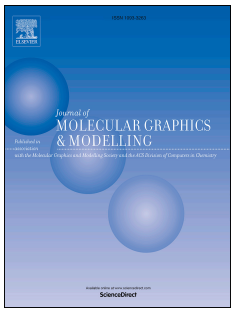
Journal Pre-proof



Identification of viable TCDD access pathways to Human AhR PAS-B ligand binding domain

Mosè Casalegno, Guido Raos, Guido Sello

PII: S1093-3263(21)00055-3

DOI: https://doi.org/10.1016/j.jmgm.2021.107886

Reference: JMG 107886

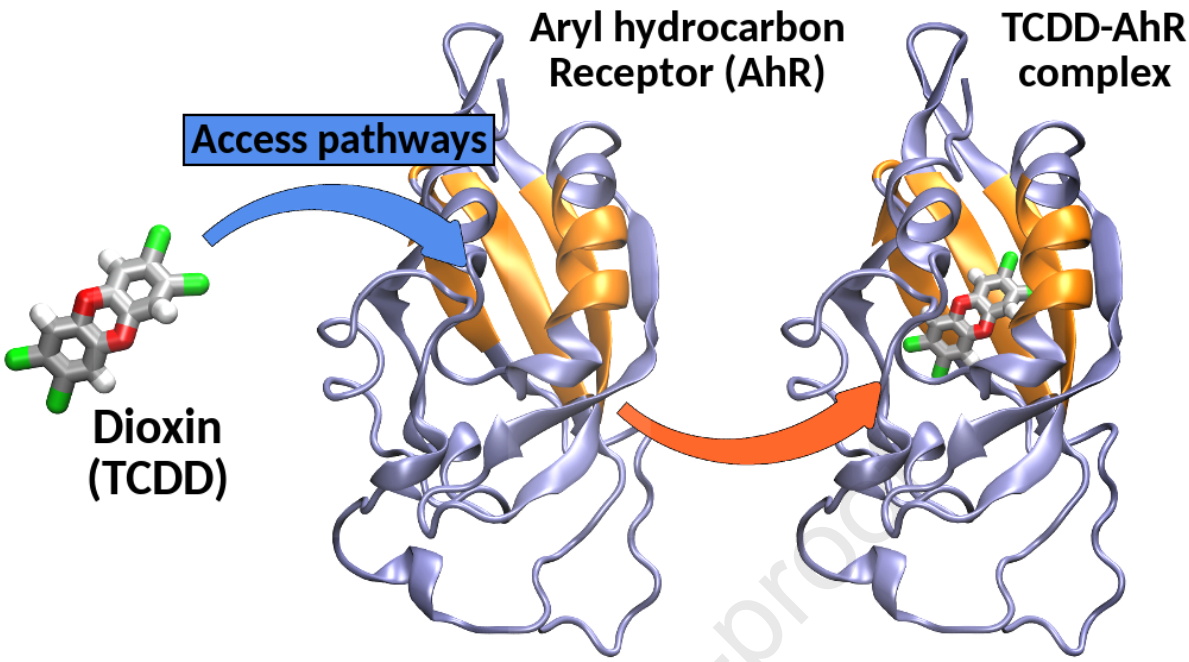
To appear in: *Journal of Molecular Graphics and Modelling*

Received Date: 28 November 2020

Revised Date: 6 February 2021

Accepted Date: 22 February 2021

Please cite this article as: M. Casalegno, G. Raos, G. Sello, Identification of viable TCDD access pathways to Human AhR PAS-B ligand binding domain, *Journal of Molecular Graphics and Modelling*, https://doi.org/10.1016/j.jmgm.2021.107886.



Identification of viable TCDD access pathways to Human AhR PAS-B ligand binding domain

Mosè Casalegno[a,*], *Guido Raos*[a], *Guido Sello*[b,*]

[a]Dipartimento di Chimica, Materiali e Ingegneria Chimica "G. Natta", Politecnico di Milano, Via L. Mancinelli 7, 20131 Milano, Italy.

[b]Dipartimento di Chimica, Università degli Studi di Milano, via Golgi 19, I-20133 Milano, Italy.

ABSTRACT

Unintentionally released in the environment as by-products of industrial activities, dioxins, exemplified by 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD), represent a primary concern for human health. Exposure to these chemicals is known to produce a broad spectrum of adverse effects, including cancer. The main mechanism of action of TCDD in humans involves binding to the Aryl hydrocarbon Receptor (AhR). Although qualitatively established, TCDD capture by the AhR remains poorly characterized at the molecular level. Starting from a recently developed structural model of the human AhR PAS-B domain, in this work we attempt the identification of viable TCDD access pathways to the human AhR ligand binding domain by means of molecular

dynamics. Based on the result of metadynamics simulations, we identify two main regions that may potentially serve as access paths for TCDD. For each path, we characterize the residues closely interacting with TCDD, thereby suggesting a possible mechanism for TCDD capture. Our results are reviewed and discussed in the light of the available information about Human AhR structure and functions.

INTRODUCTION

The term “dioxins” identifies a group of about two hundred poly-chlorinated aromatic chemicals, characterized by high toxic potency and persistence in the environment [1-4]. Exposure to dioxins has been linked to a number of adverse effects on the nervous, immune and endocrine [5-8] systems. Additionally, 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD), recognized as the most toxic representative of this class of chemicals, was classified as a human carcinogen in 1997 [9].

The mode of action of TCDD has been the subject of a number of experimental studies [10-14]. TCDD toxicokinetics begins with the absorption of TCDD into cell membranes, followed by its diffusion in the cell interior. It is known that TCDD binds to the Aryl hydrocarbon Receptor (AhR) [15], a protein which modulates the transcriptional activation of many genes, such as those involved in xenobiotic metabolism [16], regulates a number of physiological functions, including the development of pathologies like cancer [17,18].

TCDD binding to the AhR PAS-B domain triggers the release of the partner proteins and AhR hetero-dimerization with the AhR nuclear translocator protein, ARNT [19-21]. The resulting complex migrates to the nucleus, where it binds to specific DNA sequences, known as

xenobiotic-response elements, thereby encoding the transcription of metabolising enzymes[22,23]. Beside this well-established canonical signaling pathway, others have also been identified and reported [24,25].

Despite being qualitatively established, several aspects of the above route remain only partially understood at the molecular level, due to the complexity of the systems investigated, and the limited availability of experimental data. Significant efforts have gone in developing atomistically detailed models of the TCDD binding site, the ligand binding domain (LBD). Early experimental investigations identified the LBD in a hydrophobic region encompassing the PAS-B domain, a relatively small domain consisting of approximately 110 amino-acids [26-30]. In the absence of experimentally determined AhR structures, structural models of the LBD have been developed for a number of species [31-36] using computational approaches, such as homology modeling and functional analysis, taking advantage of the structural similarities across the members of the bHLH-PAS family [37]. The analysis of such models with molecular docking techniques provided important insights about the LBD structure and the residues mostly involved in TCDD binding. Quite remarkably, these residues were found to be highly conserved across different species with high TCDD responsiveness [32,38,39], thereby suggesting that ligand binding could be the main factor responsible for biological activity.

This idea was further exploited in similar studies mostly aimed at finding the set of molecular factors – collectively referred as binding mode – that could discriminate between high and low affinity ligands [40-43]. Of particular interest was the possibility of ranking the toxicity of compounds on the basis of the binding energies [44,45]. Former attempts in this direction were only partially successful, also due to the use of fixed receptor conformations in docking simulations [46,47]. Furthermore, although different ligands could effectively exploit different

key interactions with the LBD residues, the differences were often too small to justify the observed differences in the experimental binding affinities [48]. Some studies [49,50] also evidenced a poor correlation between the experimental binding affinities and those estimated by commercially available docking programs.

Some significant improvements were obtained with the introduction of more refined computational protocols [51], such as docking ensemble [33] and molecular dynamics (MD) [52,53,48]. The latter has been found especially useful in this context, thanks to the possibility to fully account for the receptor's conformational flexibility. Standard MD provides an efficient way of sampling the structure around minimum-energy configuration of ligand-receptor complexes. However, such methods cannot access the time scale of rare events, such as binding/unbinding processes. To overcome this limitation, enhanced sampling techniques, such as metadynamics [54-58], have been developed where the system's potential energy is biased during the simulation, thereby encouraging the system to explore the configurational space beyond minimum-energy conformations.

Early applications of metadynamics focused on docking in solution [59], the calculation of binding energies [60]. Later on, the method was further developed to deal with many aspects of ligand-protein association [61], including ligand binding kinetics [62-64]. The latter is of special interest in this context. Indeed, given the ligand-protein bound configuration, one may simulate the unbinding process to obtain a set of exit pathways which may be followed reversibly, and therefore eventually exploited as access pathways in ligand binding [59,60,65,66,67,68]. This approach is computationally more efficient than the search of access pathways from unbound configuration, as it bypasses the need to sample the potentially many access points to the receptor interior.

In this work, we combine standard MD and metadynamics in the attempt to identify viable TCDD binding/unbinding pathways to the human AhR (hAhR). Our aim is also to provide some background knowledge on this aspect, which has received little attention in the literature. Due to the absence of experimentally determined structures of the hAhR, the present investigation focuses on a sequence consisting of 146 residues, representative for the PAS-B domain, for which a structural model was developed via homology modeling [35]. Starting from this model and using the computational route outlined above, we identify two binding pathways for TCDD. The analysis of the free energy surfaces suggests that access from these paths may be thermodynamically feasible. For each pathway, we hypothesize a mechanism by identifying the residues most likely to be involved in TCDD capture. Finally, we discuss path accessibility in the light of the existing literature.

METHODS

System setup. As a structural model for the present investigation, we considered a sequence of 146 residues (between Pro275 and Phe420) of the hAhR, for which a three dimensional atomistic structural model was developed by Salzano et al.[35]. Hereafter, we chose to name this sequence PAS-B_{SAL}, since it substantially overlaps with that defining the PAS-B domain (Pro275-Leu386). The structure was pre-processed by adding explicit hydrogen atoms. The TCDD/PAS-B_{SAL} complex was then prepared by docking the TCDD molecule into the putative binding cavity suggested in the same study. The complex was placed in an orthorhombic simulation box, with size of 9x9x9 nm, and solvated with 17998 water molecules. The OPLS-AA force field [69-73] was used to model intra and inter molecular interactions. The positive charges on the protein were neutralized by adding 6 Cl^- ions to the solution. Additional 24 Na^+/Cl^- ion pairs were also added to simulate NaCl physiological concentration (0.15 M). For water, the TIP3P force field

was adopted [74]. All simulations were carried out with the GROMACS (v. 5.0.4) program suite [75]. Electrostatic interactions were accounted via the Particle-Mesh-Ewald method [76] with a Fourier grid spacing of 0.12 nm.

MD simulations: equilibration of the TCDD/PAS-B_{SAL} complex. The TCDD/PAS-B_{SAL} complex was equilibrated via a short (2 ns) simulation at constant temperature and pressure (NPT ensemble). The system temperature was kept constant at 310 K using the velocity rescaling algorithm developed by Bussi et al. [77], with a time constant of 1 ps. The pressure (1.0 atm) was controlled via anisotropic coupling to Berendsen barostat [78], with time constant of 4 ps and isothermal compressibility of $4.6 \cdot 10^{-5}$ atm^{-1}.

After the equilibration, a 100-ns production run was performed at constant temperature and pressure replacing the Berendsen barostat with the Parrinello-Rahman barostat [79]. In contrast to the former, the latter was found to give statistically more accurate results [80]. The post-processing of the 100-ns NPT trajectory, consisting of ten thousand frames, was carried out using different programs developed in our group. In many cases, to simplify the analysis, all frames were first aligned to the first one in the trajectory using the Kearsley algorithm [81]. The alignment was always performed on all C and N main chain atoms.

Well-tempered metadynamics simulations: TCDD/PAS-B_{SAL} unbinding. The TCDD/PAS-B_{SAL} unbinding dynamics was simulated using well-tempered metadynamics [57] approach (WT, hereafter), as implemented within the PLUMED 2.1 package [82]. In order to simplify the analysis of the results, the collective variables were defined relatively to an internal coordinate system, made up by three orthogonal axes (x, y, and z) centered at the position defined by the center-of-mass of the 13 residues indicated by Salzano et al. [35]. For better clarity, this

selection, grouping the residues closely interacting with the bound TCDD molecules, will be hereafter indicated as LBD_{SAL}. An internal coordinate system was then developed starting from the coordinates of three points, namely the LBD_{SAL} center-of-mass, the position of the backbone carbon atom in residue Phe351, and the position of the backbone carbon atom in residue Val381. A detailed explanation is given in the SI.

Running metadynamics simulations requires the choice of a set of collective variables (CVs) representative of the described system [60] and able to distinguish between different system's conformations, which in the present case were the bound and the unbound configurations. We considered three CVs, namely: 1) the distance between the TCDD and the LBD_{SAL} center-of-masses (d_{LBD}); 2) the angle between the TCDD long molecular axis (defined by the vector connecting the midpoints of the chlorine atom pairs at the opposite sides of the molecule) and the reference z axis (θ_{ZCl}); 3) the angle between the TCDD short molecular axis (defined by the vector connecting the two oxygen atoms) and the reference z axis (θ_{ZOx}). Some preliminary runs were performed to optimize the metadynamics parameters, with the aim at finding a reasonable compromise between sampling accuracy and computational efficiency. In all production runs, the height of the Gaussian hills was set at 3.0 kJ/mol. The hills width was set to 0.2 nm for the first CV and 4 degrees for the remaining two CVs. The hills deposition rate was set at 1/200 fs^{-1}, the biasfactor at 12.0.

Path collective variable metadynamics simulations: calculation of free energy surfaces (FES) and identification of the key residues along the paths. The FES associated with the two paths found in the unbinding simulations (see previous paragraph), namely P1 and P2, were calculated performing path metadynamics simulations along the unbinding trajectories. The

simulations were performed using the path collective variable approach [58], as implemented in the PLUMED 2.3 package [80].

The trajectories for these simulations were developed from those obtained from unbinding dynamics simulations (see the previous paragraph) as follows. For each path, one trajectory was prepared, comprising a variable number of "main" frames, which were linearly interpolated to obtain a larger number of smoothly connected frames. Starting from the unbound configuration (e.g. with the TCDD in bulk water), a set of consecutive main frames was selected on the basis of two descriptors, namely d_{LBD} and the root mean square distance between PAS-B_{SAL} carbon and nitrogen atoms. For two consecutive frames, the latter quantity was calculated after structure alignment, using the Kearsley algorithm [81] on all PAS-B_{SAL} carbon and nitrogen atoms. The number of interpolating frames was selected so as to obtain evenly spaced configurations in terms of the above descriptors. For P1, a trajectory consisting of 93 frames was obtained interpolating 12 main frames. For P2, 9 main frames were used to generate a trajectory consisting of 95 frames overall.

The FES were calculated as a function of two variables, namely the progress along the path, S, and the square distance from the path, Z. S is a dimensionless parameter, with values ranging from 1 to N, being N the number of frames in the path trajectory. The unit measure of Z is nm^2.

Each metadynamics run, performed at constant temperature (310 K) and pressure (1 atm), had an overall duration of 1.6 μs for each path. The following simulation parameters were optimized after some preliminary runs. The value of λ (i.e. the "temperature" factor [82]) was set at 50.0 nm-2. Geometry—adaptive Gaussian hills were adopted during the simulations, with a starting height of 3.0 kJ/mol. The hills width was set at 0.5 for S and 0.01 nm^2 for Z. Hill deposition was

performed at every 200 MD steps. The biasfactor was set at 12.0. To prevent the TCDD from being trapped in AhR regions far from the path, an additional quadratic upper wall was set at Z = 0.1 nm^2, with force constant equal to 1000 kJ/mol/nm^2. The FES reconstruction was performed with a resolution of 300 bins along both S and Z axes, and setting the width for S and Z at 0.5 and 0.01 nm^2, respectively. The FES were reweighted using the method proposed by Branduardi and coworkers [83], developed for WT simulations with adaptive Gaussians. The FES convergence was assessed by checking the hills height and the comparing the FES obtained at different simulation times (see SI for details).

The above parameters were also used to perform two sets (12 for each path) of short path collective metadynamics simulations aimed at identifying the residues closely interacting with TCDD along the paths. The path trajectories were the same described above and used in the FES calculation. Also those simulation were performed starting with the TCDD in the water phase (e.g. in the unbound configuration). Each simulation was allowed to run the time required for the TCDD to reach the LBD center (about 5 ns). The MD trajectories were then post-processed to extract the information about the interaction between TCDD and the closely interacting residues (see below).

RESULTS and DISCUSSION

The hAhR PAS-B domain, PAS-B_{SAL}, considered in this work consisted of 146 residues, spanning the region between amino acids Pro275 and Phe420 [35]. During a preliminary stage of our work, the TCDD was placed in the binding site and the resulting complex was solvated and shortly relaxed as described above. A 100-ns MD production run was then performed in order to

fully relax the TCDD/PAS-B_{SAL} complex and provide the necessary starting points for the search of potential access path candidates. Meanwhile, this simulation gave us the possibility to test the force field reliability in relation with the results obtained by Salzano et al. [35].

In order to perform a preliminary analysis of this simulation, we selected the residues closely interacting with TCDD within the binding pocket. A first selection was made by considering the average center-of-mass distance between TCDD and each PAS-B_{SAL} residue, d_{RES} hereafter. All residues within 0.5 nm from TCDD to within one standard deviation were selected, giving a set of 29 residues. For better clarity, this selection will be hereafter denoted as LBD_{CRS}. For those residues we calculated the average interaction energy with TCDD, E_{LJ}. Here, we only considered the Lennard-Jones energy, being the contributions coming from the electrostatic energy much less significant. Table 1 collects the selected residues along with the values of d_{RES} and E_{LJ} with the corresponding standard deviations. The calculation was performed averaging over 10^4 frames, corresponding to 100 ns of simulation time.

We first note that all the 13 residues, LBD_{SAL} hereafter, identified in Ref. [35], as those more closely interacting with TCDD in the bound configuration (i.e. Thr289, His291, Leu308, Leu315, Phe324, Cys333, His337, Met340, Phe351, Leu353, Ala367, Val381, Gln383), were included in our selection. Residues His291, Phe324, Cys333, His337, Phe351, Val381, and Gln383, had short-range interactions with TCDD within the binding pocket (with $d_{RES} \leq 0.3$ nm) and interesting average energies ($E_{LJ} \approx -10$ kJ/mol). For residues Thr289, Leu315, Met340, Leu353, and Ala367, the interaction energy was only moderate ($E_{LJ} \approx -5$ kJ/mol), yet with short average distances ($d_{RES} \leq 0.35$ nm). Only one residue, namely Leu308, was characterized by large average distance (0.59 ± 0.09 nm), and small interaction energy with a high standard deviation (-0.39 ± 0.43 kJ/mol).

Notably, for some of the residues not included in the above list, namely Phe295, Pro297, Ile325, Met348, Ser365, d_{RES} and E_{LJ} values likely suggested these residues could play a role in TCDD binding. Among these residues, Phe295, Pro297, and Ile325 were previously identified as target residues in a mutagenesis study performed on mouse AhR [32]. In a previous computational study [36], Met348 was already recognized as a residue interacting with the benzene ring in polychlorinated biphenyls. To the best of our knowledge, the role of Ser365 within the hAhR has never been discussed before and would deserve further experimental or computational investigation. A graphical representation of the equilibrated structure of the TCDD/PAS-B_{SAL} complex along with some representative residues surrounding the binding cavity is shown in Figure 1.

Table 1. The selection of PAS-B_{SAL} residues (LBD_{CRS}) for which the average minimum pairwise distance between TCDD and residue atoms was less or equal to 0.5 nm to within one standard deviation. Marked residues (*) were also present in the selection made in Ref. [35].

Residue	d_{RES} [nm]	E_{LJ} [kJ/mol]
Phe287	0.48 ± 0.11	-1.45 ± 1.18
Thr289*	0.34 ± 0.06	-4.69 ± 1.67
His291*	0.28 ±0.02	-12.12 ± 2.74
Phe295	0.32 ± 0.05	-10.17 ± 3.49
Thr296	0.53 ± 0.05	-1.09 ± 0.39
Pro297	0.30 ± 0.03	-4.68 ± 1.66
Leu308*	0.59 ± 0.09	-0.39 ± 0.43
Tyr310	0.45 ± 0.09	-1.92 ± 1.10

Leu315*	0.33 ± 0.05	-4.26 ± 1.81
Ser320	0.57 ± 0.15	-1.34 ± 2.01
Gly321	0.47 ± 0.09	-2.01 ± 2.14
Phe324*	0.31 ± 0.04	-8.88 ± 3.75
Ile325	0.30 ± 0.05	-7.81 ± 2.12
Cys333*	0.35 ± 0.08	-8.69 ± 3.22
Ser336	0.36 ± 0.10	-4.20 ± 1.98
His337*	0.29 ± 0.04	-10.51 ± 3.18
Met340*	0.33 ± 0.07	-4.02 ± 1.96
Ser346	0.40 ± 0.10	-3.83 ± 2.14
Gly347	0.35 ± 0.09	-3.65 ± 1.74
Met348	0.28 ± 0.03	-8.87 ± 4.14
Phe351*	0.27 ± 0.03	-8.59 ± 2.15
Leu353*	0.25 ± 0.03	-6.06 ± 2.16
Val363	0.51 ± 0.06	-0.75 ± 0.22
Ser365	0.27 ± 0.05	-8.29 ± 2.59
Asn366	0.42 ± 0.11	-3.93 ± 2.00
Ala367*	0.29 ± 0.03	-6.47 ± 2.79
Ile379	0.35 ± 0.06	-2.36 ± 0.92
Val381*	0.26 ± 0.02	-13.56 ± 2.56
Gln383*	0.30 ± 0.04	-10.31 ± 2.83

The trajectory of the TCDD/PAS-B_{SAL} complex obtained from the NPT simulation was used as a starting point for the identification of viable access pathways to the LBD. Here, we simulated TCDD unbinding in order to identify possible exit pathways. We shall show below that these can be followed reversibly and therefore exit pathways may eventually also serve as entry pathways

for TCDD. Compared to the direct approach, the latter is computationally less demanding and overcomes the necessity to sample the potentially many access points to the LBD interior. In order to overcome the high free energy barrier between the bound and unbound states, we adopted WT metadynamics [57]. The Gaussian bias potential was applied on three collective variables describing the orientation and the position of TCDD with respect to an internal coordinate system, with origin at the PAS-B_{SAL} center (see Methods for details).

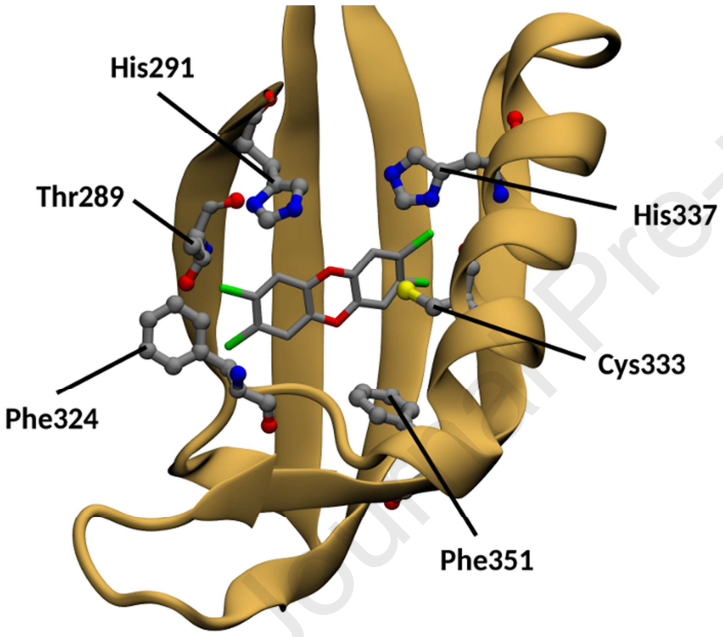


Figure 1. A three-dimensional view of the TCDD/PAS-B_{SAL} complex highlighting some of the residues listed in Table 1. Hydrogens have been omitted for clarity.

In order to minimize the bias due to the choice of the starting configuration, metadynamics simulations were performed starting from three different frames drawn from the 100-ns NPT dynamics. For each frame, we performed twenty independent metadynamics simulations, thus 60 overall. In all simulations, the TCDD molecule left PAS-B_{SAL} within few nanoseconds, along two different paths, hereafter denoted as P1 (observed in 24 runs) and P2 (observed in 36 runs).

Figure 2 provides a graphical representation of both paths along with an example trajectory followed by TCDD during the unbinding process. The secondary structure elements have been labeled according to the nomenclature generally adopted for PAS structures in the literature [84]. The N-terminal β-strands, referred to as Aβ and Bβ, are followed by three small helices (Cα, Dα, and Eα), and the helical connector (Fα). The C-terminal strands of the β-sheet are labeled as Gβ, Hβ, Iβ, followed by the α-helix Lα.

Along path P1, TCDD exits PAS-B_{SAL} crossing a U-shaped region consisting of the Fα-helix and one Gβ-strand, overall comprising the residues between Asp329 and Phe351. In this case, the undocking dynamics was characterized by TCDD displacements within the binding pocket, and changes in the Gβ conformation, which, at some point, allowed the TCDD to leave PAS-B_{SAL}. These conformational rearrangements were slightly different in different simulations, so that TCDD left PAS-B_{SAL} through different exit points between Fα and Gβ. In the remaining 36 runs, the TCDD exited along P2. This pathway is delimited by some residues in the region connecting the Aβ and Bβ strands (from Lys292 to Thr296), two alpha helices, namely Fα (residues Arg339-Gly344) and Lα (residues Ala416-Phe420), and two β-strands, here labeled as Hβ (residues Leu369-Gly374) and Iβ (residues Arg375-Ile379). It should be noted that within this pathway, the position of the TCDD upon exiting the PAS-B_{SAL} varied significantly, partly due to the mobility of the Lα helix, located in the C-terminus part of the protein, partly to the conformational flexibility of the H and I β-loops.

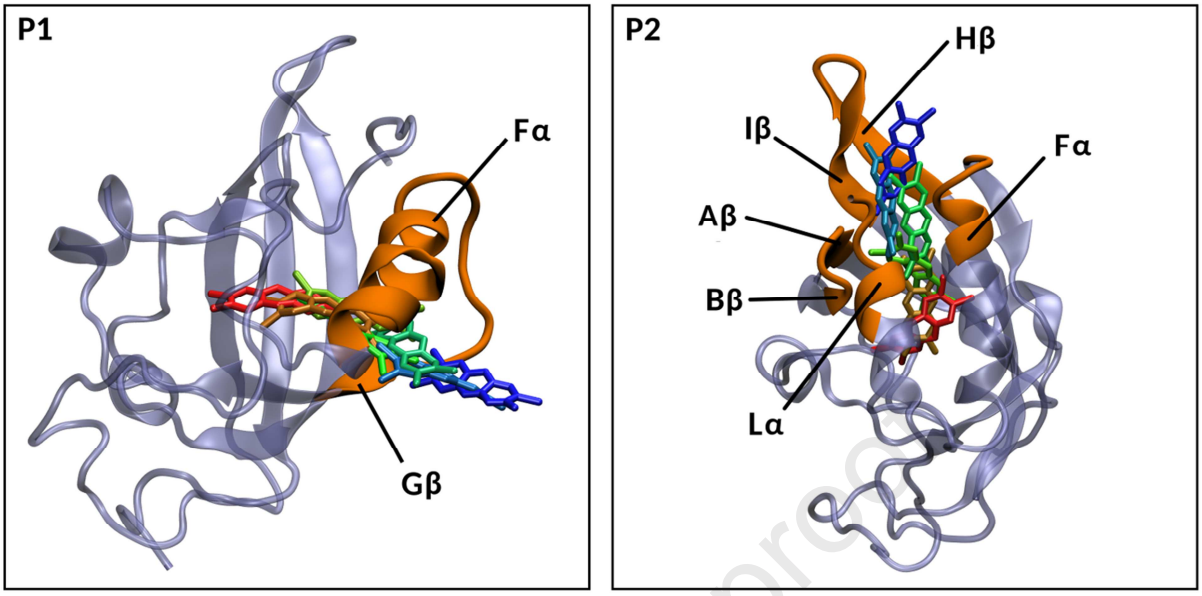


Figure 2. Three-dimensional sketches of P1 (left) and P2 (right), the paths obtained from the unbinding metadynamics simulations. The colored residues highlight the outer regions of the unbinding pathways. The starting (red) and the final (blue) TCDD positions have been reported, along with some intermediate ones. For clarity, only one PAS-B_{SAL} structure is displayed for each path. Secondary structures labels are given in the text.

To determine whether P1 and P2 could provide thermodynamically viable access pathways for TCDD, we calculated the free energy surfaces (FES) associated to both paths, by repeatedly simulating TCDD binding and unbinding along trajectories assembled from those obtained during the previous unbinding MD simulations (see Methods for details).

Starting from the unbound state, corresponding to the TCDD molecule in bulk water, metadynamics MD simulations were then performed by adding Gaussians along the path collective variables S and Z, describing the progress along the path and the square distance from the path, respectively. Figures 3 and 4 show for each path the resulting FES, which were collected over 1.6 μs. For the sake of discussion, some simplified molecular sketches of the system have been reported, which were extracted from the corresponding MD trajectories.

Figure 3 shows the free energy surface of P1. The dimensionless parameter S ranged from 1, corresponding to the unbound state, to N = 93, corresponding to the TCDD/PAS-B_{SAL} complex. Three regions, A1, B1, and C1 can be identified, corresponding to free energy basins. The former, A1, is located at about S = 20, corresponding to d_{LBD} = 1.5 nm. This minimum can be associated with the stabilizing interactions of TCDD with the outer residues of P1, taking place during the early stage of TCDD capture (a detailed discussion on the residues involved follows below). A second free energy basin, B1, is clearly visible at S ≈ 50 (d_{LBD} ≈ 0.9 nm). This distance along the path corresponds to a point where the TCDD is about to enter the binding cavity. Finally, the last basin, C1, is characterized by high S values (S > 80) and short distances (d_{LBD} < 0.2 nm), corresponding to the TCDD/PAS-B_{SAL} bound state. The FES for P2 is reported in Figure 4.

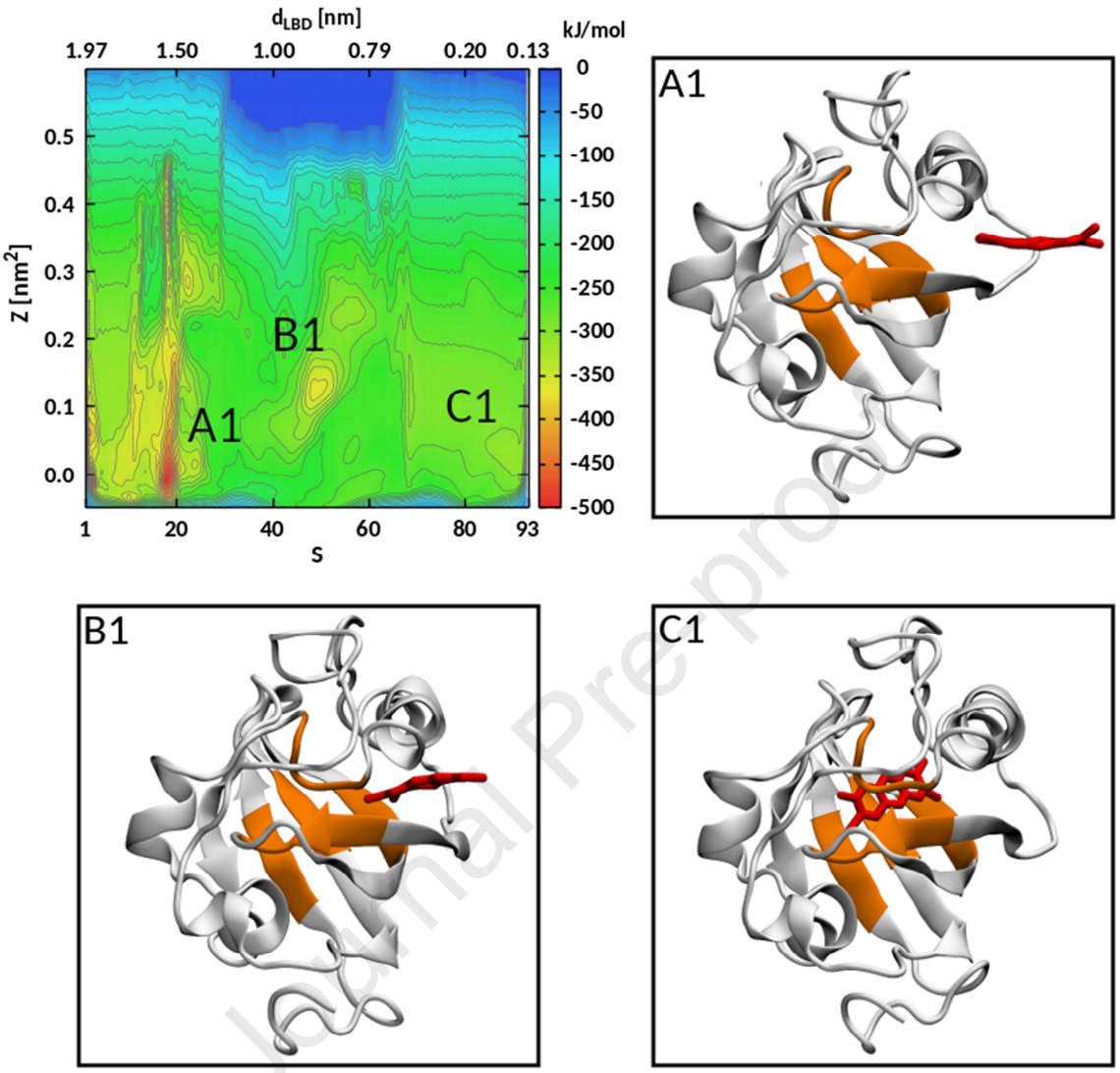


Figure 3. Two dimensional maps of the free-energy surfaces obtained for P1 as a function of S and Z (see text). Energies are in kJ/mol. The values of d_{LBD}, are also reported for the frames in the path trajectory. Three main energy basins namely, A1, B1, and C1, are highlighted. Simplified sketches of molecular structures representative of such regions are also reported, showing the position of the TCDD molecule (red) relative to the P1 residues (orange).

Also here three main energy basins can be recognized, although the free energy wells are generally shallower than those observed in P1. Basin A2 is characterized by values of S and d_{LBD} spanning a wide range ($15 < S < 30$ and $1.6\ nm < d_{LBD} < 1.3\ nm$). The extension of this basin is justified by the large conformational flexibility of this region, which includes the C-terminus chain. A detailed analysis on the conformational flexibility of the PAS-B_{SAL} is provided below. Basins B2 and C2 partially overlap, being separated by a shallow barrier. As we shall show in more detail below, basin B2 can be reasonably associated with the stabilizing interaction between TCDD and two PAS-B_{SAL} residues. Instead C2 corresponds to the TCDD/PAS-B_{SAL} bound configuration, similar to C1.

The decrease in free energy on going from bulk water to basins A1 and A2 suggests that both paths offer thermodynamically viable accesses to TCDD. The depths of these basins are quite different, as the depth of the A1 basin is -465.1 kJ/mol, whereas -268.5 kJ/mol for A2.

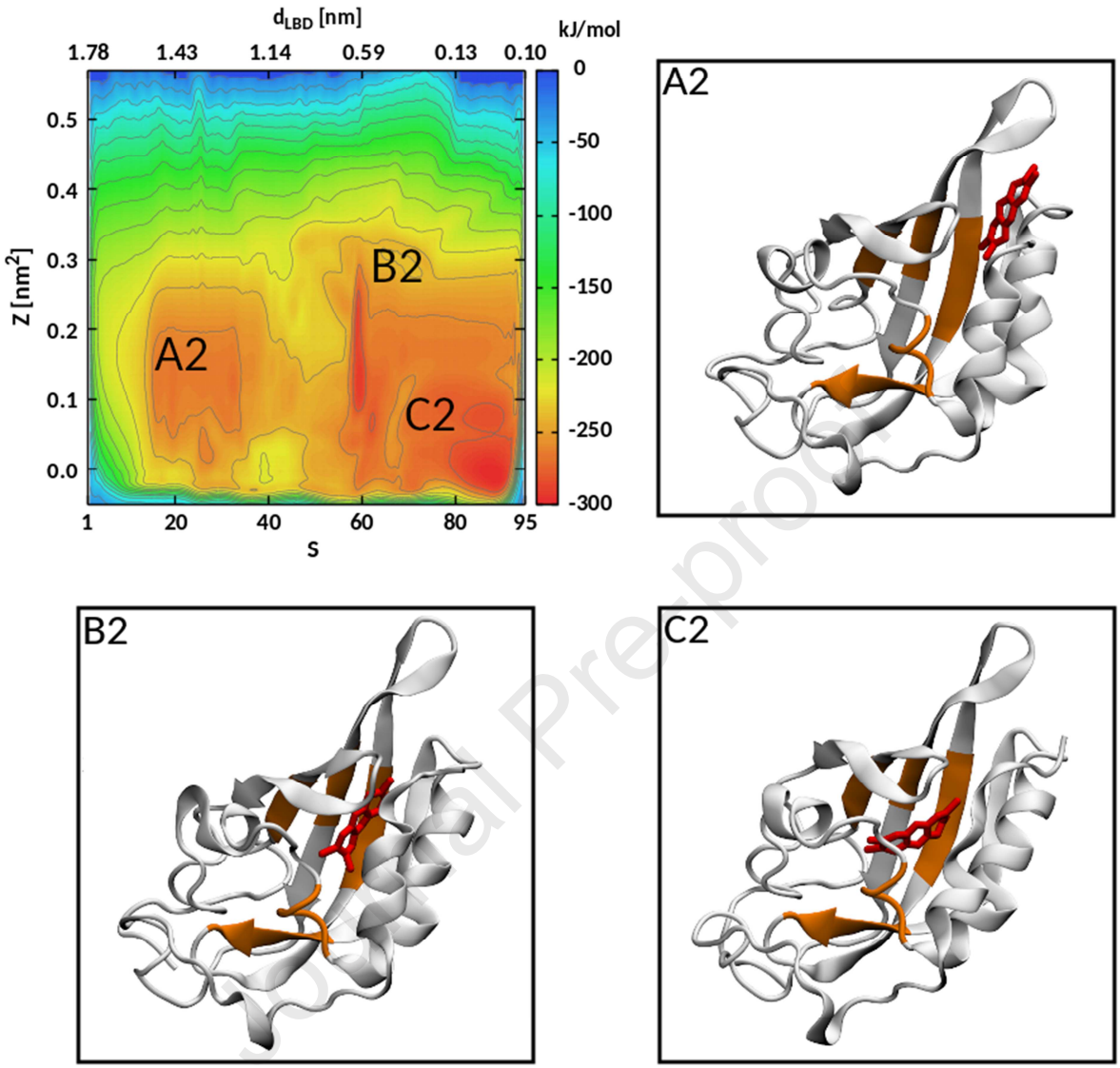


Figure 4. Two dimensional maps of the free-energy surfaces obtained for P2 as a function of S and Z (see text). Energies are in kJ/mol. The values of d_{LBD}, are also reported for the frames in the path trajectory. Three main energy basins namely, A2, B2, and C2, are highlighted. Simplified sketches of molecular structures representative of such regions are also reported, showing the position of the TCDD molecule (red) relative to the P2 residues (orange).

While the FES provide an overall view of the free energy landscape associated with TCDD binding, they do not convey detailed information about the residues involved in this process. To gain this information, we performed a set of 12 additional short path collective metadynamics simulations for each path. The path trajectories used to describe TCDD binding were the same used in FES calculation. Again, the simulations started with the TCDD molecule in the water phase. In all cases, TCDD binding took few nanoseconds to complete (usually less than 5 ns). The corresponding MD trajectories were post processed in order to extract relevant information about the intermolecular interactions between TCDD and each PAS-B_{SAL} residue. For each frame in a trajectory we considered the Lennard-Jones contributions to the intermolecular energy (E_{LJ}). The electrostatic interaction energies were significantly smaller and were neglected. Additionally, the distances between the TCDD and the LBD_{SAL} center-of-masses (d_{LBD}) were calculated. The energy values obtained across the different simulations were averaged by binning d_{LBD} values. The bin size was set at 0.05 nm.

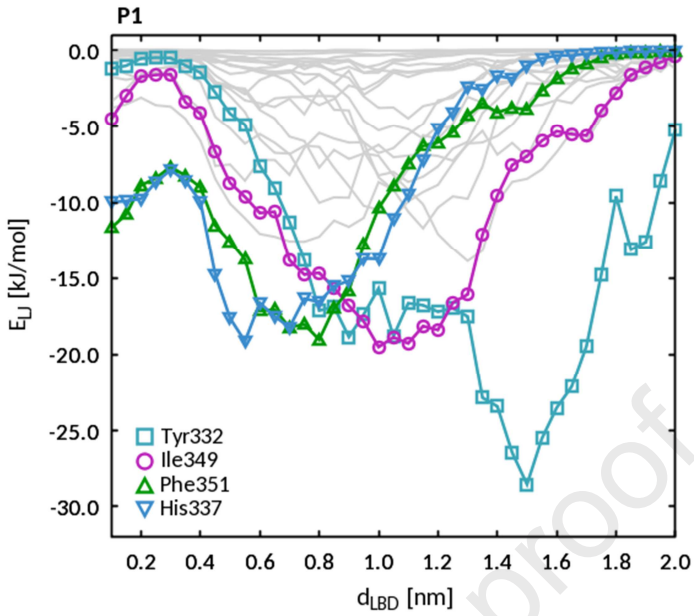


Figure 5. Average Lennard-Jones energies plotted as a function of the TCDD/LBD_{SAL} distance (d_{LBD}) for the residues from Asp329 to Phe351 along P1. The profiles of the residues more strongly interacting with TCDD are highlighted. Energies are in kJ/mol.

Figure 5 compares the E_{LJ} profiles of the residues from Asp329 to Phe351, highlighting those more strongly interacting with TCDD along the path P1. These residues were Tyr332, Ile349, His337 and Phe351. The positions of the energy minima suggest that those residues likely come into play at different stages of TCDD capture. In order to better rationalize energy profiles, and hypothesize a possible mechanism, we reported in Figure 6 the molecular sketches of four representative frames extracted from one metadynamics run. Very similar mechanisms were observed in the remaining runs. Upon approaching the binding domain, the first residue establishing a close interaction with TCDD was Tyr332. Consistently with Figure 5, this interaction took place when the TCDD distance from the LBD center was about 1.5 nm.

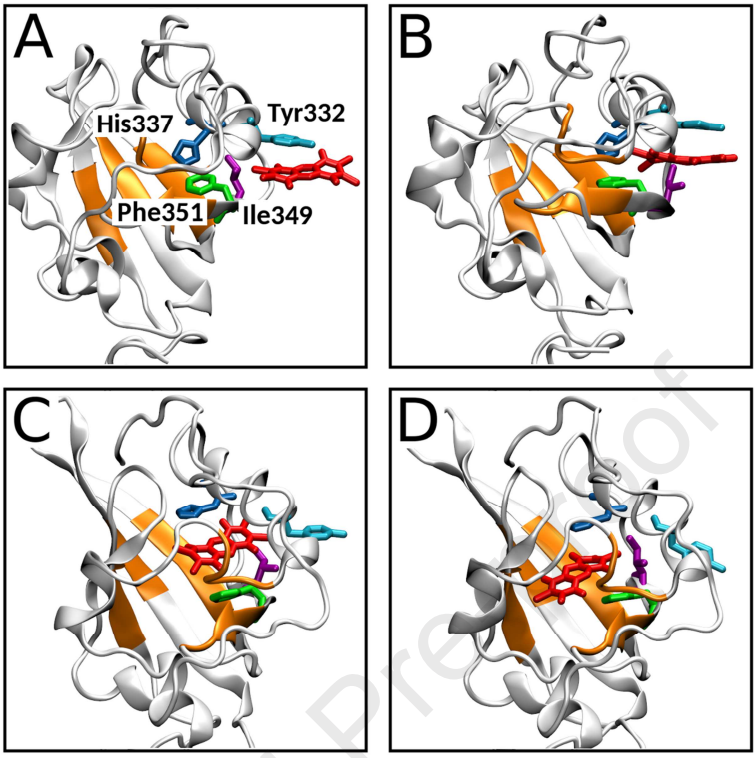


Figure 6. Representative MD frames illustrating the dynamics of TCDD capture by the PAS-B_{SAL} along P1. The four residues closely interacting with TCDD are highlighted. For better clarity, all P1 residues identified in Figure 2 are highlighted and hydrogen atoms have been omitted.

As the dioxin molecule proceeds along the entry path, the interaction with Tyr332 is complemented by that with Ile349 (see frame B). The potential wells of these two interactions in the region between 0.9 nm < d_{LBD} < 1.3 nm are close to each other, suggesting that both may trigger TCDD capture at this early stage. Subsequently (frame C), the residues His337 and Phe351 come into play (d_{LBD} < 0.8 nm), pulling the TCDD molecule into the binding site. These

residues equally contributed to TCDD capture, further pulling the molecule into the LBD (frame D).

A similar analysis was performed for P2 residues (Lys292-Thr296, Arg339-Gly344, Ala416-Phe420, Leu369-Gly374, Arg375-Ile379). To ensure an exhaustive analysis, the LBD residues were also included in this selection. According to Figure 7, where the corresponding LJ energy profiles are compared, these residues strongly interacting with TCDD were Leu293, Phe295, His337, Met340, and Ile341. Leu293 and Phe295 are located in the region connecting the Aβ and Bβ strands, while the remaining residues are all in the Fα helix. As suggested by the E_{LJ} profile, Leu293 and Ile341 were both involved in the early stage of TCDD capture, with similar interaction energies. This is clearly visible in Figure 8 (frame A), where some MD frames have been reported to support the capture mechanism. The interactions with these residues are gradually replaced by those with Phe295, Met340 and His337 as TCDD approaches the LBD center (d_{LBD} < 1.2 nm). The interaction with residue Met340 is limited in energy contribution and path length; the corresponding energy reaches its maximum value (on the average) for d_{LBD} = 1.2 nm, corresponding to the situation depicted in Figure 8B, and then rapidly decreases.

By contrast, the strength of the interaction between TCDD and residues Phe295 and His337 increases along the path, reaching a maximum at about d_{LBD} = 0.8 nm, where TCDD is located between both residues (see Figure 8C). Beyond this point, the interaction energy with His337 remains close to the minimum value (E_{LJ} = -23.86 kJ/mol) to increase as TCDD approaches the LBD center (Figure 8, frame D). The existence of a stabilizing interaction between TCDD and PAS-B_{SAL} residues in this region is consistent with the FES reported above for this path (see Figure 4B).

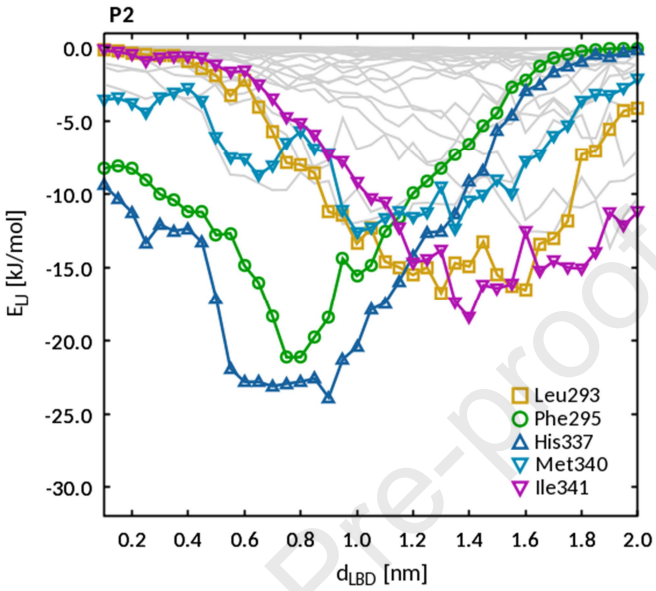


Figure 7. Average Lennard-Jones energies plotted as a function of the TCDD-LBD_{SAL} distance (d_{LBD}) for P2 residues (Lys292-Thr296, Arg339-Gly344, Ala416-Phe420, Leu369-Gly374, Arg375-Ile379). Energies are in kJ/mol.

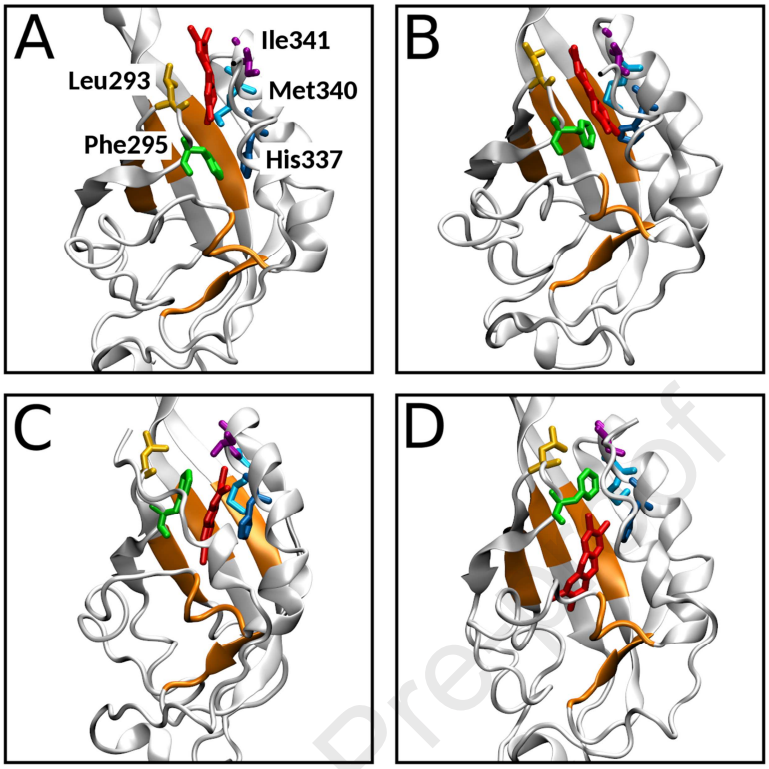


Figure 8. Representative MD frames illustrating the dynamics of TCDD capture by PAS-B_{SAL} along P2. The five residues closely interacting with TCDD are highlighted. For better clarity, all P2 residues identified in Figure 2 are highlighted and hydrogen atoms have been omitted.

Interestingly, some of the residues found, i.e. Phe295, His337, and Met340, coming into play at the late stage of TCDD capture, were originally present in the selection proposed in Ref. 30, suggesting they may play a role in the TCDD/PAS-B_{SAL} complex stabilization. In addition, His337 was already selected in P1 as a residue strongly interacting with TCDD. Some considerations can be made about the accessibility of P2 compared to P1. As far as our structural model of the hAhR is considered, while entry from P1 occurs through a well-defined region of PAS-B_{SAL}, access through P2 seems to be affected by the conformational flexibility of the C-terminus chain, Lα.

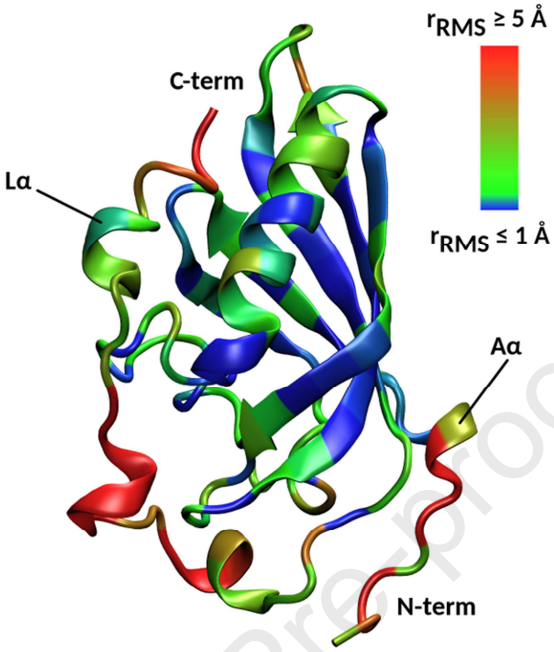


Figure 9. Three-dimensional structure of PAS-B_{SAL} showing the average root mean square displacements (r_{RMS}) of each residue.

To further investigate this aspect, we calculated the average root-mean-square displacement, r_{RMS}, of each residue in the PAS-B_{SAL} domain with enclosed TCDD. The displacements were calculated after aligning all the trajectory frames of the 100-ns NPT simulation to the starting one to eliminate the spurious effects of translational and rotational diffusion. Residue displacements were then computed from the corresponding atomic displacements, with respect to their center of mass, and averaged over the whole trajectory. Figure 9 graphically illustrates the result. As expected, the C- and N-terminus residues show large spatial displacements, whereas the internal β-sheets and α-helix do not move much. In particular, Lα residues (Ala416-Phe420), have high r_{RMS} values (e.g. 4.47 Å for residue Phe420). The conformational flexibility of these residues may eventually facilitate TCDD access through P2. Yet, only the availability of a

structural model describing the domain beyond C-terminus will provide more detailed insights about the effective accessibility of P2. Vice versa, P1 residues (Fα and Gβ) are conformationally less mobile, as expected. In this case, molecular flexibility is less critical.

CONCLUSIONS

In this work, we have attempted the identification of viable TCDD access pathways to the human AhR LBD by means of MD simulations. Two candidate paths were identified based on metadynamics simulations of TCDD/PAS-B_{SAL} unbinding. These paths spanned two different regions of PAS-B_{SAL}, the former located in the Fα-Gβ segment, the latter involving a more complex region, comprising the Aβ-Bβ connector, two alpha helices, (Fα and Lα) and two β-strands, (Hβ and Iβ). According to the FES, both paths appear to provide thermodynamically viable pathways for TCDD access. This outcome is further confirmed by the presence of residues establishing strong intermolecular interactions with TCDD, not just within the LBD, but throughout the binding-unbinding transition.

The validity of our findings is limited to the structural model we have considered (PAS-B_{SAL}), which only partially describe the hAhR structure and neglect the effect of the partner proteins, such as Hsp90 and XAP2. As evidenced by Szollosi and co-workers [48] the presence of such proteins might significantly affect hAhR accessibility and should be considered in the evaluation of the ligand binding mechanisms. In the same study, based on visual inspection, the authors have recognized the region above Fα, which resembles our P1, as a possible access path. Meanwhile, the other pathway they suggested was not found in our study. Based on the existing literature, we have no indication about the second path we have identified, P2. The presence of

free-energy minima and stabilizing interactions between TCDD and some residues suggest this path as a potentially viable one.

Hopefully, the availability of more refined structural models, will soon allow us to test the validity of our findings. At the same time, our study provides former insights on a topic that, so far, has received very little attention in the literature. In the recent past, the lack of studies in this direction was also due to the absence of adequate computational resources and tools. The advent of enhanced sampling techniques, like metadynamics, has opened new avenues in the characterization of ligand-receptor complexes. Yet, the application of these methods to biological systems is still computationally very demanding, due to the necessity to sample a large configurational space, characterized by the presence of several local energy minima. We hope that this work, in line with our previous efforts on TCDD absorption by lipid membranes [85,86], will contribute to the development of a full molecular picture describing TCDD toxicokinetics.

ASSOCIATED CONTENT

The Supporting Information is available free of charge at:…...

Details about the choice of the collective variables used in metadynamics calculations and the convergence of free energy surfaces (in PDF format).

AUTHOR INFORMATION

Corresponding Authors

Mosè Casalegno - Dipartimento di Chimica, Materiali e Ingegneria Chimica "G. Natta", Politecnico di Milano, Via L. Mancinelli 7, 20131 Milano, Italy; Email: mose.casalegno@polimi.it.

Guido Sello - Dipartimento di Chimica, Università degli Studi di Milano, via Golgi 19, I-20133 Milano, Italy; Email: guido.sello@unimi.it.

Authors

Guido Raos - Dipartimento di Chimica, Materiali e Ingegneria Chimica "G. Natta", Politecnico di Milano, Via L. Mancinelli 7, 20131 Milano, Italy; Email: guido.raos@polimi.it.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS. The Authors acknowledge the CINECA award under the ISCRA initiative (project AHRLBD, n. HP10CVZMGV), for the availability of high performance computing resources and support.

ABBREVIATIONS

TCDD, 2,3,7,8-tetrachlorodibenzo-p-dioxin; AhR, Aryl hydrocarbon Receptor; LBD, ligand binding domain; hAhR, human AhR; MD, molecular dynamics.

SOFTWARE. The following programs have been used in figure making: VMD [87] (version 1.3.1) for drawing the molecular structures (Figures 1, 2, 3, 4 ,6, 8, and 9). All plots have been made with Gnuplot [88] (Figures 3, 4, 5, and 7). GIMP [89] (version 2.8) was used in all figures for image editing.

REFERENCES

1 Dopico M, Gómez A (2015) Review of the current state and main sources of dioxins around the world. Journal of the Air and Waste Management Association 65:1033-1049

2 Malisch R, Kotz, A. (2014) Dioxins and PCBs in feed and food - Review from European perspective. Sci Tot Env 491-492:2-10

3 Schiavon M, Torretta V, Rada EC, Ragazzi M (2016) State of the art and advances in the impact assessment of dioxins and dioxin-like compounds. Environ Monit Assess 2016;188:1-20

4 Kanan S, Samara F (2018) Dioxins and furans: A review from chemical and environmental perspectives. Trends in Environmental Analytical Chemistry 17:1-13

5 White SS, Birnbaum LS (2009) An Overview of the Effects of Dioxins and Dioxin-like Compounds on Vertebrates, as Documented in Human and Ecological Epidemiology. J. Environ. Sci. Health. Part C. 27:197-211

6 Bruner-Tran KL, Gnecco J, Ding T, Glore DR, Pensabene V, Osteen KG (2017) Exposure to the Environmental Endocrine Disruptor TCDD and Human Reproductive Dysfunction: Translating Lessons from Murine Models. Reproductive Toxicology 68:59-71

7 Humblet O, Birnbaum L, Rimm E, Mittleman MA, Hauser R (2008) Dioxins and Cardiovascular Disease Mortality. Environ. Health Perspect. 116:1443-1448

8 Yoon K, Kwack SJ, Kim HS, Lee BM (2014) Estrogenic endocrine-disrupting chemicals: molecular mechanisms of actions on putative human diseases. J. Toxicol. Environ. Health B Crit. Rev 17(3):127-174

9 Steenland K, Bertazzi P, Baccarelli A, Kogevinas M (2004) Dioxin Revisited: Developments Since the 1997 IARC Classification of Dioxin as a Human Carcinogen. Environ. Health Perspect 112:1265-1268

10 Whitlock JP. Mechanistic aspects of dioxin action (1993) Chem. Res. Toxicol 6:754-763.

11 Tavakoly Sany SB, Hashim R, Salleh A, Rezayi M, Karlen DJ, Razavizadeh BBM, Abouzari-lotf E (2015) Dioxin risk assessment: mechanisms of action and possible toxicity in human health. Environ. Sci. Pollut. Res 22:19434-19450

12 Mandal PK (2005) Dioxins: a review of its environmental effects and its aryl hydrocarbon receptor biology. J. Comp. Physiol. B 175:221-230

13 Poland A, Knutson JC (1982) 2,3,7,8-tetrachlorodibenzo-p-dioxin and related halogenated aromatic hydrocarbons: examination of the mechanism of toxicity. Ann. Rev. Pharm. Toxicol 22:517-554

14 Knutsen HK, Alexander J, Barregård L, Bignami M, Brüschweiler B, Ceccatelli S, Cottrill, B, Dinovi M, Edler L, Grasl-Kraupp B, Hogstrand C, Nebbia CS, Oswald IP, Petersen A, Rose M, Roudot AC, Schwerdtle T, Vleminckx C, Vollmer G, Wallace H, Fürst P, Håkansson H, Halldorsson T, Lundebye AK, Pohjanvirta R, Rylander L, Smith A, van Loveren H, Waalkens-

Berendsen I, Zeilmaker M, Binaglia M, Gómez Ruiz JÁ, Horváth Z, Christoph E, Ciccolallo L, Ramos Bordajandi L, Steinkellner H, Hoogenboom LR (2018) Risk for animal and human health related to the presence of dioxins and dioxin-like PCBs in feed and food. EFSA Panel on Contaminants in the Food Chain (CONTAM) 16(11):e05333

15 Tian J, Feng Y, Fu H, Xie HQ, Jiang JX, Zhao B (2015) The Aryl Hydrocarbon Receptor: A Key Bridging Molecule of External and Internal Chemical Signals. Environ. Sci. Technol. 49:9518-9531

16 Nguyen LP, Bradfield CA (2008) The Search for Endogenous Activators of the Aryl Hydrocarbon Receptor. Chem. Res. Toxicol. 21:102-116

17 Leclerc D, Pires ACS, Guillemin GJ, Gilot D (2021) Detrimental activation of AhR pathway in cancer: an overview of therapeutic strategies. Curr. Opin. Immunol. 70: 15-26

18 Baker JR, Sakoff JA, McCluskey A (2020) The aryl hydrocarbon receptor (AhR) as a breast cancer drug target. Med. Res. Rev. 40:972-1001

19 Fukunaga BN, Probst MR, Reisz-Porszasz S, Hankinson O (1995) Identification of functional domains of the aryl hydrocarbon receptor. J. Biol. Chem. 270:29270-29278

20 Lindebro M, Poellinger L, Whitelaw M (1995) Protein-protein interaction via PAS domains: role of the PAS domain in positive and negative regulation of the bHLH/PAS dioxin receptor-Arnt transcription factor complex. EMBO J 14:3528-3539

21 Swanson HI (2002) DNA binding and protein interactions of the AHR/ARNT heterodimer that facilitate gene activation. Chem. Biol. Interact 141:63-76

22 Denison MS, Soshilov AA, He G, DeGroot DE, Zhao B (2011) Exactly the same but different: promiscuity and diversity in the molecular mechanisms of action of the aryl hydrocarbon (dioxin) receptor. Toxicol. Sci 124:1-22

23 Nukaya M, Lin BC, Glover E, Moran SM, Kennedy GD, Bradfield CA (2010) The Aryl Hydrocarbon Receptor-interacting Protein (AIP) Is Required for Dioxin-induced Hepatotoxicity but Not for the Induction of the Cyp1a1 and Cyp1a2 Genes. J. Biol. Chem 285:35599-35605

24 Guyot E, Chevallier A, Barouki R, Coumoul X (2013) The AhR twist: ligand-dependent AhR signaling and pharmaco-toxicological implications. Drug Discov Today 18:479-486

25 Wright EJ, De Castro KP, Joshi AD, Elferink CJ (2017) Canonical and non-canonical aryl hydrocarbon receptor signaling pathways. Curr Opin Toxicol 2:87-92.

26 Bonati L, Corrada D, Tagliabue SG, Motta S (2017) Molecular modeling of the AhR structure and interactions can shed light on ligand-dependent activation and transformation mechanisms. Curr Opin Toxicol 2:42-49

27 Coumailleau P, Poellinger L, Gustafsson JÅ, Whitelaw ML (1995) Definition of a minimal domain of the dioxin receptor that is associated with Hsp90 and maintains wild type ligand binding affinity and specificity. J Biol Chem 270:25291-25300

28 Ema M, Ohe N, Suzuki M, Mimura J, Sogawa K, Ikawa S, Fujii-Kuriyama Y (1994) Dioxin binding activities of polymorphic forms of mouse and human arylhydrocarbon receptors. J Biol Chem 269:27337-27343

29 Goryo K, Suzuki A, Carpio CAD, Siizaki K, Kuriyama E, Mikami Y, Kinoshita K, Yasumoto K, Rannug A, Miyamoto A, Fujii-Kuriyama Y, Sogawa K (2007) Identification of

amino acid residues in the Ah receptor involved in ligand binding. Biochem Biophys Res Commun 354:396-402

30 Procopio M, Lahm A, Tramontano A, Bonati L, Pitea D (2002) A model for recognition of polychlorinated dibenzo-p-dioxins by the aryl hydrocarbon receptor. Eur J Biochem 269:13-18

31 Pandini A, Denison MS, Song Y, Soshilov AA, Bonati L (2007) Structural and functional characterization of the aryl hydrocarbon receptor ligand binding domain by homology modeling and mutational analysis. Biochemistry 46:696-708

32 Pandini A, Soshilov AA, Song Y, Zhao J, Bonati L, Denison MS (2009) Detection of the TCDD binding-fingerprint within the Ah receptor ligand binding domain by structurally driven mutagenesis and functional analysis. Biochemistry 48:5972-5983

33 Motto I, Bordogna A, Soshilov AA, Denison MS, Bonati L(2011) A new aryl hydrocarbon receptor homology model targeted to improve docking reliability. J Chem Inf Model 51:2868-2881

34 Odio C, Holzman SA, Denison MS, Fraccalvieri D, Bonati L, Franks DG, Hahn ME, Powell WH (2013) Specific ligand binding domain residues confer low dioxin responsiveness to AHR1β of xenopus laevis. Biochemistry 52:1746-1754

35 Salzano M, Marabotti A, Milanesi L, Facchiano A (2011) Human aryl-hydrocarbon receptor and its interaction with dioxin and physiological ligands investigated by molecular modeling and docking simulations. Biochem Biophys Res Commun 413:176-181

36 Cao F, Li X, Ye L, Xie Y, Wang X, Shi W, Qian X, Zhu Y, Yu H (2013) Molecular docking, molecular dynamics simulation, and structure-based 3D-QSAR studies on the aryl

hydrocarbon receptor agonistic activity of hydroxylated polychlorinated biphenyls. Environ Toxicol Pharmacol 36:626-635

37 Fribourgh JL, Partch CL (2017) Assembly and function of bHLH-PAS complexes. Proc Natl Acad Sci 114:5330-5332

38 Fraccalvieri D, Soshilov AA, Karchner SI, Franks DG, Pandini A, Bonati L, Hahn ME, Denison MS (2013) Comparative Analysis of Homology Models of the Ah Receptor Ligand Binding Domain: Verification of Structure-Function Predictions by Site-Directed Mutagenesis of a Nonfunctional Receptor. Biochemistry 52:714-725

39 Hirano M, Hwang JH, Park HJ, Bak SM, Iwata H, Kim EY (2015) In silico analysis of the interaction of avian aryl hydrocarbon receptors and dioxins to decipher isoform-, ligand-, and species-specific activations. Environ Sci Technol 49:3795-3804

40 Bisson WH, Koch DC, O'Donnell EF, Khalil SM, Kerkvliet NI, Tanguay RL, Abagyan R, Kolluri SK (2009) Modeling of the aryl hydrocarbon receptor (AhR) ligand binding domain and its utility in virtual ligand screening to predict new AhR ligands. J Med Chem 52:5635-5641

41 O'Donnell EF, Saili K, Koch D, Kopparapu P, Farrer D, Bisson W, Mathew L, Sengupta S, Kerkvliet N, Tanguay R, Kolluri S (2010) The anti-Inflammatory drug leflunomide is an agonist of the aryl hydrocarbon receptor. PloS One 5(10):e13128

42 Murray IA, Flaveny CA, Chiaro CR, Sharma AK, Tanos RS, Schroeder JC, Amin SG, Bisson WH, Kolluri SK, Perdew GH (2011) Suppression of cytokine-mediated complement factor gene expression through selective activation of the Ah receptor with 3',4'-dimethoxy-α-naphthoflavone. Mol Pharmacol 79:508-519

43 Gerlach CV, Das SR, Volz DC, Bisson WH, Kolluri SK, Tanguay RL (2014) Mono-substituted isopropylated triaryl phosphate, a major component of firemaster 550, is an AHR agonist that exhibits AHR-independent cardiotoxicity in zebrafish. Aquat Toxicol 154:71-79

44 Panda R, Cleave ASS, Suresh P (2014) In silico predictive studies of mAHR congener binding using homology modelling and molecular docking. Toxicol Ind Health 30:765-776

45 Gadhwal M, Patil S, D'Mello P, Joshi A (2013) Homology modeling of aryl hydrocarbon receptor and docking of agonists and antagonists. Int J Pharm Pharm Sci 5:76-81

46 Bordogna A, Pandini A, Bonati L (2011) Predicting the accuracy of protein-ligand docking on homology models. J Comput Chem 32:81-98

47 Cozzini P, Kellogg GE, Spyrakis F, Abraham DJ, Costantino G, Emerson A, Fanelli F, Gohlke H, Kuhn LA, Morris GM, Orozco M, Pertinhez TA, Rizzi M, Sotriffer CA (2008) Target flexibility: an emerging consideration in drug discovery and design. J Med Chem 51:6237-6255

48 Szöllősi D, Erdei Á, Gyimesi G, Magyar C, Hegedűs T (2016) Access path to the ligand binding pocket may play a role in xenobiotics selection by AhR. PloS Comp Biol 11(1): e0146066

49 Gohlke HG, Klebe G (2002) Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Angew Chem Int Edit 41:2644-2676.

50 Kim R, Skolnick J (2008) Assessment of programs for ligand binding affinity prediction. J Comput Chem 29:1316-1331.

51 Jogalekar AS, Reiling S, Vaz RJ (2010) Identification of optimum computational protocols for modeling the aryl hydrocarbon receptor (AhR) and its interaction with ligands. Med Chem Lett 20:6616-6619.

52 Nuti R, Gargaro M, Matino D, Dolciami D, Grohmann U, Puccetti P, Fallarino F, Macchiarulo A (2014) Ligand binding and functional selectivity of l-tryptophan metabolites at the mouse aryl hydrocarbon receptor (mAhR). J Chem Inf Model 54:3373-3383

53 Xing Y, Nukaya M, Satyshur KA, Jiang LV, Stanevich V, Korkmaz EN, Burdette L, Kennedy GD, Cui Q, Bradfield CA (2012) Identification of the Ah-receptor structural determinants for ligand preferences. Toxicol Sci 129:86-97

54 Barducci A, Bonomi M, Parrinello M (2011) Metadynamics. WIREs Comput Mol Sci 1:826-843

55 Micheletti C, Laio A, Parrinello M (2004) Reconstructing the density of states by history-dependent metadynamics. Phys Rev Lett 92:170601

56 Barducci A, Bussi G, Parrinello M (2008) Well-tempered metadynamics: a smoothly-converging and tunable free-energy method. Phys Rev Lett 100:020603.

57 Bonomi M, Parrinello M (2010) Enhanced sampling in the well-tempered ensemble. Phys Rev Lett 104:190601.

58 Branduardi D, Gervasio FL, Parrinello M (2007) From A to B in free energy space. J Chem Phys 126:054103.

59 Gervasio FL, Laio A, Parrinello M (2005) Flexible docking in solution using metadynamics. J Am Chem Soc 127:2600-2607

60 Fidelak J, Juraszek J, Branduardi D, Bianciotto M, Gervasio FL (2010) Free-energy-based methods for binding profile determination in a congeneric series of CDK2 inhibitors. J Phys Chem B 114:9516-9524

61 Cavalli A, Spitaleri A, Saladino G, Gervasio FL (2015) Investigating drug-target association and dissociation mechanisms using metadynamics-based algorithms. Acc Chem Res 48:277-285

62 Casasnovas R, Limongelli V, Tiwary P, Carloni P, Parrinello M (2017) Unbinding kinetics of a p38 MAP kinase type II inhibitor from metadynamics simulations. J Am Chem Soc 139(13): 4780-4788

63 Tiwary P, Limongelli V, Salvalaglio M, Parrinello M (2015) Kinetics of protein-ligand unbinding: predicting pathways, rates, and rate-limiting steps. Proc Natl Acad Sci 112:E386-E391

64 Matsunaga Y, Fujisaki H, Terada T, Furuta T, Moritsugu K, Kidera A (2012) Minimum free energy path of ligand-induced transition in adenylate kinase. PloS Comp Biol 8:e1002555

65 Deng Y, Roux B (2009) Computations of standard binding free energies with molecular dynamics simulations. J Phys Chem B 113(8):2234-2246

66 D'Annessa I, Raniolo S, Limongelli V, Di Marino D, Colombo G. (2019) Ligand binding, unbinding, and allosteric effects: deciphering small-molecule modulation of HSP90. J Chem Theory Comput 15:6368-6381

67 Limongelli V, Bonomi M, Parrinello M (2013) Funnel metadynamics as accurate binding free-energy method. PNAS 10:6358-6363

68 Capelli R, Carloni P, Parrinello M (2019) Exhaustive search of ligand binding pathways via volume-based metadynamics, J Phys Chem Lett 10:3495-3499

69 Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118:11225-11236

70 Jorgensen WL, McDonald NA (1998) Development of an all-atom force field for heterocycles. Properties of liquid pyridine and diazenes. Theochem 424:145-155

71 Rizzo RC, Jorgensen WL (1999) OPLS All-Atom model for amines: resolution of the amine hydration problem. J Am Chem Soc 121:4827-4836

72 McDonald NA, Jorgensen WL (1998) Development of an all-atom force field for heterocycles. Properties of liquid pyrrole, furan, diazoles, and oxazoles. J Phys Chem B 102:8049-8059

73 Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. J Phys Chem B 105:6474-6487

74 Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML(1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79:926-935

75 Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) GROMACS: fast, flexible, and free. J Comput Chem 26:1701-1718

76 Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an Nlog(N) method for Ewald sums in large systems. J Chem Phys 98(12):10089-10092

77 Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126:014101

78 Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1994) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684-3690

79 Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: a new molecular dynamics method. J App Phys 52:7182-7190

80 Shirts MR (2013) Simple quantitative tests to validate sampling from thermodynamic ensembles. J Chem Theory Comput 9(2):909-926

81 Kearsley SK (1989) On the orthogonal transformation used for structural comparisons. Acta Crystallogr A 45:208-210

82 Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G (2014) PLUMED 2: new feathers for an old bird. Comp Phys Comm 185:604-613

83 Branduardi D, Bussi G, Parrinello M (2012) Metadynamics with adaptive gaussians. J Chem Theory Comput 8(7):2247-2254

84 Gong W, Hao B, Mansy SS, Gonzalez G, Gilles-Gonzalez MA, Chan MK (1998) Structure of a biological oxygen sensor: A new mechanism for heme-driven signal transduction, Proc Natl Acad Sci 95:15177-15182

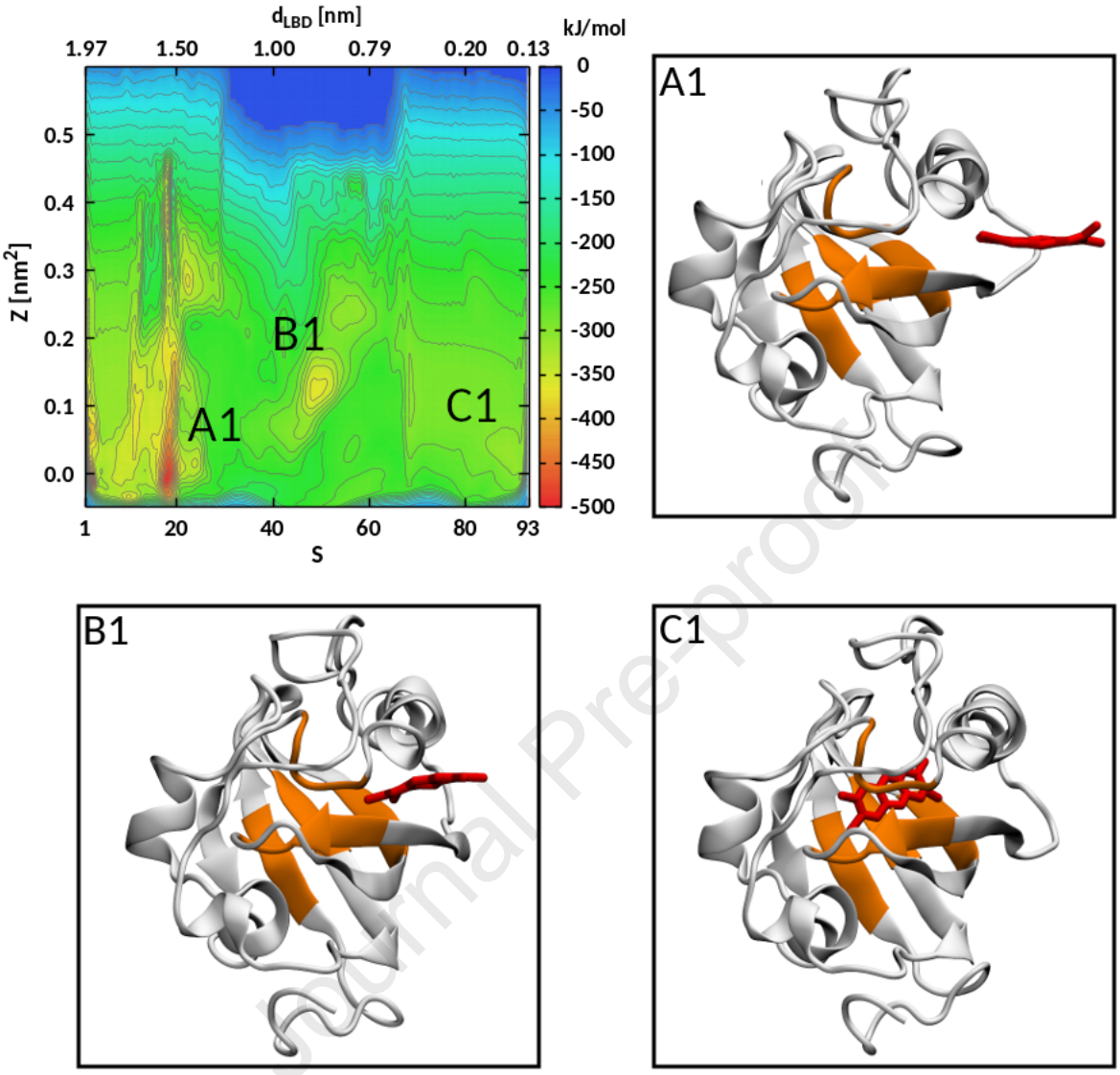
85 Casalegno M, Raos G, Sello G (2015) Hydrophobic aggregation and collective absorption of dioxin into lipid membranes: insights from atomistic simulations. Phys Chem Chem Phys 17:2344-2348

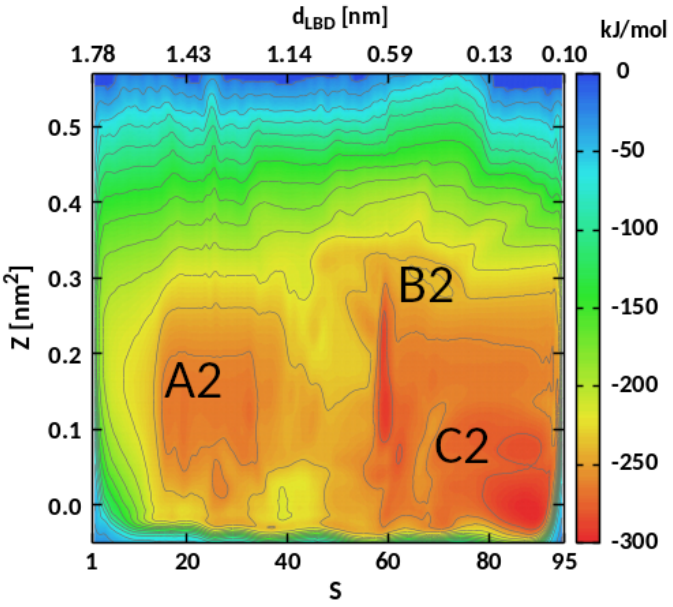
86 Casalegno M, Raos G, Sello G (2016) From dioxin to dioxin congeners: understanding the differences in hydrophobic aggregation in water and absorption into lipid membranes by means of atomistic simulations. Phys Chem Chem Phys 18:17731-17739

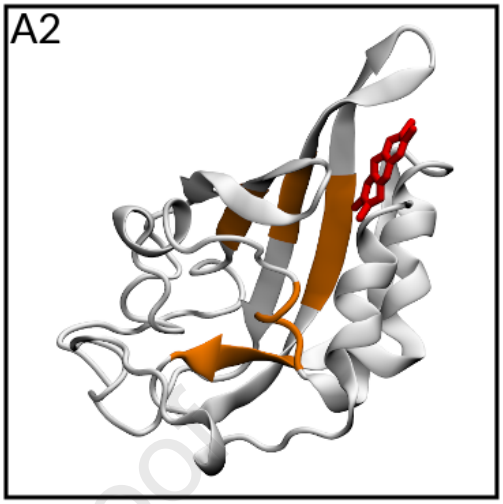
87 Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Molec Graphics 14:33-38

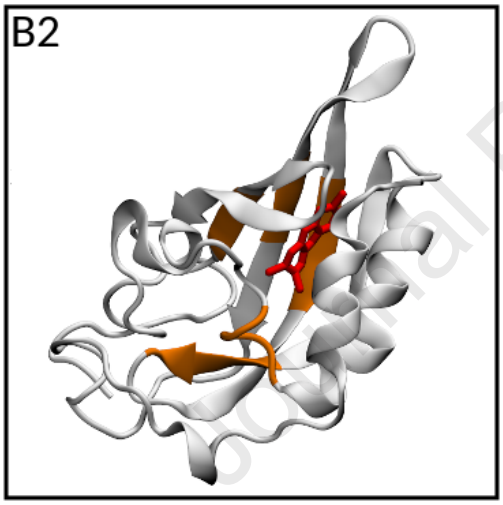
88 Gnuplot 5.0: an interactive plotting program. http://gnuplot.sourceforge.net/. Accessed: 4 April 2018

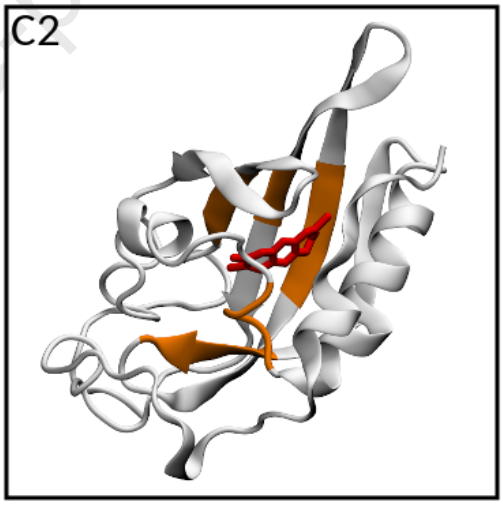
89 Gnu Image Manipulation Program (GIMP). http://www.gimp.org/. Accessed: 4 April 4 2018











Highlights

2,3,7,8-tetrachlorodibenzo-p-dioxin, simply known as dioxin, is a primary concern for human health.

The main mechanism of action of dioxin in humans, involving binding to the Aryl hydrocarbon Receptor, is poorly understood at the molecular level.

Standard and biased molecular dynamics approaches are used to identify two access pathways to the Aryl hydrocarbon Receptor.

For each pathway, a mechanism is hypothesized by identifying the key residues likely to be involved in dioxin capture.

Declaration of interests

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

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: