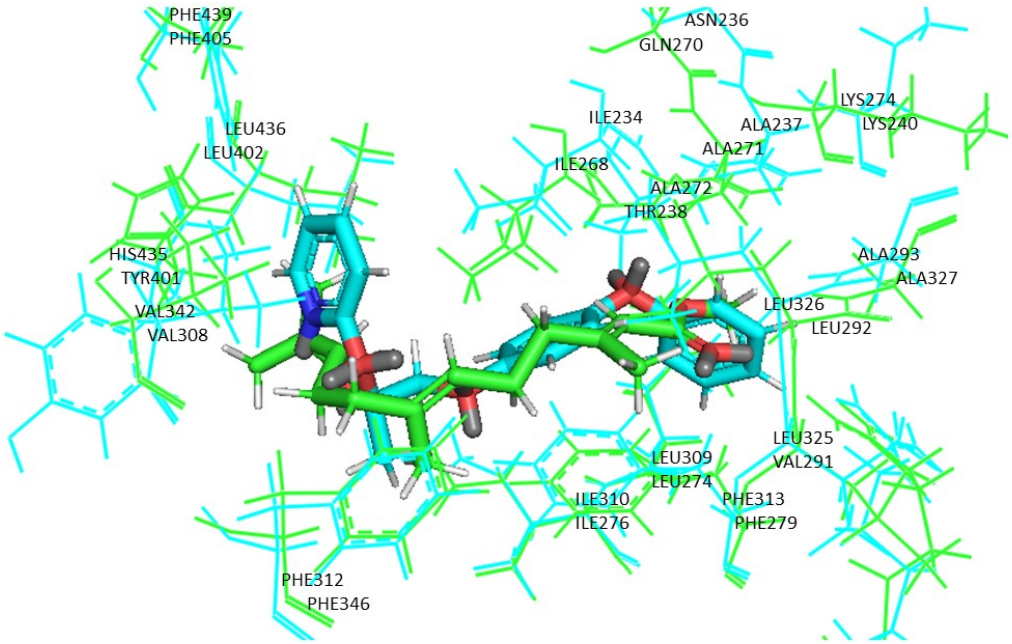


Figure 2. The binding pocket residues of RXRα (in blue) and USP (in green) in complex with pyriproxyfen and juvenile hormone III, respectively.

Comparison of *A. mellifera* and *H. sapiens* Models

Two monomers of *A. mellifera*, USP and EcR, and *H. sapiens*, RXRα (PDB ID: 1FM6) and FXR (PDB ID: 4QE6), were superimposed on the RXRα-FXR dimer structure of *H. sapiens* (PDB ID: 6A5Y) in order to place them at the right distance for the formation of the dimer. Through literature research, information was obtained on the interface surface of the two bee monomers and on the presence of amino acids important in the interaction of monomers for the formation of the dimer in some insects *B. ovis*, *T. castaneum*, and *H. virescens* (Gilbert, 2012). The structural similarities of the EcR-USP of *A. mellifera* and RXRα-FXR of *H. sapiens* were analysed, and specifically the interactions between the two monomers, to assess whether there is preservation in the interaction for the formation of the dimer. The intermolecular interactions are grouped into polar and nonpolar interactions. The nonpolar interactions include Van der Waals's contacts and hydrophobic interactions with a distance cut-off of 4.5 Å, while the polar interactions include charged interactions (5.5 Å cut-off) and hydrogen bonds (4.0 Å cut-off). Close examination of the two dimers interfaces reveals that the residues present in the interface surfaces and residues that promote the interaction between monomers are preserved except for one residue: Asn206 in the EcR structure is replaced by His445 in the FXR structure. The interactions between both USP and EcR and RXRα and FXR were stabilized by a combination of hydrophobic and electrostatic interactions of the monomers: (i) in *A. mellifera* the nitrogen atom of Arg191 in USP makes an electrostatic interaction with the oxygen atom in EcR, while Pro188, Leu184, and Leu185 in USP form hydrophobic interactions respectively with Arg202 and Leu195 in EcR; (ii) in *H. sapiens* Leu419, Pro423, and Leu430 in RXRα make hydrophobic interactions respectively with Leu434, Arg441, and His445 in FXR (Figure 3) (Kojetin et al., 2015) (Zheng et al., 2018). RXRα and FXR interact via the conserved asymmetric dimer interface composed mainly of H11 in each monomer (Wang et al., 2018). Comparing the structures, the interface among all dimers is very similar, both in terms of the distance between monomers and in the secondary structures involved in the interaction between them.

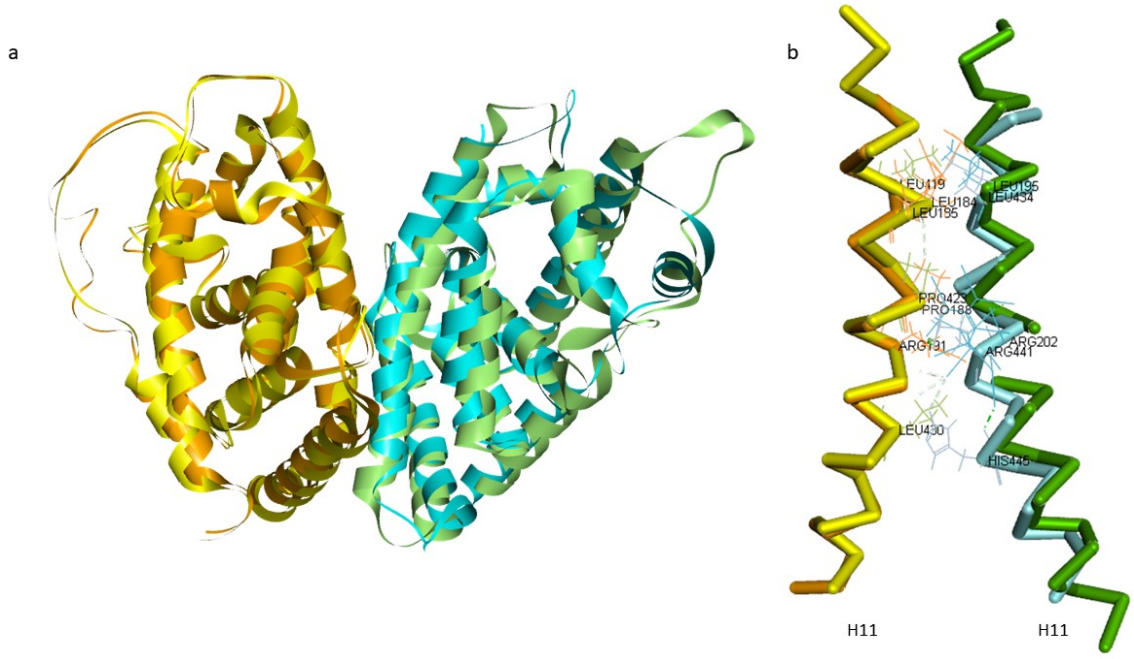


Figure 3. RXRα-FXR and USP-EcR dimers. (a) Alignment of the dimer of *A. mellifera*, USP (yellow) and EcR (green), and *H. sapiens*, RXRα (orange) and FXR (cyan). (b) Focus on the intermolecular interactions mediated by helices 11. Key residues that form the core hydrophobic interface of the parallel coiled are labelled.

Molecular Docking

Molecular docking was carried out using two software, GOLD (http://www.ccd.cam.ac.uk. Archived on 28 November 2019) and AutoDock (Trott and Olson, 2009). The purpose of using more software is to have the possibility to compare multiple scoring functions, which have different protein–ligand interaction assessment methods. This allowed us to have more reliable results because, if three scoring functions provide a positive result, the prediction turns out to be more valid. The combination of different scoring functions allows to reduce the number of false positive, leading to more reliable results. 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) (Charifson et al., 1999). In particular, the GOLD software was used to generate the poses of each chemical, which were scored with Gold Score, Chem Score, and Hint Score, while the AutoDock software was used to generate and to score the poses with the internal scoring function. For each chemical four scoring values were obtained and their binding affinities were scored in comparison to the respective natural ligand, used as a reference compound. The use of multiple scoring functions allowed the

comparison of the results obtained and the calculation of a consensus score. The natural ligands chosen were: 9-cis-retinoic acid for RXRα, chenodeoxycholic acid (JN3) for FXR, juvenile hormone III (JHIII) for USP, and 20-hydroxyecdysone for EcR (Jones and Sharp, 1997) (Nakagawa and Henrich, 2009b) (Parks et al., 1999) (Sasorith et al., 2002) (Wang et al., 1999) (Wang et al., 2008). Docking results of the ligands are illustrated in Table 1.

Table 1. Molecular docking results of the natural ligands, pyriproxyfen, and 4'-OH-pyriproxyfen against *A. mellifera* and *H. sapiens*' monomers.

RXRα				
Ligand	Gold Score	Chem Score	Hint Score	Affinity
9-cis-retinoic acid	72.96	39.72	1374.1	-9.9
Pyriproxyfen	59.58	35.02	1376.3	-9.8
4'-OH-pyriproxyfen	64.52	35.82	1279.4	-9.8
FXR				
Chenodeoxycholic acid	74.56	32.1	1920.2	-11.2
USP				
Juvenile hormone III	49.02	26.73	841.04	-6.9
Pyriproxyfen	60.29	32.14	1178.7	-9.0
4'-OH-pyriproxyfen	56.15	29.75	1136.0	-9
EcR				
20-hydroxyecdysone	77.58	25.98	-2580,88	-9.7

The approach of these software is to dock/score thoroughly all possible positions of each ligand in the binding site. The docking of the molecules was successful as indicated by the statistically significant scores, except for the Hint Score of 20-hydroxyecdysone. 20E was considered the natural ligand of EcR. It is known that the natural ligand, by definition, interacts and binds with the protein.

Analyses were made to assess if the problem was the modeled structure of EcR. A new Hint Score calculation was carried out using the crystal structure of the ligand-binding domains of the *T. castaneum* heterodimer EcR-USP (PDB ID: 2NXX) bound to Ponasterone A (P1A) obtaining a negative result. We decided not to consider the Hint Score for the docking of the EcR monomer because we achieved a negative Hint Score using both natural ligand (20E) and ligand inside the pocket of the crystallized structure (P1A). Moreover, by using multiple scoring functions to obtain a consensus and having three scoring functions out of four that predicted a positive interaction, probably HINT cannot reliably predict possible interaction in the case of the EcR monomer. As shown

in Table 1 the molecular docking values indicated that the binding of pyriproxyfen and 4'-OH-PPF with RXRα and USP monomers is stronger than that of 9-cis-RA and JHIII, respectively. As shown in Figure 4a, 4'-OH-PPF interacts through the same binding mode of 9-cis-RA: the formation of a hydrogen bond between the oxygen of the two ligands and active site residue Arg316 (Chitranshi et al., 2019) (Dawson and Xia, 2012) (Gampe et al., 2000). In the case of pyriproxyfen, the hydrogen bond is not present, but there are many small hydrophobic interactions with residues in RXRα: Ala272 and Val349.

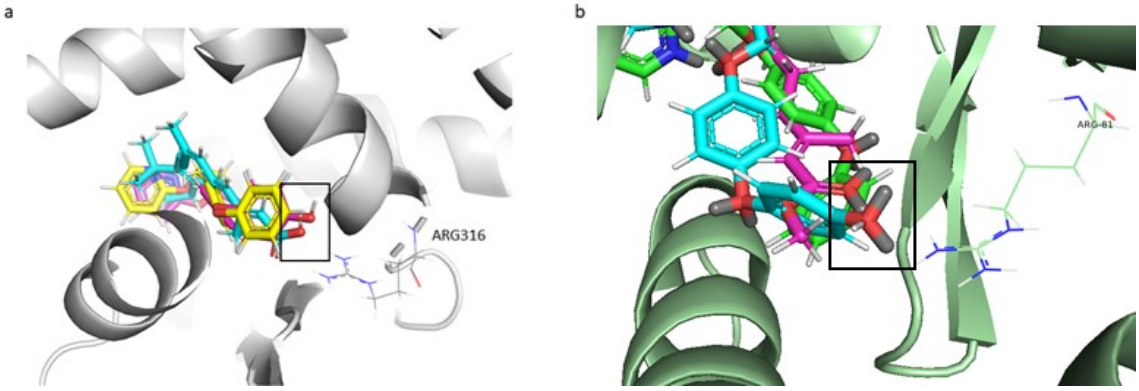


Figure 4. Ligand-receptor interactions. (a) The human hydrogen bond formed between the oxygen of 9-cis-RA (cyan) and 4'-OH-PPF (magenta) and Arg316. The box highlights the oxygen present in 9-cis-RA and the 4'-OH-PPF, but not in the pyriproxyfen (yellow). (b) The bee hydrogen bond formed between the oxygens of the juvenile hormone (cyan) and 4'-OH-PPF (magenta) and Arg81. The box highlights the oxygen present in the juvenile hormone and 4'-OH-PPF. The pyriproxyfen (green) does not have the oxygen for the hydrogen bond.

Moreover, in bees as we discovered for human, the ligand interacts with an arginine residue present in the USP active site. Arg81 of the *A. mellifera* is in the same position as Arg316 of RXRα which, as said before, makes important interaction for the ligand binding. As shown in Figure 4b the interaction is a hydrogen bond between the Arg81 and the oxygens of the 4'-OH-PPF and the juvenile hormone. Pyriproxyfen does not have this interaction because of the presence of a benzenic ring in the position instead of oxygen. This benzenic ring does not interact with the arginine but makes interactions with the Ala36 as in humans. Among all the interactions some are present between residues important for the ligand binding, like Ile33 and Val107, and all three ligands.

Molecular Dynamic Simulations

To evaluate the stability and the mechanism of interaction of pyriproxyfen and 4'-OH-PPF with *A. mellifera* and *H. sapiens*' dimers, 250 ns of molecular dynamic (MD) simulations were carried out for six different complexes: (i) RXRα-FXR with 9-cis-RA and JN3, respectively; (ii) RXRα-FXR with pyriproxyfen and JN3, respectively; (iii) RXRα-FXR with 4'-OH-PPF and JN3, respectively; (iv) USP-EcR with JHIII and 20E, respectively; (v) USP-EcR with pyriproxyfen and 20E, respectively; vi) USP-EcR with 4'-OH-PPF and 20E, respectively. The root-main-square-deviation (RMSD) of the protein backbone was used to monitor conformational changes and, hence, the stability of each system during the total simulation run. From Figure 5a, it can be seen that the RMSD value of the protein backbone (RXRα-FXR) for the three *H. sapiens* systems increased ranging from 1.0–3.5 Å and ultimately attained equilibrium at about 50 ns. Upon binding, the averaged RMSD for the complex of RXRα-FXR with 9-cis-RA, pyriproxyfen, and 4'-OH-PPF was 3.04, 2.95, and 3.14 Å, respectively. As we can see from the graph, the dimer in complex with 9- cis-RA and with 4'-OH-PPF have the same constant and stable trend. To get insights into the stability of the systems, the RMSD value of the RXRα monomer backbone was calculated. From Figure 5b, it can be seen that RXRα is more stable when in complex with 4'-OH-PPF (RMSD average 2.64 Å) than when in complex with pyriproxyfen (RMSD average 3.01 Å) and natural ligand (RMSD average 2.81 Å). This is probably due to the fact that 4'-OH-PPF establishes with a protein residue a hydrogen bond reducing the conformational flexibility of RXRα compared to the pyriproxyfen. Thus, the RMSD of each ligand, 9-cis-RA, pyriproxyfen, and 4'-OH-PPF, with respect to the initial positions of the ligand atoms was evaluated for each complex. Figure 5c shows the RMSD plot of 9-cis-RA, pyriproxyfen, and 4'-OH-PPF molecules present in the active site of RXRα. During the simulation, after 40 ns, there is no significant fluctuation in the 9-cis-RA and 4'-OH-PPF molecules when they are present in RXRα; the corresponding maximum RMSD value is 2.4 Å and 2.9 Å, respectively. The stability of these two molecules in the binding cavity is due to the hydrogen bonds that limit the fluctuations of the protein–ligand complexes. Contrary to these two molecules, in the pyriproxyfen-RXRα complex, this trend is found to be different, wherein the RMSD of the pyriproxyfen molecule is relatively greater when compared with the other two systems.

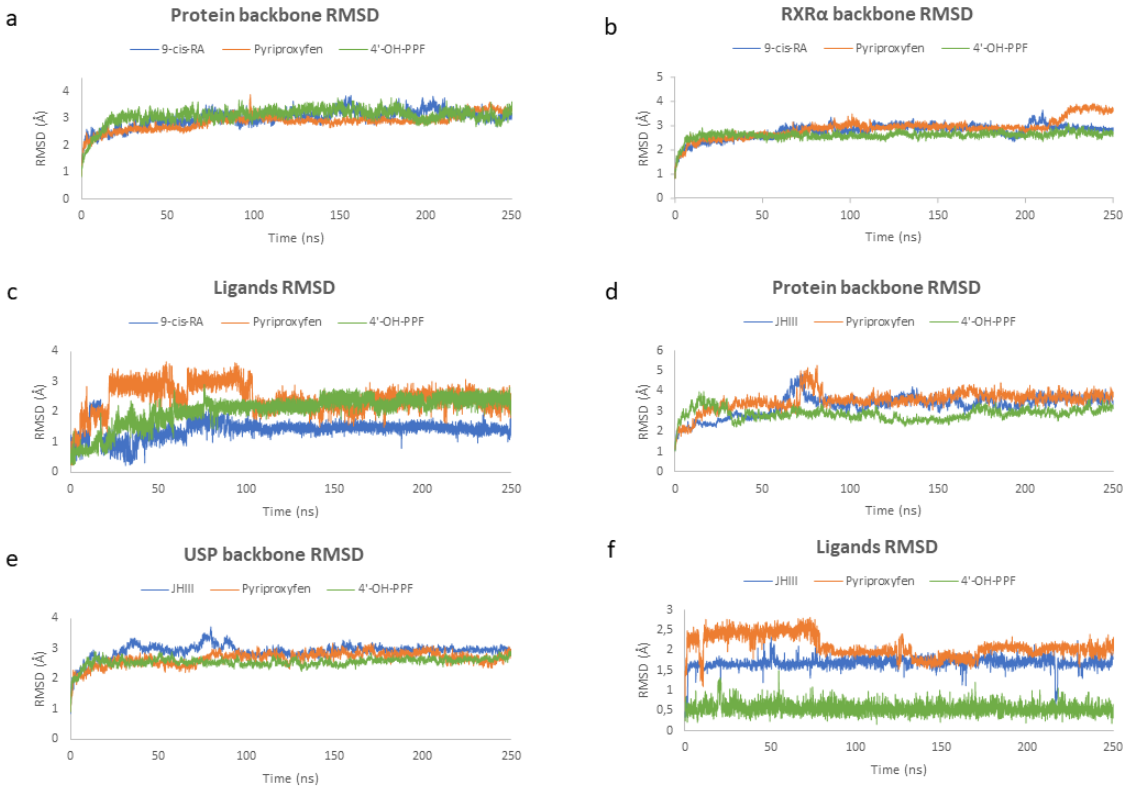


Figure 5. RMSD results graphics in *Homo sapiens* and *Apis mellifera*. RMSD of protein backbone of *H. sapiens* (a) and *A. mellifera* (d), RMSD of RXRα (b) and USP (e) monomer, and heavy atoms of the ligands of *H. sapiens* (c) and *A. mellifera* (f).

In the case of *A. mellifera*, the RMSD value of the protein backbone (USP-EcR) for the two systems (the dimer in complex with juvenile hormone and the dimer in complex with pyriproxyfen) increased ranging from 1.0–5 Å and ultimately attained equilibrium at about 80 ns, while for the dimer in complex with 4'-OH-PPF increased ranging from 1.0–3.5 Å and ultimately attained equilibrium at about 40 ns (Figure 5d). Upon binding, the averaged RMSD for the complex of USP-EcR with JHIII, pyriproxyfen, and 4'-OH-PPF was 3.29, 3.56, and 2.87 Å, respectively. As we can see from the graph, the dimer in complex with 4'-OH-PPF has the same constant and stable trend for the molecular dynamic simulation. To get insights into the stability of the systems, the RMSD value of the RXRα monomer backbone was calculated. The Figure 5e shows that USP is more stable with the 4'-OH-PPF (RMSD average 2.55 Å) and with pyriproxyfen (RMSD average 2.66 Å) than natural ligand (RMSD average 2.91 Å). Thus, the RMSD of each ligand, juvenile hormone, pyriproxyfen, and 4'-OH-PPF, with respect to the initial positions of the ligand atoms was evaluated for each complex (Figure 5f). During the simulation, there is no significant fluctuation in the 4'-OH-PPF molecules when they are

present in USP. The stability of this molecule in the binding cavity is due to the hydrogen bonds that contribute to the small fluctuations of the protein–ligand complex. A similar trend is seen in the pyriproxyfen graph where at about 80 ns there is an RMSD decrease and it ultimately attains equilibrium.

The results obtained for *H. sapiens* and *A. mellifera* show that in both cases the pyriproxyfen and the 4'-OH-pyriproxyfen are stable during the dynamic simulations, and there is no difference between the dimer in complex with the natural ligands and the dimer in complex with the pesticides. One difference, in both cases, is that the dimer in complex with the 4'-OH-pyriproxyfen results to be more stable with respect to the natural ligands and the pyriproxyfen and the ligands have no significant fluctuation due to the hydrogen bonds presents in both cases.

Comparing the formation and persistence of hydrogen bonds network between the ligands and the protein, it is worth to note that the interaction between 9-cis-retinoic acid and RXRα, and 4'-OH-pyriproxyfen and RXRα is characterized by the formation and breaking of two hydrogen bonds: one occurs only for few nanoseconds at the beginning of the simulation and the second one is maintained during the total simulation. In fact, the residues involved in the ligand-binding during the simulation are equal for the two complexes. Contrarywise, pyriproxyfen establishes with residues of RXRα only weak and small hydrophobic interactions resulting in a major instability of the system. The interaction between the oxygen both 9-cis-RA and 4'-OH-PPF and Arg316 breaks at the beginning of the simulation, while the hydrogen bond between 9-cis-RA and 4'-OH-PPF and Ala327 and Asn306 is mostly stable for the total simulation run, contributing to the 42.50% and 49.58% followed by the interaction with Ile268 (1.25%) and Ala327 (2.08%) respectively. These interactions are explained by the graphs showing the distances between the residues of the protein and 9-cis-retinoic acid and 4'-OH-pyriproxyfen (Figure 6a,b). The same situation of RXRα is detected for USP. As we can see in Figure 6c,d the interaction between the oxygen of both the juvenile hormone and 40-OH-PPF and Arg81 breaks at the beginning of the simulation. After that, JHIII and 4'-OH-PPF form other hydrogen bonds with some residues in the binding cavity: Ala92 (53.00%) and Cys197 (0.04%) in the case of juvenile hormone and Ala92 (48.28%), Thr37 (0.12%), Ala36 (0.04%), and Gln40 (0.22%) in the case of 4'-OH-pyriproxyfen.

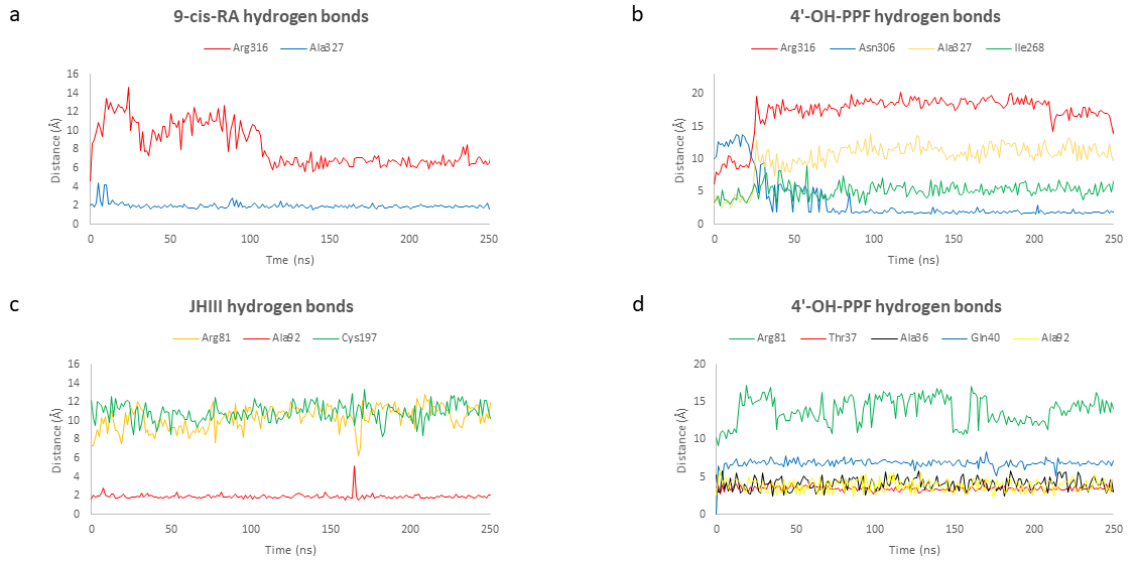


Figure 6. Distances between 9-cis-retinoic acid (a) and 4'-OH-pyriproxyfen (b) and the residues of RXRα involved in the hydrogen bond interactions with the ligands. Distances between juvenile hormone (c) and 4'-OH-pyriproxyfen (d) and the residues of USP involved in the hydrogen bond interactions with the ligands.

In addition, the root main square fluctuation (RMSF) of the six complexes was monitored to analyze the local mobility of protein residues. As shown in Figure 7, the three *H. sapiens* (a) complexes and the three *A. mellifera* (b) complexes had a similar trend. However, in the case of RXRα-FXR in the regions corresponding to the amino acids from 324 to 330 and from 453 to 460, greater fluctuations were evident in the 9-cis-RA complex compared to the other two systems (Figure 8). This is due probably to the major instability of the RXRα-9-cis-RA system.

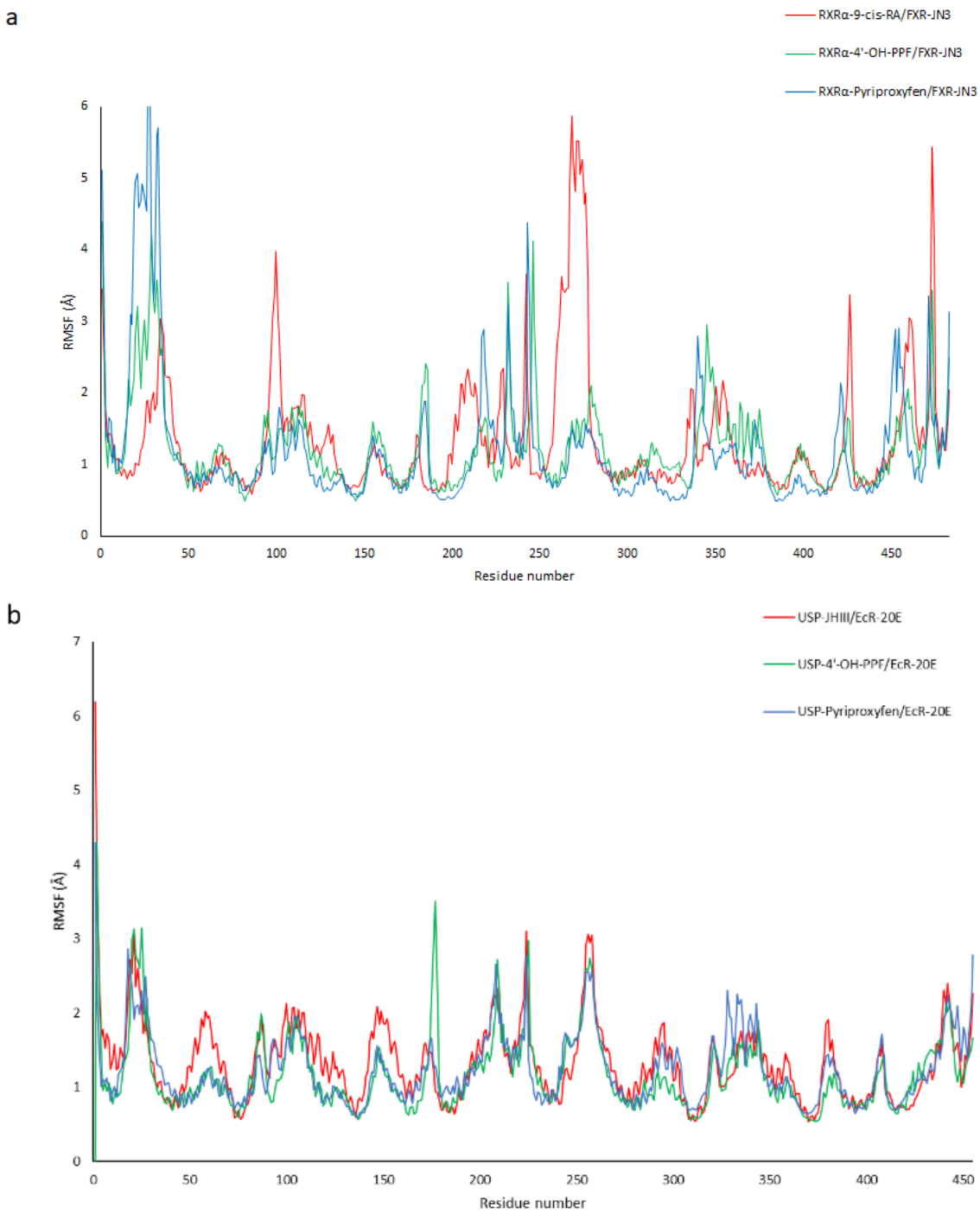


Figure 7. RMSF of the three *H. sapiens* (a) complexes and the three *A. mellifera complexes* (b) obtained by molecular dynamic simulations.

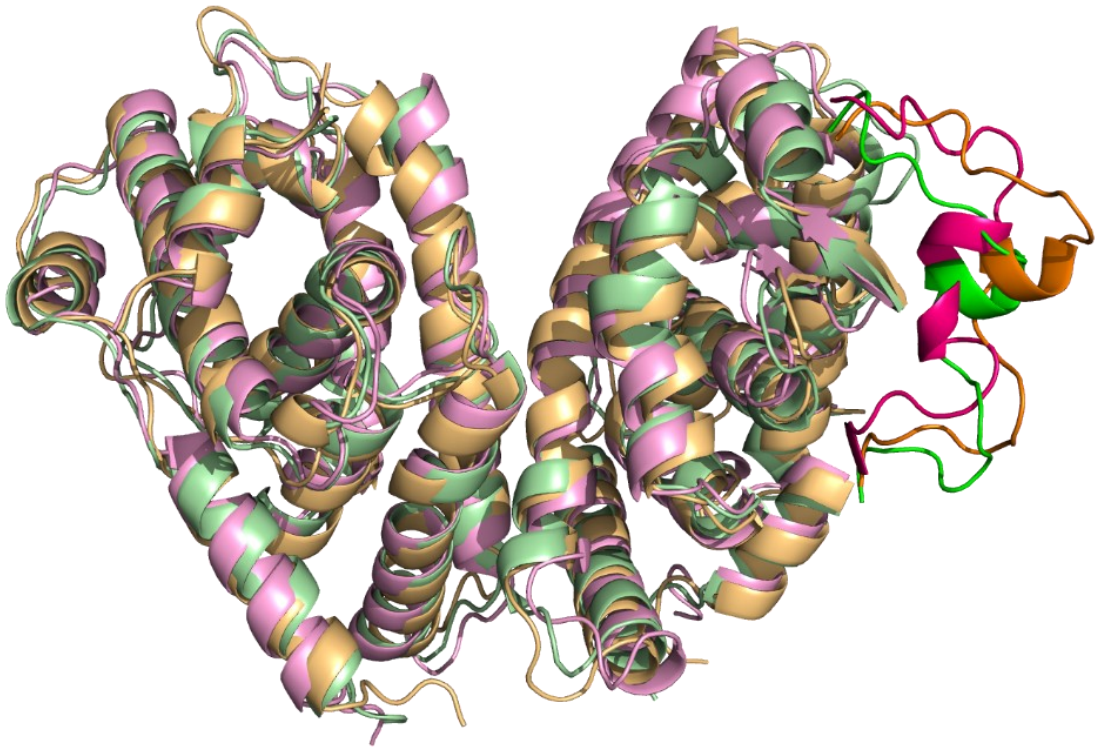


Figure 8. Alignment of RXRα-FXR in complex with 9-cis-RA and JN3 respectively at 0 ns, 125 ns, and 250 ns. Only relevant conformational changes have been highlighted.

Conclusions

Pesticides are widely used in agriculture worldwide. The increase of the pesticides used and their persistence in air, soil, and water is the reason why these compounds are associated with human disease and bee disease and mortality. An important nuclear receptor in bees is the ecdysone receptor that regulates the development and behavior of bees through its activation induced by hormones such as 20-hydroxyecdysone and juvenile hormone. This receptor, composed of two monomers, EcR and USP, is an ortholog of the human FXR-RXRα.

The purpose of this paper was to use *in silico* techniques for the prediction of endocrine interference of pyriproxyfen and its metabolite 4'-OH-pyriproxyfen on the *A. mellifera* USP-EcR dimer and the *H. sapiens* ortholog RXRα-FXR. Docking results, both for humans and bees, predicted a protein–ligand interaction for the two compounds that can be considered as possible binders for the USP and RXRα monomers. Our results show that the 4'-OH-pyriproxyfen, like the natural ligand, makes an important hydrogen bond with an Arg residue (Arg316 for humans and Arg81 for bees) that is known to be an important residue for the ligand binding. The molecular dynamic simulation allows us to

analyze the stability of the dimers and of the ligands inside the binding pocket and the interactions changing and conservation during the 250 ns of simulation. Our results show how the two compounds

are stable inside the binding pockets and comparing the simulation that we studied there are no significant differences between the dimers binding the two compounds and the dimers binding the natural ligand. We also showed that the 4'-OH-pyriproxyfen seems to be more stable with respect to the pyriproxyfen, both in humans and bees.

In conclusion, these *in silico* analyses revealed a possible interaction of the two compounds, pyriproxyfen and 4'-OH-pyriproxyfen, with RXRα-FXR and USP-EcR dimers. These interactions and possible binding to the monomers can affect the normal function of the dimers. We demonstrated the endocrine interference of these two compounds and we explained the possible mechanism of action; *in vitro* studies should be carried out to evaluate the biological effects of pyriproxyfen and its metabolite.

Acknowledgments: The molecular dynamic simulations were performed thanks to Iscra-C grant of 50,000 computers hours on High Performance Computing system, CINECA, Bologna, Italy.

References

Abdourahime, H., Anastassiadou, M., Arena, M., Auteri, D., Barmaz, S., Brancato, A., Bura, L., Carrasco Cabrera, L., Chaideftou, E., Chiusolo, A., Court Marques, D., Crivellente, F., De Lentdecker, C., Egsmose, M., Fait, G., Ferreira, L., Gatto, V., Greco, L., Ippolito, A., Istace, F., Jarrah, S., Kardassi, D., Leuschner, R., Lostia, A., Lythgo, C., Messinetti, S., Miron, I., Molnar, T., Padovani, L., Parra Morte, J.M., Pedersen, R., Raczyk, M., Reich, H., Ruocco, S., Saari, K.E., Santos, M., Serafimova, R., Sharp, R., Stanek, A., Streissl, F., Sturma, J., Szentes, C., Terron, A., Tiramani, M., Vagenende, B., Vainovska, P., Villamar-Bouza, L., 2019. Peer review of the pesticide risk assessment of the active substance pyriproxyfen. EFSA J. 17. https://doi.org/10.2903/j.efsa.2019.5732

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., Pennsylvania, T., Park, U., 1990. Basic Local Alignment Search Tool 2Department of Computer Science. J Mol Biol 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Bateman, A., 2019. UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res.

47, D506–D515. https://doi.org/10.1093/nar/gky1049

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

Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., Schwede, T., 2017. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. Sci. Rep. 7, 1–15. https://doi.org/10.1038/s41598-017-09654-8

Bienert, S., Waterhouse, A., De Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L., Schwede, T., 2017. The SWISS-MODEL Repository-new features and functionality. Nucleic Acids Res. 45, D313–D319. https://doi.org/10.1093/nar/gkw1132

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. https://doi.org/10.1016/j.scitotenv.2019.133896

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

Carvalho, F.P., 2017. Pesticides, environment, and food safety. Food Energy Secur. 6, 48–60. https://doi.org/10.1002/fes3.108

Celli, G., Maccagnani, B., 2003. Honey bees as bioindicators of environmental pollution. Bull. Insectology 56, 137–139.

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. https://doi.org/10.1021/jm990352k

Chen, Y.W., Wu, P.S., Yang, E.C., Nai, Y.S., Huang, Z.Y., 2016. The impact of pyriproxyfen on the development of honey bee (Apis mellifera L.) colony in field. J. Asia. Pac. Entomol. 19, 589–594. https://doi.org/10.1016/j.aspen.2016.06.005

Chitranshi, N., Dheer, Y., Kumar, S., Graham, S.L., Gupta, V., 2019. Molecular docking, dynamics, and pharmacology studies on bexarotene as an agonist of ligand-activated transcription factors, retinoid X receptors. J. Cell. Biochem. 120, 11745–11760. https://doi.org/10.1002/jcb.28455

Clayton, G.M., Peak-Chew, S.Y., Evans, R.M., Schwabe, J.W.R., 2001. The structure of the ultraspiracle ligand-binding domain reveals a nuclear receptor locked in an

inactive conformation. Proc. Natl. Acad. Sci. U. S. A. 98, 1549–1554. https://doi.org/10.1073/pnas.98.4.1549

Dawson, M.I., Xia, Z., 2012. The retinoid X receptors and their ligands. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids 1821, 21–56. https://doi.org/10.1016/j.bbalip.2011.09.014

Devillers, J., Devillers, H., 2020. Lethal and sublethal effects of pyriproxyfen on apis and non-apis bees. Toxics 8, 1–21. https://doi.org/10.3390/toxics8040104

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

Faber, D., 2020. Poisoning the World for Profit: Petro-Chemical Capital and the Global Pesticide Crisis. Capital. Nature, Social. 31, 1–17. https://doi.org/10.1080/10455752.2020.1829794

Fiaz, M., Martínez, L.C., Plata-Rueda, A., Goncalves, W.G., De Souza, D.L.L., Cossolin, J.F.S., Carvalho, P.E.G.R., Martins, G.F., Serrão, J.E., 2019. Pyriproxyfen, a juvenile hormone analog, damages midgut cells and interferes with behaviors of Aedes aegypti larvae. PeerJ 2019, 1–21. https://doi.org/10.7717/peerj.7489

Fisher, A., Colman, C., Hoffmann, C., Fritz, B., Rangel, J., 2018. The Effects of the Insect Growth Regulators Methoxyfenozide and Pyriproxyfen and the Acaricide Bifenazate on Honey Bee (Hymenoptera: Apidae) Forager Survival. J. Econ. Entomol. 111, 510–516. https://doi.org/10.1093/jee/tox347

Gampe, R.T., Montana, V.G., Lambert, M.H., Miller, A.B., Bledsoe, R.K., Milburn, M. V., Kliewer, S.A., Willson, T.M., Xu, H.E., 2000. Asymmetry in the PPARγ/RXRα crystal structure reveals the molecular basis of heterodimerization among nuclear receptors. Mol. Cell 5, 545–555. https://doi.org/10.1016/S1097-2765(00)80448-7

Gilbert, L., 2012. Insect Endocrinology, Insect Endocrinology. https://doi.org/10.1016/C2009-0-62089-4

Guex, N., Peitsch, M.C., Schwede, T., 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. Electrophoresis 30, 162–173. https://doi.org/10.1002/elps.200900140

Henrich, V.C., Burns, E., Yelverton, D.P., Christensen, E., Weinberger, C., 2003. Juvenile hormone potentiates ecdysone receptor-dependent transcription in a mammalian cell culture system. Insect Biochem. Mol. Biol. 33, 1239–1247.

https://doi.org/10.1016/j.ibmb.2003.08.002

Hristov, P., Shumkova, R., Palova, N., Neov, B., 2020. Factors associated with honey bee colony losses: A mini-review. Vet. Sci. 7, 1–16. https://doi.org/10.3390/vetsci7040166

Jones, G., Jones, D., Teal, P., Sapa, A., Wozniak, M., 2006. The retinoid-X receptor ortholog, ultraspiracle, binds with nanomolar affinity to an endogenous morphogenetic ligand. FEBS J. 273, 4983–4996. https://doi.org/10.1111/j.1742-4658.2006.05498.x

Jones, G., Sharp, P.A., 1997. Ultraspiracle: An invertebrate nuclear receptor for juvenile hormones. Proc. Natl. Acad. Sci. U. S. A. 94, 13499–13503. https://doi.org/10.1073/pnas.94.25.13499

Kaur, R., Mavi, G.K., Raghav, S., Khan, I., 2019. Pesticides Classification and its Impact on Environment. Int. J. Curr. Microbiol. Appl. Sci. 8, 1889–1897. https://doi.org/10.20546/ijcmas.2019.803.224

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

Kemper, J.K., 2011. Regulation of FXR transcriptional activity in health and disease: Emerging roles of FXR cofactors and post-translational modifications. Biochim. Biophys. Acta - Mol. Basis Dis. 1812, 842–850. https://doi.org/10.1016/j.bbadis.2010.11.011

Kielmanowicz, M.G., Inberg, A., Lerner, I.M., Golani, Y., Brown, N., Turner, C.L., Hayes, G.J.R., Ballam, J.M., 2015. Prospective Large-Scale Field Study Generates Predictive Model Identifying Major Contributors to Colony Losses. PLoS Pathog. 11, 1–20. https://doi.org/10.1371/journal.ppat.1004816

Kim, K.H., Kabir, E., Jahan, S.A., 2017. Exposure to pesticides and the associated human health effects. Sci. Total Environ. 575, 525–535. https://doi.org/10.1016/j.scitotenv.2016.09.009

Kojetin, D.J., Matta-Camacho, E., Hughes, T.S., Srinivasan, S., Nwachukwu, J.C., Cavett, V., Nowak, J., Chalmers, M.J., Marciano, D.P., Kamenecka, T.M., Shulman, A.I., Rance, M., Griffin, P.R., Bruning, J.B., Nettles, K.W., 2015. Structural mechanism for signal transduction in RXR nuclear receptor heterodimers. Nat. Commun. 6. https://doi.org/10.1038/ncomms9013

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a

program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291. https://doi.org/10.1107/S0021889892009944

Laskowski, R.A., Rullmann, J.A.C., 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

Liu, H., Li, P., Wang, P., Liu, D., Zhou, Z., 2019. Toxicity risk assessment of pyriproxyfen and metabolites in the rat liver: A vitro study. J. Hazard. Mater. 121835. https://doi.org/10.1016/j.jhazmat.2019.121835

Md Meftaul, I., Venkateswarlu, K., Dharmarajan, R., Annamalai, P., Megharaj, M., 2020. Pesticides in the urban environment: A potential threat that knocks at the door. Sci. Total Environ. 711, 134612. https://doi.org/10.1016/j.scitotenv.2019.134612

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

Monneret, C., 2017. What is an endocrine disruptor? Comptes Rendus - Biol. 340, 403–405. https://doi.org/10.1016/j.crvi.2017.07.004

Nakagawa, Y., Henrich, V.C., 2009a. Arthropod nuclear receptors and their role in molting. FEBS J. 276, 6128–6157. https://doi.org/10.1111/j.1742-4658.2009.07347.x

Nakagawa, Y., Henrich, V.C., 2009b. Arthropod nuclear receptors and their role in molting 276, 6128–6157. https://doi.org/10.1111/j.1742-4658.2009.07347.x

Neov, B., Georgieva, A., Shumkova, R., Radoslavov, G., Hristov, P., 2019. Biotic and abiotic factors associated with colonies mortalities of managed honey bee (Apis mellifera). Diversity 11, 1–16. https://doi.org/10.3390/d11120237

Oro, A.E., McKeown, M., Evans, R.M., 1990. Relationship between the product of the Drosophila ultraspiracle locus and the vertebrate retinoid X receptor. Nature 347, 298–301. https://doi.org/10.1038/347298a0

Parks, D.J., Blanchard, S.G., Bledsoe, R.K., Chandra, G., Consler, T.G., Kliewer, S.A., Stimmel, J.B., Willson, T.M., Zavacki, A.M., Moore, D.D., Lehmann, J.M., 1999. Bile acids: Natural ligands for an orphan nuclear receptor. Science (80-.). 284, 1365–1368. https://doi.org/10.1126/science.284.5418.1365

Patel, V., Pauli, N., Biggs, E., Barbour, L., Boruff, B., 2021. Why bees are critical for achieving sustainable development. Ambio 50, 49–59.

https://doi.org/10.1007/s13280-020-01333-9

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera - A visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. https://doi.org/10.1002/jcc.20084

Phillips, J.C., Hardy, D.J., Maia, J.D.C., Stone, J.E., Ribeiro, J. V., Bernardi, R.C., Buch, R., Fiorin, G., Hénin, J., Jiang, W., McGreevy, R., Melo, M.C.R., Radak, B.K., Skeel, R.D., Singharoy, A., Wang, Y., Roux, B., Aksimentiev, A., Luthey-Schulten, Z., Kalé, L. V., Schulten, K., Chipot, C., Tajkhorshid, E., 2020. Scalable molecular dynamics on CPU and GPU architectures with NAMD. J. Chem. Phys. 153. https://doi.org/10.1063/5.0014475

Rosas, L.G., Eskenazi, B., 2008. Pesticides and child neurodevelopment. Curr. Opin. Pediatr. 20, 191–197. https://doi.org/10.1097/MOP.0b013e3282f60a7d

Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: A unified platform for automated protein structure and function prediction. Nat. Protoc. 5, 725–738. https://doi.org/10.1038/nprot.2010.5

Sabarwal, A., Kumar, K., Singh, R.P., 2018. Hazardous effects of chemical pesticides on human health–Cancer and other associated disorders. Environ. Toxicol. Pharmacol. 63, 103–114. https://doi.org/10.1016/j.etap.2018.08.018

Sasorith, S., Billas, I.M.L., Iwema, T., Moras, D., Wurtz, J.M., 2002. Structure-based analysis of the ultraspiracle protein and docking studies of putative ligands. J. Insect Sci. 2. https://doi.org/10.1673/031.002.2501

Sullivan, J.J., Goh, K.S., 2008. Environmental fate and properties of pyriproxyfen. J. Pestic. Sci. 33, 339–350. https://doi.org/10.1584/jpestics.R08-02

Trott, O., Olson, A.J., 2009. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31, NA-NA. https://doi.org/10.1002/jcc.21334

Truman, J.W., 2019. The Evolution of Insect Metamorphosis. Curr. Biol. 29, R1252–R1268. https://doi.org/10.1016/j.cub.2019.10.009

Velarde, R.A., Robinson, G.E., Fahrbach, S.E., 2006. Nuclear receptors of the honey bee: Annotation and expression in the adult brain. Insect Mol. Biol. 15, 583–595. https://doi.org/10.1111/j.1365-2583.2006.00679.x

Wang, H., Chen, J., Hollister, K., Sowers, L.C., Forman, B.M., 1999. Endogenous Bile Acids Are Ligandsfor the Nuclear Receptor FXR/BAR 3, 543–553.

Wang, N., Zou, Q., Xu, J., Zhang, J., Liu, J., 2018. Ligand binding and heterodimerization

with retinoid X receptor (RXR) induce farnesoid X receptor (FXR) conformational changes affecting coactivator binding. J. Biol. Chem. 293, 18180–18191. https://doi.org/10.1074/jbc.RA118.004652

Wang, Y.D., Chen, W.D., Moore, D.D., Huang, W., 2008. FXR: A metabolic regulator and cell protector. Cell Res. 18, 1087–1095. https://doi.org/10.1038/cr.2008.289

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, 407–410. https://doi.org/10.1093/nar/gkm290

Wilson, T.G., 2004. The molecular site of action of juvenile hormone and juvenile hormone insecticides during metamorphosis: How these compounds kill insects. J. Insect Physiol. 50, 111–121. https://doi.org/10.1016/j.jinsphys.2003.12.004

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2014. The I-TASSER Suite: protein structure and function prediction. Nat. Methods 12, 7–8. https://doi.org/10.1038/nmeth.3213

Yoshino, H., Kaneko, H., Nakatsuka, I., Yamada, H., 1996. Metabolism of pyriproxyfen. 3. In vitro metabolism in rats and mice. J. Agric. Food Chem. 44, 1578–1581. https://doi.org/10.1021/jf950510q

Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9, 1–8. https://doi.org/10.1186/1471-2105-9-40

Zheng, W., Lu, Y., Tian, S., Ma, F., Wei, Y., Xu, S., Li, Y., 2018. Structural insights into the heterodimeric complex of the nuclear receptors FXR and RXR. J. Biol. Chem. 293, 12535–12541. https://doi.org/10.1074/jbc.RA118.004188

CHAPTER 8

General Discussion and Conclusions

The scientific evidences show that that many substances present in food can interfere with hormone levels or/and their action inducing toxic effects on human and animals health. These substances, called endocrine disrupting chemicals (EDCs), are exogenous molecules that are able to interfere with the function of hormonal systems and produce a range of diseases in humans and animals. Most EDCs are chemicals produced by industry and released into the environment, such as plasticized, polychlorinated biphenyls (PCBs), and dioxins, but some can also be produced by plants or fungi, such as mycotoxins. The group of molecules acting as EDCs is highly heterogeneous and they can affect the endocrine systems of an organism in several ways by i) mimicking natural hormones, ii) antagonizing their action, or iii) modifying their synthesis, metabolism, and transport through their interference with cellular targets, such as nuclear receptors. Due to the hazardous effects of EDCs on human and animals health, experts, international organizations, and scientists all over the world have been made to explore and decipher the entire class of endocrine disruptors, from legislative to molecular point of view. However, testing all the possible endocrine disruptors against all the potential targets related to endocrine disruption is an important but also expensive, long, and difficult, and sometimes impossible, task, also due to the limited availability of suitable bioassays. Therefore, new approaches are needed to help to identify potentially harmful chemicals for human and animal health as quickly as possible. One of these is represented by computational methods applied to EDCs research, either in the identification of new EDCs or pointing in the right direction when the mechanism of action is already known.

In this context, the aim of this PhD thesis has been to detect the possible endocrine disruptors in food using *in silico* methods. The first part of this thesis is addressed at understanding and investigating the role of *in silico* methods in food safety to predict the interaction between food contact chemicals and different targets involved in human diseases and decipher their mechanism of biding. The second part is focused on giving several applications of these methods to determine the possible binding of these chemicals with nuclear receptors.

In Chapter 3, the use of docking techniques as an alternative to animal tests is illustrated using recently published cases studies. Computational methods, such as repository or database design, virtual screening, and molecular docking, are well-known approaches in medicinal chemistry, but they are a relatively recent challenge in the food science field. In fact, these techniques can be applied to a wide range of food safety problems, such as

the discovery of new possible food contaminants. Moreover, *in silico* methods can be used to study the interaction of a food chemical with a protein receptor involved in human diseases and/or to decipher and understand their mechanism of binding, or the competition between a natural ligand of a protein and a food molecule, or to design of chemosensors that are able to include a toxin in a cavity to take away the hazardous molecules from water or food, or to decipher the activity of a dimer against a monomer. An important milestone for screening/docking approaches is the availability of a 3D Database to collect the huge amount of food contact chemicals to make it possible to test these compounds otherwise unfeasible with traditional *in vitro* tests. All these techniques can be used to control food contaminants and to prevent foodborne diseases. In fact, ensuring food safety is a significant challenge to protect public health. In this chapter, computational techniques are explained in order to give an overview of their utility in food safety challenges.

In Chapter 4, alternative animal tests are discussed that can be used to screen food contact chemicals against nuclear receptors and to evaluate their possible endocrine disruption. At the beginning of this chapter, a description of nuclear receptors has been given in order to understand their structure, function, and their associated diseases. Firstly, an overview of *in vitro* bioassays, such as ligand binding assays, isothermal titration calorimetry, differential scanning fluorimetry, and reporter gene assays, currently accepted by different agencies involved in this field, are given. Although *in vitro* studies are common usage for screening endocrine disrupting compounds, the huge amount of food contact chemicals highlights the importance of alternative methods, such as computational techniques, that can predict EDCs in a faster, safer, and better way. For this reason, in the second part of this chapter, the *in silico* methods, such as ligand-based virtual screening, molecular docking, consensus scoring, and molecular dynamics simulation, are discussed. Finally, in the last section, some real case studies in which *in silico* methods are applied together with the *in vitro* tests are illustrated in order to show how these methods can be used together for detecting endocrine disrupting chemicals in food and for preventing the relate foodborne diseases.

After providing the basic information for understanding computational methods uses, nuclear receptors structures and functions, the relative diseases, and food contact chemicals, the following chapters want to give real examples of the application of these concepts.

In Chapter 5, a combination of highly curated database and computational methods (in particular molecular docking and consensus scoring) has been used in order to screen a large number of food contact chemicals against the nuclear receptors. In the last years, given the high number of chemicals produced (only in the EU the production of chemicals hazardous to health were 211 million tonnes in 2019), the possible negative effects on human and animal health are dramatically increased. Given the impossible to use *in vitro* tests for testing all the possible hazardous chemicals, in this manuscript, a possible solution is given in order to screen millions of compounds reducing time and cost. In fact, given the high number of chemicals potentially harmful to the human body, the traditional animal tests are not enough to allow an appropriate check in a short time and with little expense. *In silico* methods permit to screen a huge number of molecules in a very short time (from 0,5 seconds to 1/2 minutes depending on the molecular docking program). On the other hand, the computational methods do not want to replace the *in vitro* tests (and they cannot replace them), but they should be considered as useful and preliminary methods to screen a huge number of molecules in a short time and with low costs. In this way, only the substances positive to the screening will be tested with *in vitro* and *in vivo* methods. An important pillar of computational methods is the availability of highly curated databases from which to retrieve and download structural, chemical, and regulative information about a substance. In this context, this manuscript gives an example of the application of these two *in silico* approaches, Big Data, and screening, in order to demonstrate their reliability in the food safety field. Given the breadth of information about food contact chemicals, their retrieval is very challenging. To this end, the foodchem database has been designed to store different types of information, such as identification names, chemical-structures properties, three-dimensional structures, and a link to European authorities, in order to retrieve all food contact chemicals information in a single place, provide a tool to accelerate the computational method, and consider the available legislative information. This work may be very useful in several applications: for implementing the EU Chemicals Strategy for Sustainability, for developing novel, inherently non-hazardous materials, and for screening for likely hazardous chemicals, as well as screening for chemicals that likely have no hazard properties and which should be the focus of further investigations. These likely non-hazardous chemicals should be more widely used in products like food contact articles instead of those that are hazardous. Focus on this work, a huge number of heterogeneous molecules from a chemical and structural point of view are docked against 31 nuclear receptors in order to discover their

potential endocrine disrupting activity. To do this, a robust consensus scoring approach using two different docking software and four different scoring functions is used. The results show that more than 50% of food contact chemicals are good interactors of liver X receptor β (LXRβ) (the nuclear receptor with the highest number of food contact chemicals that fall in the high interactor group), 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). Moreover, considering the class of food contact chemicals which have a greater number of molecules able to interfere with the endocrine system, almost the totality of dioxins, furans, and PCBs molecules can interact with more than 15 nuclear receptors with high binding affinity, followed by pesticides and phthalates sub-classes.

In chapter 6, the same approach described in the previously chapter was applied to a small subset of data, the mycotoxins, a burning issue in the last years. The mycotoxins contamination of foods and feeds (every year over 25% of global agricultural products are contaminated by mycotoxins) is a significant problem. Mycotoxins are toxic secondary metabolites produced by fungi that can contaminate food before/after harvest or/and during storage as a result of crops infection. These metabolites have negative effects not only on humans and animals health (they cause acute and chronic toxic effects), leading to an increased loss of human and animal life, and health care and veterinary care costs but also on economic and industrial sectors (the agricultural and industrial losses are estimated around the billions of dollars). Several international strategies and regulations have been proposed in order to control the mycotoxins occurrence in food and feed. Unfortunately, these actions are not adequate to limit the overall mycotoxin dietary exposure, safeguarding human and animal health. Over 300 mycotoxins have been identified and reported, but until now, risk assessments and toxicological studies have been carried out only on some major mycotoxins, such as aflatoxin, fumonisin, and zearalenone, while for all the rest, the toxicological and chemical studies are not sufficient to understand their effects. In this chapter, a possible strategy is given in order to deal with the rapid identification of mycotoxins as possible endocrine disruptor molecules. A preliminary evaluation of the endocrine disruptor activity of 328 mycotoxins is addressed using *in silico* structural approaches (molecular docking) with the support of statistical analysis. The results show that the majority of the mycotoxins are predicted as a possible good binder for all the considered nuclear

receptors. In particular, six of them, ochratoxin A, zearalenone, α- and β-zearalenol, aflatoxin B1, and alternariol, have shown to be putative endocrine disruptors chemicals for nuclear receptors, such as the pregnane X receptor (PXR), farnesoid X receptor (FXR), and liver X receptor (LXR), as highlighted from different studies. This preliminary study provides a relatively rapid and inexpensive method highlighting that can be seen as a powerful tool for a preliminary endocrine disruption evaluation of a large number of compounds.

In the last chapter (Chapter 7), an emerging problem of these last years is discussed: the growing bees mortality in the world due to pesticides. Bees, mainly *Apis mellifera*, are the major pollinators of wild plants and crops in terrestrial ecosystems and for these reasons, they are essential for humans. These pollinators are declining as a result of different factors, such as habitat loss and degradation, climate changes, pesticides, pathogens, and others. In the last years, a lot of bees massacres occurred due to the pesticides on the fields. Just to give one example, in 2019 about ten million bees died due to pesticides used for the cultivation of corn in Lombardia. But not only bees can die for the use inappropriate of pesticides: nearly 300,000 non-invasive species and humans die worldwide every year. According to a joint report of the World Health Organization (WHO) and the United Nations Environment Programme (UNEP), roughly 200,000 people in the world die, and around three million are poisoned each year by pesticides. In this chapter, the attention is brought to two pesticides, pyriproxyfen and 4'-OH-pyriproxyfen, which are insect growth regulators (IGRs) and active ingredients of several insecticides. Toxicological studies are not sufficient to demonstrate that pyriproxyfen is an endocrine disruptor for mammals. Otherwise, in the case of bees, the proofs indicated a high risk for the larva. In fact, pyriproxyfen can block the development of larvae and thus increases mortality, or, in the case of sublethal doses, can affect the behavior of bees and create malformations at the adult organism level. Given the above, in this chapter, the possible negative effects on human and bees health were investigated applying molecular modeling techniques. In particular, the mechanism of binding of pyriproxyfen and its metabolite, 4'-OH-pyriproxyfen, against ultraspiracle protein-ecdysone receptor (USP-EcR) dimer of *A. mellifera* and the relative heterodimer farnesoid X receptor-retinoid X receptor alpha FXR-RXRα) of *H. sapiens* were investigated. Once the USP and the EcR models were built and the interactions in the USP-EcR and RXRα-FXR dimers interface were studied, molecular docking has been carried out, in order to predict

and evaluate the structural physical interactions between the receptor and the pesticides. In the end, the mechanism of binding of pyriproxyfen and its metabolite with USP-EcR dimer and FXR-RXRα were studied using molecular dynamic simulations. The results revealed that pyriproxyfen and its metabolite, the 4'-OH-pyriproxyfen, stabilize both bees and human dimer. More in detail, the molecular dynamics simulations show how both pyriproxyfen and the 4'-OH-pyriproxyfen are stable inside the binding pockets of the dimers. Comparing the simulations, the dimers binding the two compounds, pyriproxyfen and the 4'-OH-pyriproxyfen, and the dimers binding the natural ligand, juvenile hormone III and 9-cis-retinoic acid. In conclusion, these *in silico* analyses reveal a possible interaction of the two compounds, pyriproxyfen and 4'-OH-pyriproxyfen, with RXRα-FXR and USP-EcR dimers demonstrating the probable endocrine interference of action of these two pesticides against bees and humans.

Conclusions

Food is made up of chemicals. These chemicals have several origins. They may be naturally occurring, such as toxins produced by plants, animals, or microorganisms, intentionally added, such as food additives or flavourings, and unintentionally added, such as pesticides that accidentally contaminate the food being processed or food contact chemicals that come into contact with food. Chemicals in food are generally harmless and present at very low levels. However, chemicals can have a variety of toxicological properties which can cause negative effects on human and animal health. As a result, in order to protect consumers from exposure to potentially hazardous substances, national and international food safety standards are developed based on credible scientific risk assessments that define "safe" exposure levels. In addition to the active/parent drug, which is generally well-characterized in terms of bioavailability and toxicological qualities, the consumer is exposed to a wide spectrum of compounds emerging from metabolic and degradation processes. In the majority of cases, very little information is available on the toxicological characteristics of these substances. Alternative (nonanimal) assessment methodologies are required to support analyses of the toxicological profile of chemicals in food, including metabolites and degraded substances. Computational techniques that predict bioavailability and toxicity based on chemical structure are particularly appealing for cost-effectiveness, efficiency, and animal welfare. Computational toxicology is a fast-developing field that is receiving increasing attention

from national and international bodies such as the European Commission and the European Food Safety Authority (EFSA).

In this Doctoral Thesis, it has been demonstrated how computational approaches may aid particularly in the early stages when there is little understanding of the active components, molecular target, or mechanism of action. These techniques allow for the analysis of larger sets of molecules, such as the virtual screening used to select molecules of dataset repositories identifying the molecules which may potentially have a harmful activity. Based on this knowledge, these compounds may be analysed with subsequent *in vitro* and *in vivo* studies. Computational methods can help to elucidate the mechanisms of action of molecules of interest by giving information on molecule binding modes with macromolecular targets. Moreover, *in silico* approaches are useful for predicting how a molecules system will evolve following changes in parameters such as temperature or pressure, and how these modifications can affect the studied molecules. Nonetheless, the use of computational techniques in food safety is still at an early stage but with great potential if they are viewed as future challenges and opportunities in food and feed safety fields.

Work in progress

During the PhD period, some new collaborations and researches were raised still ongoing.

1. Consensus scoring combined with machine learning to identify potential endocrine disruptors among food contact chemicals (Manuscript in preparation).
   In this work, in collaboration with Martyn T. Smith (Professor of Toxicology and the Kenneth Howard and Marjorie Witherspoon Kaiser Endowed Chair in Cancer Epidemiology at University of California, Berkeley) and his teams, Jane Muncke (Managing Director and Chief Scientific Officer of the Food Packaging Forum, Zurich, Switzerland), and Ksenia Groh (Group Leader in Bioanalytics, Department of Environmental Toxicology, Eawag, Dübendorf, Switzerland), a combination of Machine Learning (ML) and molecular docking approaches were applied to identify potential endocrine disruptors among 4847 food contact chemicals. Firstly, a predictive tool, NR-Toxpred, was used to screen 4847 chemicals against nine different nuclear receptors in order to predict their binding potential. 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 that modulate nuclear receptors. NR-Toxpred uses machine learning to classify the given chemical as a binder or non-binder for a chosen receptor, and for nine of the receptors, it predicts whether the active binder is an agonist or antagonist. Secondly, molecular docking and consensus scoring approaches were used to identify the substances of very high concern (SVHCs). Specifically, two docking software and four scoring functions have been applied to reach a statistical consensus prediction obtained using MATLAB software.
2. Thermodynamic characterization of the interaction between retinoid X receptor alpha (RXRα) and pyriproxyfen by Isothermal Titration Calorimetry (ITC).
   The aim of this research, in collaboration with Professor Emilia Fisicaro (Department of Food and Drug, University of Parma) and Professor Angelo Bolchi (Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma), was to characterize the thermodynamics of RXRα-pyriproxyfen interactions. Isothermal titration calorimetry (ITC) is a physical technique used to determine the thermodynamic parameters of interaction, including the changes in free energy, enthalpy, entropy, and heat capacity (Wright, Vincent, and Fernandez 2007). This research is only at the beginning. Firstly, the expression vector of human RXRα LBD (residues 220-462)

was purchased and transfected into *Escherichia coli* cells. At the end, the protein was purified using His-tag reaching a concentration of 121 uM. The buffer utilized to stock the protein contained 25 mM Tris, pH 7.5, and 0.15 M NaCl. A first ITC experiment was performed but a first problem was found. In fact, the pyriproxyfen is only soluble in organic solvents, such as ethanol and dimethyl sulfoxide (DMSO), but the first induces high heat change and the second denatures the protein. Several experiments must be performed specially to find the correct conditions of analysis.

References

Valsecchi, C., Grisoni, F., Motta, S., Bonati, L., & Ballabio, D. (2020). NURA: A curated dataset of nuclear receptor modulators. *Toxicology and Applied Pharmacology*, *407*(August), 115244. https://doi.org/10.1016/j.taap.2020.115244

Wright, Edward, Jeremy Vincent, and Elias J. Fernandez. 2007. "Thermodynamic Characterization of the Interaction between CAR-RXR and SRC-1 Peptide by Isothermal Titration Calorimetry." *Biochemistry* 46 (3): 862–70. https://doi.org/10.1021/bi061627i.

About the author

Name: Giulia
Surname: Spaggiari
Home address: Via Guglielmo Marconi, 100, Felino, Parma, Italy
Telephone number: +393495140032
E-mail: giulia.spaggiari@outlook.it

Giulia Spaggiari was born on 15th June 1994 in Montecchio Emilia (Italy) and she has performed her academic studies at the University of Parma obtaining both bachelor's and master's degrees in Food Science and Technology. 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 Professor Pietro Cozzini. Her research focus was on the detection of endocrine disruptors in food using *in silico* methods.

She is especially fascinated by the complexity of the molecular systems underlying the food and the interactions between it and our body, hidden behind gestures as simple as drinking a glass of water. These systems power our lives but require all our efforts to turn into safe. Here is where she wants to put her energy, taking every step with excitement, and find the best solutions.

Publications

- **Published**
  - Spaggiari, G., Morelli, G., Riani, M., Cozzini, P., 2022. A synergism of *in silico* and statistical approaches to discover new potential endocrine disruptor mycotoxins. Toxicol. Appl. Pharmacol. 435, 115832. https://doi.org/10.1016/j.taap.2021.115832
  - Cozzini, P., Cavaliere, F., Spaggiari, G., Morelli, G., Riani, M., 2021. Computational methods on food contact chemicals: Big data and *in silico* screening on nuclear receptors family. Chemosphere 292. https://doi.org/https://doi.org/10.1016/j.chemosphere.2021.133422

- Cavaliere, F., Spaggiari, G., Cozzini, P., 2021. Computational Applications on Food Contact Chemicals as Nuclear Receptor Binders. In: Badr M.Z. (eds) Nuclear Receptors. Springer, Cham. https://doi.org/10.1007/978-3-030-78315-0_12
- Spaggiari, G., Cavaliere, F., Cozzini, P., 2021. The Application of *in silico* Methods on Umami Taste Receptor. In: . Handbook of Experimental Pharmacology. Springer, Berlin, Heidelberg. https://doi.org/10.1007/164_2021_515
- 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
- Cavaliere, F., Spaggiari, G., Cozzini, P., 2021. Molecular docking: a contemporary story about food safety. In M. S. Coumar (Ed.). Molecular Docking for Computer-Aided Drug Design: Fundamentals, Techniques, Resources and Applications (pp. 479-492). Pondicherry, India. https://doi.org/10.1016/B978-0-12-822312-3.00025-4
- Verhagen, H., Alonso-Andicoberry, C., Assunção, R., Cavaliere, F., Eneroth, H., Hoekstra, J., Koulouris, S., Kouroumalis, A., Lorenzetti, S., Mantovani, A., Menozzi, D., Nauta, M., Poulsen, M., Rubert, J., Siani, A., Sirot, V., Spaggiari, G., Thomsen, S. T., Trevisan, M., Cozzini, P., 2021. Risk-benefit in food safety and nutrition – Outcome of the 2019 Parma, 141. https://doi.org/10.1016/j.foodres.2020.110073
- Spaggiari, G., Di Pizio, A., Cozzini, P., 2020. Sweet, umami and bitter taste receptors: State of the art of *in silico* molecular modeling approaches. Trends in Food Science and Technology, 96, 21–29. https://doi.org/10.1016/j.tifs.2019.12.002

Courses, conferences, and workshops:

- Parma Summer School "Food Safety Aspects of Integrated Food Systems" - Parma - 28-30 September 2021
- IPAM *in silico* - 2021 - Corso teorico-pratico sull'applicazione dei metodi computazionali nel Replacement - Italian Platform on Alternative Methods - IPAM - 02 July 2021 - 16 July 2021 - 10 September 2021 - 17 September 2021

- High-Performance Molecular dynamics - Cineca, Bologna - 7-9 April 2021
- Intensive School for Advanced Graduate Studies Machine Learning - University of Pavia - 7-17 September 2020
- Parma Summer School “Risk-Benefit in Food Safety and Nutrition” - Parma - 11-13 June 2019
- Fundamental of programming language - Parma - 2018-2019

Teaching:

- IPAM *in silico* - 2021- Corso teorico-pratico sull’applicazione dei metodi computazionali nel Replacement - “MolDock – 1, modelling interactions by Molecular Docking - a tool for structure based drug discovery and toxicology” - Italian Platform on Alternative Methods - IPAM - 9 July 2021
- Teaching support of Molecular Modelling course - University of Parma
- Exams support of General Chemistry - University of Parma