Research Article

Bisphenol A Binding Promiscuity: A Virtual Journey through the Universe of Proteins

Elena Lo Piparo[1#], Lydia Siragusa[2#], Frederic Raymond[3], Giovanna Ilaria Passeri[2], Gabriele Cruciani[4]
and Benoît Schilter[1]

[1]Food Safety Research, Nestlé Research, Lausanne, Switzerland; [2]Molecular Discovery Limited, London, UK; [3]Multi-Omics Profiling, Nestlé Research, Lausanne, Switzerland; [4]Department of Chemistry, Biology and Biotechnology, University of Perugia, Perugia, Italy

Abstract

Significant efforts are currently being made to move toxicity testing from animal experimentation towards human-relevant, mechanism-based approaches. In this context, the identification of the molecular target(s) responsible for mechanisms of action is an essential step. Inspired by the recent concept of polypharmacology (the ability of drugs to interact with multiple targets), we argue that whole proteome virtual screening may become a breakthrough tool in toxicology as it reflects the true complexity of chemical-biological interactions. We investigated the value of performing ligand-protein binding prediction screening across the full proteome to identify new mechanisms of action for food chemicals. We applied the new approach to make a broader comparison between bisphenol A (BPA) (food-packaging chemical) and the endogenous estrogen 17β-estradiol (EST). Applying a novel, high-throughput ligand-protein binding prediction tool (BioGPS) using the Amazon Web Services (AWS) cloud (to speed-up the calculation), we investigated the value of performing *in silico* screening across the full proteome (all human and rodent x-ray protein structures available in the Protein Data Bank). The strong correlation between *in silico* predictions and available *in vitro* data demonstrated the high predictive power of the method. The most striking result was that BPA was predicted to bind to many more proteins than previously described, most of which also appear to bind EST. Our findings provide a new and unprecedented insight into the complexity of chemical-protein interactions, highlighting the binding promiscuity of BPA and its broad binding similarity to the female sex hormone, EST.

1 Introduction

Our study exploits recent discoveries in supercomputer-based drug design to support the new paradigm under development in the field of toxicology. Indeed, exploring alternative approaches to animal experimentation is inducing a revolution, moving from black box high-dose animal studies towards more mechanistic, human-relevant approaches. This is illustrated by the concept of adverse outcome pathways (AOPs), consisting of a framework linking molecular initiating events (MIEs) (based on chemical-target interactions) to a sequence of molecular, cellular, anatomical and functional events, leading to an adverse effect (Burden et al., 2015). It is thought that AOPs will transform risk assessment of chemicals, placing more emphasis on the integration of mechanistic information. In this context, when a MIE

for a compound is known, introducing ligand-protein binding prediction can be helpful to assist toxicologists in the assessment of similar compounds (e.g., very important for risk assessment of mixtures). But, as reported for drugs, for which polypharmacology, i.e., the ability of drugs to interact with multiple targets, seems to be more a rule than an exception (Ellingson et al., 2014), xenobiotics may also interact with multiple targets, potentially triggering several MIEs that lead to adverse outcomes.

Following this hypothesis, we investigated the feasibility of applying a novel *in silico* workflow based on the virtual screening of all characterized proteins to toxicology. This could represent a rapid method to identify MIEs and toxicological pathways that are potentially triggered by chemicals of interest. This idea is in line with other strategies developed under large research initiatives such as U.S. Tox21 (Thomas et al., 2018) and ToxCast (Richard et al.,

---

# contributed equally

2016), which are based on *in vitro* screening using hundreds of bioassays to identify a large spectrum of potential biological effects.

To investigate the value of our approach in addressing toxicological questions, the comparison of bisphenol A (BPA) with endogenous estrogen (EST) was selected as a proof-of-concept. BPA was chosen because it is well documented to interact with a number of protein targets, which could then serve as a basis to assess the performance of BioGPS (Siragusa et al., 2015, 2016) to predict potential MIE(s). Moreover, because of its structural similarity to the endogenous hormone estradiol (EST), the endocrine-disrupting effects of BPA have been widely discussed (Rubin, 2011). This feature inspired the search for its toxic effect uniquely through estrogen receptor (ER) activation, and consequently most BPA alternatives were designed to reduce this activity. However, a number of other, diverse targets of BPA have been discovered, suggesting the involvement of other mechanisms and potentially further MIEs (MacKay and Abizaid, 2018). Indeed, as recommended by EFSA (EFSA, 2015), more mechanistic studies to determine the modes of action of BPA are needed. Although our approach is not quantitative and only evaluates binding affinity, we believe it can be a good starting point to design further *in vitro* tests.

The main questions we addressed were whether BioGPS is suitable to identify the different known targets of BPA and how the broad *in silico* protein-binding profile of BPA compares to EST. To answer these, we widened our perspective to assess the complete 3D proteome by performing a virtual screening analysis of all the human and rodent 3D structures that were available in the Protein Data Bank (PDB) (Berman et al., 2000), the central storehouse of biomolecular structures (30,153 structures, October 2014).

To be as exhaustive as possible, we investigated not only active sites in the proteins but also secondary pockets, i.e., binding sites on the protein other than the active site. Indeed, it is increasingly evident that secondary pockets may play a role in biological responses (Liu and Nussinov, 2016; Lu et al., 2014). From a drug discovery viewpoint, the exploration of secondary sites contributes to the

design of innovative therapeutic solutions (Tschammer, 2016; Livingston and Traynor, 2018; Citro et al., 2016; Kasbekar et al., 2016). Following the same reasoning, it appears that focusing attention only on "main" active sites may be insufficient to understand the full toxicological effects, and considering the interaction with "secondary" sites may provide key elements to understanding toxic phenomena.

In silico-in vitro validation was performed using freely available information. Furthermore, to better compare BPA to EST, we generated the protein binding profile and calculated the *in silico* binding promiscuity index, defined as the percentage of proteins predicted to bind the substance of interest versus the total number of proteins tested, the general hypothesis being that the more proteins a compound interacts with, the greater is the likelihood of observing adverse effects in toxicological studies.

To summarize, the proposed approach is based on the sequential application of:

− BioGPS (Global Positioning System in Biological Space) software (Siragusa et al., 2015, 2016), implementing the virtual screening of a compound against all x-ray protein structures available in the PDB (Berman et al., 2000);
− Validation of the *in-silico* ligand-binding prediction by *in vitro* data available at PubChemBioassay platform (Wang et al., 2017) and at Comparative Toxicogenomics Database (Davis et al., 2018);
− Calculation of the *in-silico* binding promiscuity index;
− Grouping the predicted ligand-protein interactions into different families to generate the protein-binding fingerprint profile.

2 Methods

Data collection

We collected protein structures belonging to four organisms: *Homo sapiens*, *Rattus rattus*, *Rattus norvegicus* and *Mus musculus* from the PDB (Berman et al., 2000). The total number of protein structures was 30,153 (October 2014), belonging to 6,309 unique proteins. "Protein structures" refers to structures downloaded from the PDB, while "proteins" are the biological entities extracted, starting from the PDB code, from the Uniprot (Nightingale et al., 2017) database.

Cavity database generation

The BioGPS[1] protocol (Siragusa et al., 2015, 2016) for detecting and characterizing cavities was applied to each of the 30,153 structures. First, the fixpdb tool was used to prepare the input protein structure for subsequent analysis. Protein residues, solvent molecules, co-crystallized ligands, cofactors and ions contained in the PDB structure file were also properly processed. All nucleic acids, ligands, and water molecules co-crystallized with the protein were removed, while cofactors were retained (i.e., NAD, FAD, GSH). Additionally, using a defined GRID energy threshold (Goodford, 1985), Cu^{+2}, Fe^{+2}, Zn^{+2}, Mg^{+2} ions exhibiting interactions with the protein residues (excluding those ions that are artefacts) were retained.

---

[1] http://biogps.org
[2] https://www.moldiscovery.com

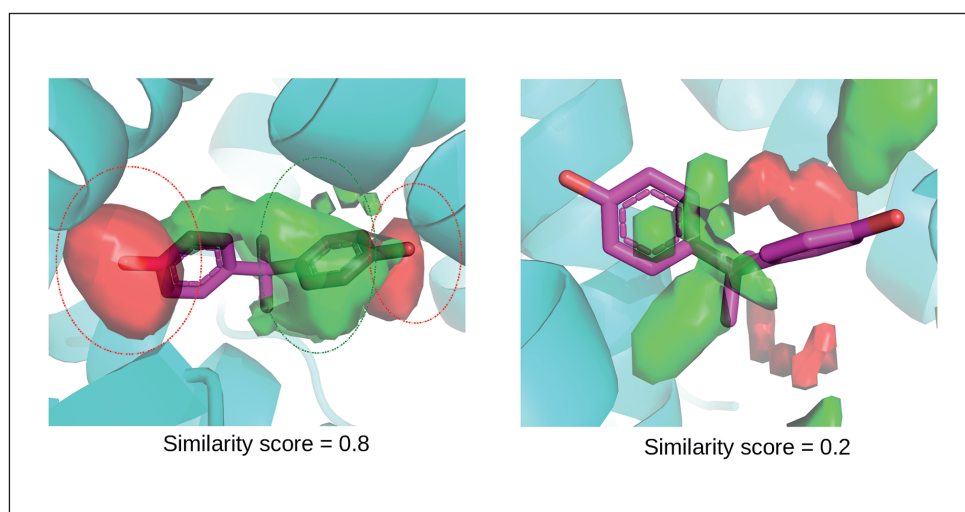Similarity score = 0.8   Similarity score = 0.2

Fig. 1: Examples of BPA fitting into a cavity with a high (left) or low (right) complementarity score
Red and green pockets indicate areas where the hydrogen bonding acceptor and hydrophobic moieties, respectively, are favorably hosted. Red dashed circles highlight that BPA-OH groups fit well into hydrogen bonding acceptor MIFs of the binding site. Green dashed circle indicates good complementarity of BPA hydrophobic moieties in hydrophobic MIFs of the binding site.

Then, the flapsite tool was applied for detecting protein cavities. In total, 199,436 cavities were collected. For each detected cavity, the molecular interaction fields (MIFs) (Baroni et al., 2007) were calculated using the default GRID probes (H (shape), DRY (hydrophobic interactions), O (H-bonding donor interactions) and N1 (H-bonding acceptor interactions). The resulting MIFs were then reduced in complexity to form a "common reference framework". These were stored in the BioGPS database to perform a pair-wise comparison with one or more template structures. By using 288 cpus (72 nodes each of 4 cpus) on Amazon Elastic Compute Cloud (Amazon EC2), a web service that provides resizable computing capacity in the cloud, the cavity database creation took 20 days.

Virtual screening
Bisphenol-A (BPA) and 17β-estradiol (EST) were used as template molecules for structure-based virtual screening against all cavities. The FLAP[2] structure-based approach (Baroni et al., 2007; Cross et al., 2010; Brincat et al., 2011; Sirci et al., 2012; Siragusa et al., 2016) was applied to evaluate the MIF volume complementarity between the template molecule and the cavity MIF. The MIF superposition is reported as a complementarity score: the higher the score, the better the ligand fits into the cavity. Score values range between 0 and 1, where 0 indicates no complementarity between ligand and cavity MIF, and 1 indicates that the ligand fits the cavity with the maximum complementarity.

Figure 1 reports an example where BPA fits well into a cavity (left) and one where the complementarity is not favorable (right). The calculated complementarity score indicates the probability that a molecule fits into a cavity, but no information on the entity of the interaction (or the activity) is contained in the score.

BPA and EST were screened against the entire database of MIF cavities (199,436 cavities). By using 288 cpus (72 nodes each of 4 cpus), the BPA and EST virtual screening took 3 days on Amazon Elastic Compute Cloud (Amazon EC2). The score values were considered for ranking the database cavities according to their complementarity with the two templates. An *in-silico* activity profile was obtained for both molecules. For each protein (identified through the Uniprot code) only the cavity with the maximum score was considered. By setting a score threshold of 0.6 to discriminate between active and inactive targets (proteins bound and not bound by the two templates), we identified 2,773 active proteins for BPA and 1,608 active proteins for EST.

PubChem bioassay
Biological assay results for small molecules are stored in the PubChem BioAssay (Wang et al., 2017) database. A PubChem BioAssay data entry includes an assay description, the corresponding target, the substances tested, the experimental values (Ki, IC_{50}, EC_{50}), and their activity (chemical probe, active, inactive, inconclusive and unspecified). We searched for biological activity data for BPA and EST. We selected as positives those targets for which BPA or EST results were "active" and where the activity outcome was derived from dose response curves following tests with multiple concentrations ("confirmatory bioassays"). We selected as negatives those targets for which BPA or EST results were "inactive". Among these proteins, we considered only those having deposited human, rat or mouse x-ray structures in the PDB.

The Comparative Toxicogenomic Database (CTD)
The CTD (Davis et al., 2018) supplies information on chemical-gene and chemical-protein interactions and their relationships with medical conditions. Data are curated by experts from published literature. We searched for BPA and EST bioactivity. A list of chemical-gene/protein interaction is available with the following designations: any, abundance, activity, binding, cotreatment. Only binding type interactions were selected, and it was manually verified that they involved a direct link between one chemical and one protein.

Protein classification and protein binding fingerprint profile
We collected the proteins with positive results in the virtual screening analysis for BPA or EST (2,773 and 1,608, respectively). By using the Panther classification system (Thomas et al., 2003; Mi et al., 2010), we grouped these proteins into different classes. We first computed the total number of proteins in the
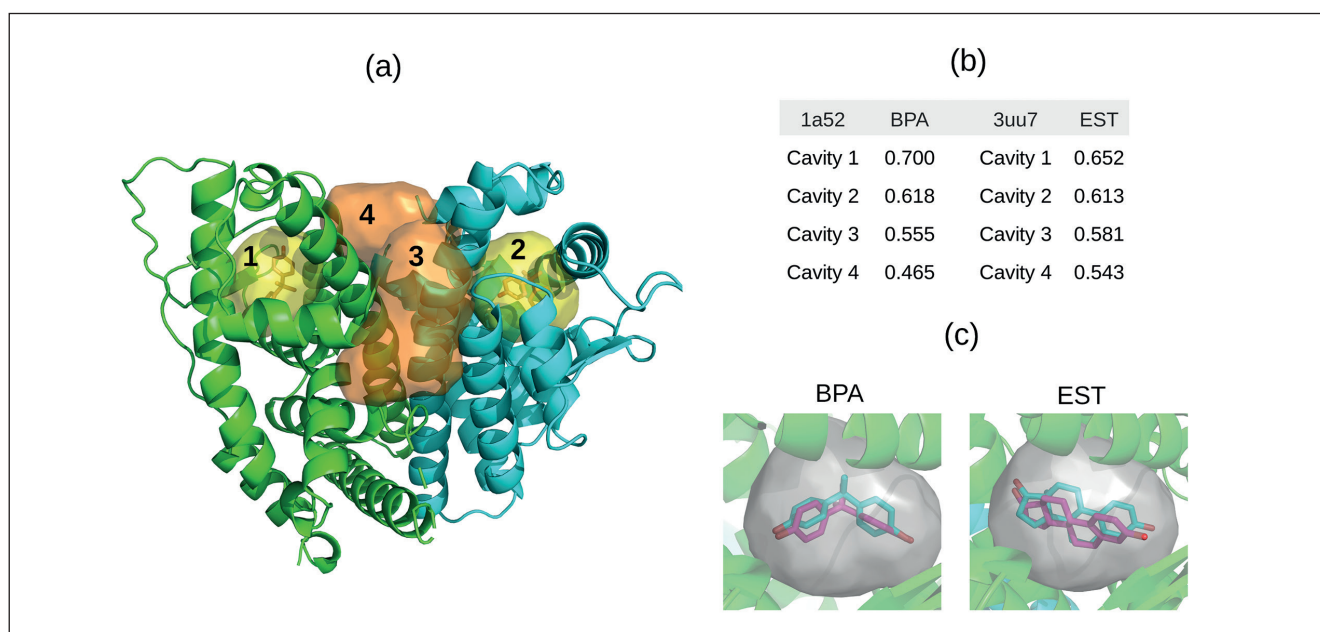
(a)

(b)

1a52	BPA	3uu7	EST
Cavity 1	0.700	Cavity 1	0.652
Cavity 2	0.618	Cavity 2	0.613
Cavity 3	0.555	Cavity 3	0.581
Cavity 4	0.465	Cavity 4	0.543

(c)

BPA   EST

Fig. 2: Binding of BPA to the ERα
(a) Cavities detected on ERα (PDBcode: 3uu7). Monomers 1 and 2 are displayed in green and cyan, respectively. Cavities 1 and 2 are well-known and extensively studied BPA binding sites; the other cavities represent "secondary" sites discovered by BioGPS. (b) BioGPS complementarity scores of BPA/EST against ERα cavities. (c) Superposition of BPA and EST poses generated by BioGPS, and x-ray structure of ERα. X-ray ligands are displayed in cyan, BioGPS poses in magenta, the ERα [left PDB code: 3UU7; right PDB code: 1A52] in green, and the binding site in grey.

screened database for each protein class. We then calculated the number of proteins predicted to bind BPA or EST from each class. To avoid bias introduced by the distribution of families among the full dataset, we then computed the ratio between the number of positive and the total number of proteins for each class. We built fingerprint graphs for BPA and for EST, reporting the number and the percentage of proteins in each class.

3 Results

3.1 Broadening the space of virtual screening
In this work, we moved the magnifying glass from the well-known active sites to the unexplored rooms in protein structures by including in the virtual screening analysis all human and rodent 3D structures freely available in the PDB (Berman et al., 2000). Using BioGPS (Siragusa et al., 2015, 2016), we identified all potential pockets on each of the 30,153 protein structures belonging to 6,309 unique proteins, finding at total of 199,436 cavities, 15,857 of which contain a drug-like small molecule binding site (defined as "active site"). We used all the cavities detected for each of the 3D characterized structures for *in silico* screening to simulate ligand binding interactions in the whole organism.

As an illustrative example, we first assessed the known interaction between BPA or EST and the estrogen receptor α (ERα). As a first step, we detected the location of all cavities in the ERα structure co-crystallized with either BPA (PDBcode: 3uu7) or with EST (PDBcode: 1a52) (Delfosse et al., 2012). As shown in

Figure 2a, ERα dimer cavities show a specular architecture: cavities 1 and 3 on one monomer (helix 1, in green) correspond to cavities 2 and 4 on the other monomer (helix 2, in cyan), respectively. Cavities 1 and 2 are well-known and extensively studied BPA binding sites; the other cavities represent "secondary" sites. BPA and EST were screened against all cavities detected in existing co-crystal structures of the ERα (BPA in 1a52, EST in 3uu7) to verify that BioGPS correctly identifies the true binding location. Scores reported in Figure 2b show that, as expected, EST and BPA scores were highest in the well-known binding sites (cavity 1 and 2).

As a further validation of the method, we extracted the superposition of BPA and EST in the ERα binding site generated by the BioGPS algorithm. The best pose for each of the two molecules was selected. We compared the two poses with the x-ray structures. The BioGPS algorithm was able to reproduce the x-ray position of the two ligands (Fig. 2c), suggesting that the MIFs overlap is correct, and thus can be used to detect putative new targets.

Interestingly, BioGPS also retrieved the space between the two monomers (cavities 3 or 4), in agreement with studies reporting likelihood of ER ligand binding at the interface of the dimer (Chakraborty et al., 2012). This confirms that unknown binding sites still can be discovered.

3.2 Bioactivity profiling of BPA and EST
BPA and EST were screened against the full panel of human, rat and mouse protein cavities collected from the PDB in order to identify all putative protein targets. This structure-based approach
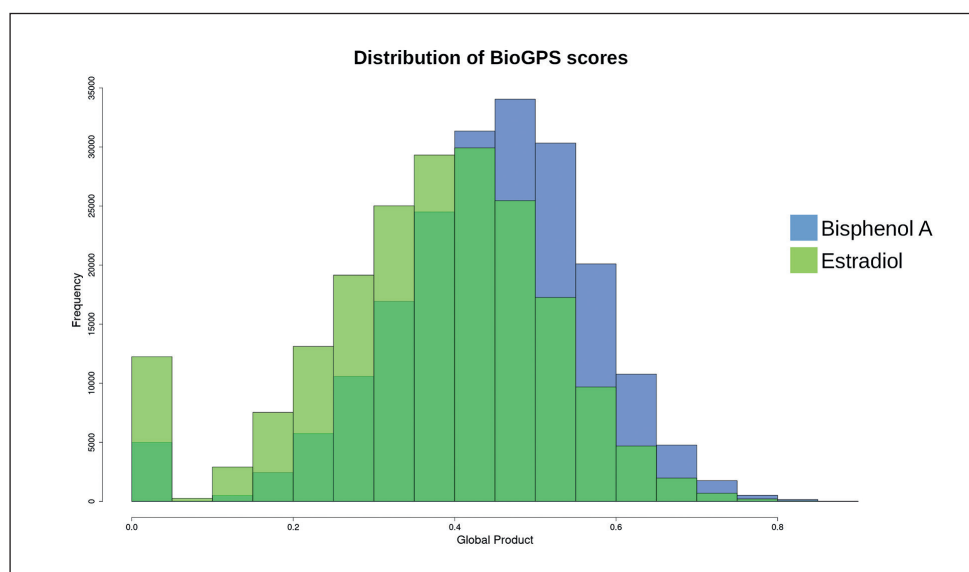
Distribution of BioGPS scores

Fig. 3: Distributions of complementarity score values obtained by virtual screening of BPA (blue bars) and EST (green bars) against all cavities

Tab. 1: Number of unique proteins, total number of structures available in the Protein Data Bank and number of cavities calculated through BioGPS together with the corresponding BPA and EST binding proteins, structures and cavities applying a cut-off of 0.6

	Total	Hits BPA	Hits EST
Proteins	6,309	2,773	1,608
Structures	30,153	10,503	5,336
Cavities	199,436	17,991	7,624

consists in evaluating the degree of complementarity that occurs between a ligand (here, BPA or EST) and a cavity: the higher the score, the better the molecule fits into the template cavity. The distribution of complementarity scores for BPA and EST is reported in Figure 3.

The distribution of complementarity score values of BPA was shifted to higher values (binding more proteins with higher scores) as compared with the distribution of values for EST, indicating a higher promiscuity of BPA as a ligand. This fact may be due to the smaller shape of BPA compared to EST and/or to its simpler structure, which may make it more easily adaptable to different cavities.

What we learned from this first analysis is that BPA and EST binding performances differ in a modest way, with the BPA active profile being more extended. In order to address this assumption, we selected from the panel those cavities that accommodate both molecules (BPA and EST) similarly to or better than ERα. To do this, we first calculated the complementarity scores obtained for BPA and EST for all ERα binding sites (Fig. S1[3]). For the two molecules, a value of 0.6 was obtained for the interaction with the

most studied pocket. Being our benchmark to identify other potentially relevant binding sites for BPA and EST, we retrieved all cavities presenting a score > 0.6.

17,991 and 7,624 cavities belonging to 2,773 and 1,608 proteins, respectively, were detected as putative binding sites for BPA and EST, respectively, (Tab. 1). Taking into account that many x-ray structures can be available for one protein (e.g., for human ERα the number of proteins is 1, the number of structures available in the PDB (October 2014) is 123, and the number of cavities is 405), the number of binding structures were, respectively 10,503 and 5,336. For the full list of screened proteins and BPA and EST hits, see supplementary file 2[4].

3.3 Computational standpoint meets *in vitro* data
To gain insights into how the putative BPA targets retrieved by the virtual screening compare to experimental data, we searched for proteins for which an interaction with BPA had been tested, i.e., experimentally active and inactive targets, in the PubChem Bioassay platform (Wang et al., 2017) and the CTD (Davis et al., 2018). Proteins for which no 3D structure was available were removed from the set. BPA was found to be experimentally active against 23 proteins and inactive against 82 proteins. Thus, for the computationally predicted proteins (about 6 thousand found either active or inactive), experimental data is only available for just over one hundred. The same strategy was applied to EST, finding 25 experimentally active targets and 92 inactive targets (Tab. 2).

To evaluate the robustness of the BioGPS classifier, we compared the *in vitro* and *in silico* activity. Figure 4 shows the receiver-operating characteristic (ROC) enrichment curves resulting from BioGPS scoring (true positive rate versus false positive rate) and the AUC (area under the curve). The AUC is a measure of the successful discrimination between known actives and inactives. The ROC curves clearly show good results of *in silico* predictions

Tab. 2: *In silico* binding prediction of the proteins known to be targeted by BPA and EST (only *in vitro* active are reported)
In green we reported proteins for which *in silico* predictions were confirmed by *in vitro* data, in red proteins for which *in silico* prediction was not confirmed by *in vitro* data.

Uniprot	Source	*In silico* prediction	Protein name
O15296	common	active	Arachidonate 15-lipooxygenase type B
O75469	common	active	Pregnane X nuclear receptor
P03372	common	active	Estrogen receptor 1
P04150	common	active	Glucocorticoid receptor
P04278	common	active	Sex hormone-binding globulin
P06401	common	active	Progesterone receptor
P07237	common	active	Protein disulfide-isomerase (PDI)
P10275	common	active	Androgen receptor
P10828	common	active	Thyroid hormone receptor beta
P11474	common	active	Estrogen related receptor alpha
P30561	common	active	Aryl hydrocarbon receptor
P37231	common	active	Peroxisome proliferator activated receptor gamma
P51449	common	active	RAR-related orphan receptor gamma
Q14994	common	active	Nuclear receptor subfamily 1 group I member 3 (CAR)
Q92731	common	active	Estrogen receptor 2
P01116	selective BPA	active	GTPase KRas
P02768	selective BPA	active	Serin albumin
P17787	selective BPA	active	Neuronal acetylcholine receptor subunit beta-2
P19652	selective BPA	active	Alpha-1-acid glycoprotein
P61769	selective BPA	active	Beta-2-microglobulin
P62508	selective BPA	active	Estrogen related receptor gamma
Q03181	selective BPA	active	Peroxisome proliferator-activated receptor delta (PPAR-delta)
Q16236	selective BPA	inactive	Nuclear factor erythroid 2-related factor 2
O95718	selective EST	inactive	Estrogen related receptor beta
P04637	selective EST	active	Tumor protein p53
P08185	selective EST	active	Serpin family A member 6
P08684	selective EST	active	Cytochrome P450 3A4
P14061	selective EST	inactive	Hydroxysteroid 17-beta dehydrogenase 1
P20813	selective EST	active	Cytochrome P450 family 2 subfamily B member 6
P49888	selective EST	active	Sulfotransferase family 1E member 1
P50172	selective EST	active	Corticosteroid 11-beta-dehydrogenase isozyme 1
P50225	selective EST	active	Sulfotransferase family 1A member 1
Q96KQ7	selective EST	inactive	Euchromatic histone-lysine N-methyltransferase 2

compared with *in vitro* data. In both cases, almost all true positives were correctly classified in the first 40% of the false positives.

Table 2 reports all experimentally positive targets. Protein hits are divided into: "common", i.e., those that were experimentally active for both BPA and EST, and "selective BPA" or "selective EST", those that were active only for one ligand and inactive or not determined for the other. Among the BPA selective targets,

nuclear factor erythroid 2 related factor 2 (Nrf2) was found to be false negative for BPA (*in vitro* active target predicted *in silico* as inactive). Searching deeply in the literature, we found that the induction of Nrf2 through BPA is considered to be mediated by nitrosylation of Keap1 (Nakamura et al., 2018), while direct binding to BPA is not reported. In conclusion, the percentage of correctly predicted active targets was 95% (23 out of 24) for BPA and
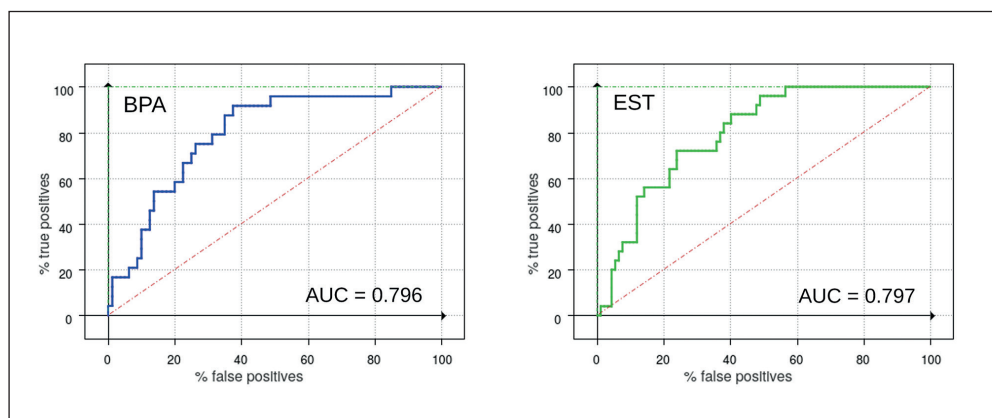
Fig. 4: Enrichment curve for BPA (blue) and EST (green)
False positive rate (x-axis) versus true positive rate (y-axis).

88% (22 out of 25) for EST. 32% and 54% of the inactive targets were correctly predicted for BPA and EST, respectively (see Tab. S1[3] for further details).

Moreover, we must emphasize that binding is a condition necessary but not sufficient to trigger an activity, and there is a distinction between the binding (that was predicted *in silico*) and the activity (experimental *in vitro* data that we compare the prediction with). In our approach, we looked not only at the very well-known binding site but also at all potential alternate ones that may or may not trigger activity. Interestingly, Graves et al. (2005) addressed the difference between true binders and non-binders for a target, defining FP hits as "decoys" and distinguishing: i) geometric decoys, when an incorrect configuration of a ligand in a binding site is predicted and ii) hit list decoys, when a compound is

predicted to bind but, upon experimental testing, is found not to bind at relevant concentrations. Indeed "inactive" means not active at a relevant concentration, therefore binding can still exist.

3.4 *In silico* binding promiscuity and protein binding fingerprint profile

We calculated the *in-silico* protein-binding promiscuity index (*isPI*), defined as the percentage of receptors that bind versus the total number of receptors tested, for BPA and EST. Values lie between 0 and 1, where 1 indicates that the ligand shows complementarity with all tested proteins.

BPA was found to be more promiscuous (*isPI* = 0.44) than EST (*isPI* = 0.25). The space of structural promiscuity is vast: thousands of potential binders were predicted. Interestingly, as
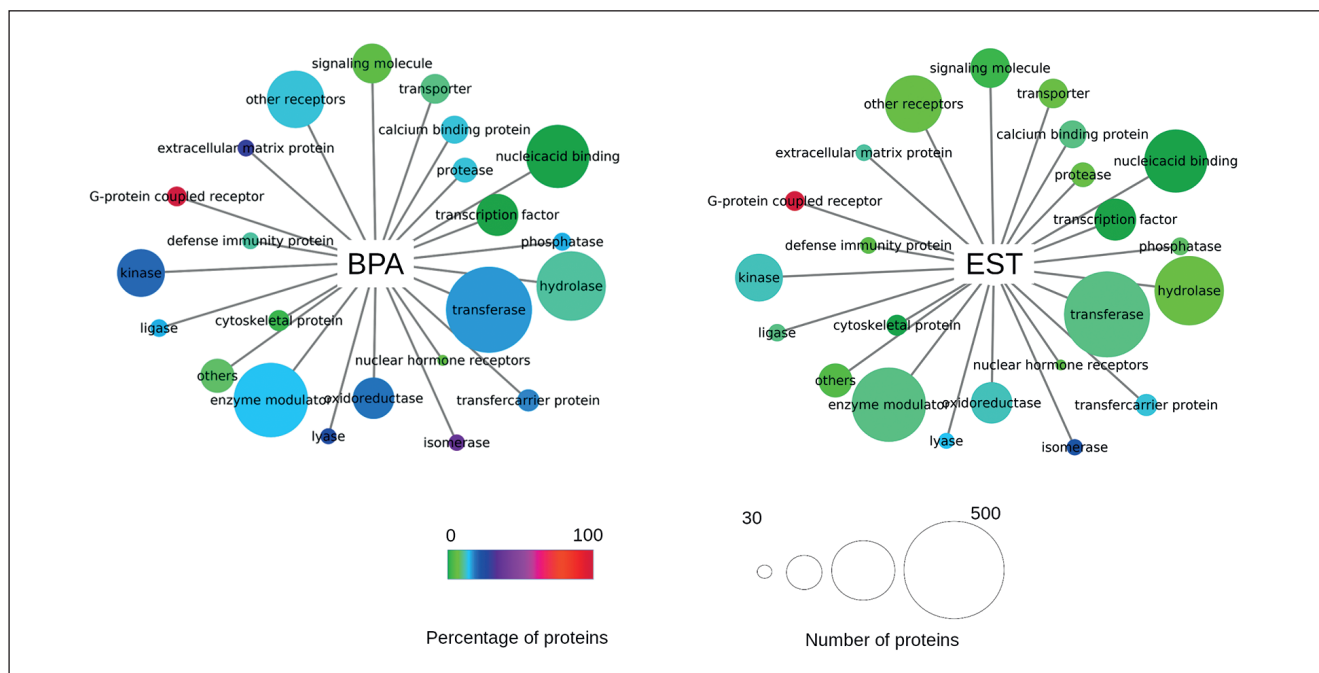

Fig. 5: Protein-binding fingerprint profiles for BPA and EST
Distributions of BPA and EST hits among protein classes are represented as circles of different dimensions. Percentage, computed for each family, of retrieved protein hits in relation to total available proteins in screened database is reported as gradient color from green to red.
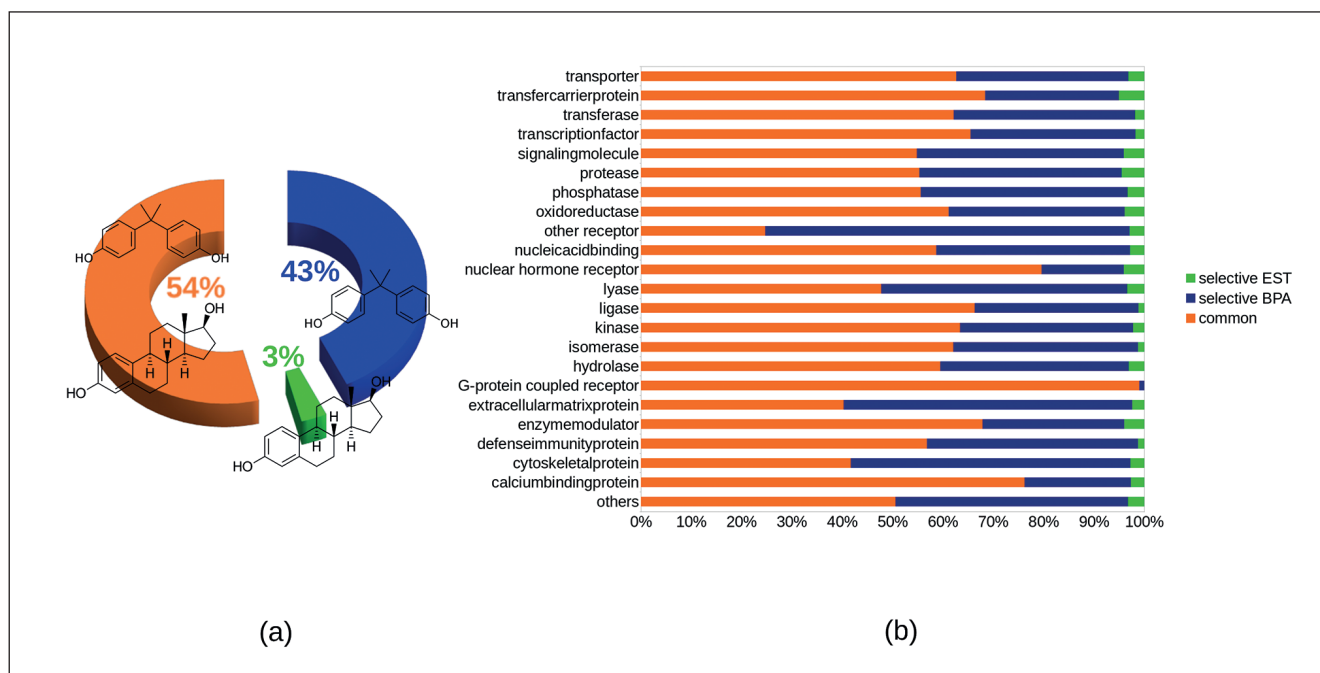
Fig. 6: Hits and protein binding selectivity classification for BPA and EST
(a) Pie chart reporting the fraction of hits detected only for BPA, only for EST, and for both (selective and common hits). (b) Classification of selective and common proteins.

already reported by other authors (MacKay and Abizaid, 2018), the binding profile does not involve solely nuclear receptors (Fig. 5). We grouped putative BPA and EST targets into different families based on Panther protein classification (Thomas et al., 2003; Mi et al., 2010): The "receptor" family was split into its subfamilies "nuclear hormone receptors", "G-protein coupled receptors", and "other receptors". Using the number of proteins for each retrieved family, we created a "protein-binding fingerprint profile" for both BPA and EST (Fig. 5). The protein families binding either molecule are represented as circles of different sizes (the larger the circle, the more proteins); furthermore the percentage of protein hits was computed for each family (indicated by a color gradient from green to red), to avoid bias introduced by the distribution of families among the dataset. We find that a large number of BPA and EST protein hits belong to the class of transferases (big circle). Nevertheless, paying attention to the percentage (blue/green circles), we can conclude that only a small part of available transferases in the screened database were actually retrieved as protein hits for BPA and EST (16.03% and 10.46%, respectively). It therefore appears that a broad array of enzymes (consisting of kinases, oxidoreductases, hydrolases, and many others), G-protein coupled receptors and transporters contributes to the *in-silico* BPA protein-binding profile.

EST exhibits a similar binding profile. Indeed, the relative dimensions of circles (Fig. 5) are very similar between the profiles (transferase > enzyme modulator > hydrolase, etc.). Nevertheless, the BPA profile shows both higher percentages (more blue) and larger absolute dimensions, reflecting its higher binding pro-

miscuity. This means that BPA tends to bind a higher percentage of proteins in the database, across different families.

We then focused the analysis on proteins that are predicted as selective targets either for BPA or EST, or for both ligands. 1,254 proteins were found to be selective hits for BPA and 81 for EST. 1,547 proteins were predicted to potentially bind both BPA and EST (Fig. 6a). We classified common and selective targets according to Panther classification (Fig. 6b). As expected, half of the protein hits were commonly targeted by BPA and EST: the high number of common hits is probably due to the similar structures they have, i.e., a hydrophobic body and two terminal OH groups. Selective BPA putative binding proteins (43%) represent a high percentage of protein hits. From this structural standpoint, BPA has a wider binding profile and potential activity pattern, compared to EST. This may be due to its simpler structure, which is more easily adaptable to many cavities. In conclusion, we learn from this *in silico* analysis that BPA is a promiscuous binder, as already discussed by other authors (MacKay and Abizaid, 2018), supporting the view that the toxicological investigation of BPA should consider this.

4 Discussion

We explored the application of a new *in silico* workflow, based on whole proteome virtual screening, to support the toxicological characterization of small molecules. The goal was to determine the nature and extent of the chemical protein binding profile and to study the feasibility to identify potential targets and mecha-

nisms of action for food chemicals. As a proof of principle, the binding profile of BPA was examined in comparison to EST. The results provided a strong correlation between BioGPS predictions and already available *in vitro* data, indicating the predictive value of the model. Unanticipated data, not reported earlier, was obtained. The virtual screening of all human and rodent 3D structures (30,153 structures) revealed high BPA protein binding promiscuity, expressed as the ratio between proteins predicted to bind and total proteins tested.

BPA was predicted to potentially interact with more than 2000 proteins, while EST was predicted to bind proteins following a very similar profile, but with lower promiscuity. Overall, these data throw a new light on protein-ligand interactions, which may profoundly challenge how modes of action and AOP are viewed and understood. The data obtained in this study are in line with the complex and numerous experimentally documented biological and toxicological effects as well as mechanistic information reported for BPA (Murata and Kang, 2018; EFSA 2015). This method may constitute a breakthrough tool reflecting the true complexity of chemical-biological interactions.

It has to be underlined that this work constitutes a first step, and that the actual application of such data, e.g., for safety evaluation, will require additional research, which is on-going. This includes the assessment of many more chemicals expressing different types of toxicity through different mechanisms and the *in vitro* screening of additional proteins that have not yet been investigated. It is important to note that the approach, as applied above, is qualitative and does not give any insight into what concentration of the substance is required to activate pathways and trigger effects. Indeed, it does not yet provide an estimate of potency. Instead, the question of potency constitutes a key research focus that will need to be considered to better exploit BioGPS in the context of risk assessment.

However, the *in-silico* protein binding fingerprint can already provide a new and broader overview on how to determine compound similarity, supporting and strengthening the application of read-across, one of the most-used alternative methods in toxicology (Hartung, 2016; Grace et al., 2017). Read-across is based on filling toxicological data gaps by extrapolating data that are available for similar chemical(s), i.e., analogue(s), and the main challenge in its application is the selection of such similar compounds. In fact, molecules cannot be considered similar in absolute terms (only based on their chemical structure), but their similarity should be evaluated with respect to their biological activity. Applying our strategy, analogue definition for use in read-across can be based not only on chemical structure or a single mechanism of action (e.g., binding to ERα in case of BPA), but on a molecule's protein binding fingerprint, which covers the full spectrum of potential MIEs. In this context, we think that our approach can be considered a step forward towards improving and strengthening the use of *in silico* prediction in a context of chemical risk assessment. It can also support priority setting to direct further toxicological studies and to better design targeted *in vitro* and *in vivo* experiments. Indeed, 3D protein binding prediction is valuable to gain insight about possible mechanisms of action concealed behind *in vivo* effects.

Our approach could be considered complementary to those developed under large research initiatives such as U.S. Tox21 (Thomas et al., 2018) and ToxCast (Richard et al., 2016), which were based on *in vitro* screening using hundreds of bioassays (815 distinct assay endpoints for 1,060 environmentally-relevant chemicals for EPA ToxCast program, while Tox21 program has run more chemicals in a smaller number of assays). Similar to the method presented here, such broad approaches provide a large spectrum of potential biological effects. However, the *in silico* approach is faster, cheaper and more comprehensive than the *in vitro* screening approach and can help to interpret and prioritize *in vitro* assays (e.g., all information regarding the protein binding profiles for BPA and EST is reported[4] and can trigger further *in vitro* screening).

In silico screening can be directly applied only to pure substances, and mixtures as such are not applicable. However, testing the known single constituents of a mixture with BioGPS could support grouping them according to their binding profiles; this could help decide on the application of combined toxicology principles to assess the safety of the mixture.

Moreover, our method can be used to test compounds that are difficult to obtain, such as metabolites and impurities, and molecules that are onerous to isolate or synthetize. We strongly believe that together with ToxCast and Tox21, combining the use of hundreds of *in vitro* assays for many environmental-relevant chemicals, a broad *in silico* screening of molecular targets could help to elucidate potential AOPs through which chemicals may cause adverse effects *in vivo*.

In conclusion, BioGPS appears a promising tool, not only for studying polypharmacology but also for mechanistic toxicological investigations. It is in line with current initiatives to develop alternative methods for toxicology, which rely less on high-dose animal studies and more on human-relevant, mechanistic information.

References

Baroni, M., Cruciani, G., Sciabola, S. et al. (2007). A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for Ligands and Proteins (FLAP): Theory and application. *J Chem Inf Model 47*, 279-294. doi:10.1021/ci600253e

Berman, H. M., Westbrook, J., Feng, Z. et al. (2000). The protein data bank. *Nucleic Acids Res 28*, 235-242. doi:10.1093/nar/28.1.235

Brincat, J. P., Carosati, E., Sabatini, S. et al. (2011). Discovery of novel inhibitors of the NorA multidrug transporter of Staphylococcus aureus. *J Med Chem 54*, 354-365. doi:10.1021/jm1011963

Burden, N., Sewell, F., Andersen, M. E. et al. (2015). Adverse outcome pathways can drive non-animal approaches for safety assessment. *J Appl Toxicol 35*, 971-975. doi:10.1002/jat.3165

Citro, V., Pena-García, J., den-Haan, H. et al. (2016). Identification of an allosteric binding site on human lysosomal alpha-galactosidase opens the way to new pharmacological chaperones for fabry disease *PLoS One 11*, 1-15. doi:10.1371/journal.pone.0165463

Chakraborty, S., Cole, S., Rader, N. et al. (2012). In silico design of peptidic inhibitors targeting estrogen receptor alpha dimer interface. *Mol Divers 16*, 441-451. doi:10.1007/s11030-012-9378-x

Cross, S., Baroni, M., Carosati, E. et al. (2010). FLAP: GRID molecular interaction fields in virtual screening. Validation using the DUD data set. *J Chem Inf Model 50*, 1442-1450. doi:10.1021/ci100221g

Davis, A. P., Grondin, C. J., Johnson, R. J. et al. (2018). The comparative toxicogenomics database: Update 2019. *Nucleic Acids Res 47*, D948-D954. doi:10.1093/nar/gky868

Delfosse, V., Grimaldi, M., Pons, J. L. et al. (2012). Structural and mechanistic insights into bisphenols action provide guidelines for risk assessment and discovery of bisphenol A substitutes. *Proc Natl Acad Sci U S A 109*, 14930-14935. doi:10.1073/pnas.1203574109

EFSA – European Food Safety Authority (2015). Panel on Food Contact Materials, Enzymes, Flavourings and Processing Aids (CEF). Scientific Opinion on the risks to public health related to the presence of bisphenol A (BPA) in foodstuffs. *EFSA 13*, 3978. doi:10.2903/j.efsa.2015.3978

Ellingson, S. R., Smith, J. C. and Baudry, J. (2014). Polypharmacology and supercomputer-based docking: Opportunities and challenges. *Mol Simul 40*, 848-854. doi:10.1080/08927022.2014.899699

Goodford, P. J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules *J Med Chem 28*, 849-857. doi:10.1021/jm00145a002

Grace, P., George, H., Prachi, P. et al. (2017). Navigating through the minefield of read-across tools: A review of in silico tools for grouping. *Comput Toxicol 3*, 1-18. doi:10.1016/j.comtox.2017.05.003

Graves, A. P., Brenk, R. and Shoichet, B. K. (2005). Decoys for docking. *J Med Chem 48*, 3714-3728. doi:10.1021/jm0491187

Hartung, T. (2016). Making big sense from big data in toxicology by read-across. *ALTEX 33*, 83-93. doi:10.14573/altex.1603091

Kasbekar, M., Fischer, G., Mott, B. T. et al. (2016). Selective small molecule inhibitor of the Mycobacterium tuberculosis fumarate hydratase reveals an allosteric regulatory site. *Proc Natl Acad Sci U S A 113*, 7503-7508. doi:10.1073/pnas.1600630113

Liu, J. and Nussinov, R. (2016). Allostery: An overview of its history, concepts, methods, and applications. *PLoS Comput Biol 12*, 1-5. doi:10.1371/journal.pcbi.1004966

Livingston, K. E. and Traynor, J. R. (2018). Allostery at opioid receptors: Modulation with small molecule ligands. *Br J Pharmacol 175*, 2846-2856. doi:10.1111/bph.13823

Lu, S., Li, S. and Zhang, J. (2014). Harnessing allostery: A novel approach to drug discovery. *Med Res Rev 34*, 1242-1285. doi:10.1002/med.21317

MacKay, H. and Abizaid, A. (2018). A plurality of molecular targets: The receptor ecosystem for bisphenol-A (BPA). *Horm Behav 101*, 59-67. doi:10.1016/j.yhbeh.2017.11.001

Murata, M. and Kang, J. H. (2018). Bisphenol A (BPA) and cell signaling pathways. *Biotechnol Adv 36,* 311-327. doi:10.1016/j.biotechadv.2017.12.002

Mi, H., Dong, Q., Muruganujan, A. et al. (2010). PANTHER version 7: Improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res 38*, D204-D210. doi:10.1093/nar/gkp1019

Nakamura, M., Yamanaka, H., Oguro, A. and Imaoka, S. (2018). Bisphenol A induces Nrf2-dependent drug-metabolizing enzymes through nitrosylation of Keap1. *Drug Metab Pharmacokinet 33*, 194-202. doi:10.1016/j.dmpk.2018.04.003

Nightingale, A., Antunes, R., Alpi, E. et al. (2017). The Proteins API: Accessing key integrated protein and genome information. *Nucleic Acids Res 45*, W539-W544. doi:10.1093/nar/gkx237

Richard, A. M., Judson, R. S., Houck, K. A. et al. (2016). ToxCast chemical landscape: Paving the road to 21st century toxicology. *Chem Res Toxicol 29*, 1225-1251. doi:10.1021/acs.chemrestox.6b00135

Rubin, B. S. (2011). Bisphenol A: An endocrine disruptor with widespread exposure and multiple effects. *J Steroid Biochem Mol Biol 127*, 27-34. doi:10.1016/j.jsbmb.2011.05.002

Siragusa, L., Cross, S., Baroni, M. et al. (2015). BioGPS: Navigating biological space to predict polypharmacology, off-targeting, and selectivity. *Proteins 83*, 517-532. doi:10.1002/prot.24753

Siragusa, L., Luciani, R., Borsari, C. et al. (2016). Comparing drug images and repurposing drugs with BioGPS and FLAP-dock: The thymidylate synthase case. *ChemMedChem 11*, 1653-1666. doi:10.1002/cmdc.201600121

Sirci, F., Goracci, L., Rodríguez, D. et al. (2012). Ligand-, structure- and pharmacophore-based molecular fingerprints: A case study on adenosine A (1), A (2A), A (2B), and A (3) receptor antagonists. *J Comput Aided Mol Des 26*, 1247-1266. doi:10.1007/s10822-012-9612-8

Thomas, D., Campbell, M. J., Kejariwal, A. et al. (2003). PANTHER: A library of protein families and subfamilies indexed by function. *Genome Res 13*, 2129-2141. doi:10.1101/gr.772403

Thomas, R. S., Paules, R. S., Simeonov, A. et al. (2018). The US Federal Tox21 Program: A strategic and operational plan for continued leadership. *ALTEX 35*, 163-168. doi:10.14573/altex.1803011

Tschammer, N. (2016). Allosteric modulators of the class A G protein coupled receptors. In T. Bödicke (ed.), *Protein Targetin Compounds* (185-207). *Adv Exp Med Biol 917*. Cham, Switzerland: Springer. doi:10.1007/978-3-319-32805-8_9

Wang, Y., Bryant, S. H., Cheng, T. et al. (2017). PubChem BioAssay: 2017 update. *Nucleic Acids Res. 45*, D955-D963. doi:10.1093/nar/gkw1118

Conflict of interest

The authors declare that they have no conflicts of interest.