Molecular insight into endosulfan degradation by Ese protein from Arthrobacter : Evidence-based structural bioinformatics and quantum mechanical calculations

Ernesto Andrade-Collantes, Bruno Landeros-Rivera, Yudibeth Sixto-López, Ciresthel Bello-Rios, Julia Contreras-García, José Antonio Garzón Tiznado, Abraham Pedroza-Torres, Beni Camacho-Pérez, Sarita Montaño

Molecular insight into endosulfan degradation by Ese protein from Arthrobacter: evidence-based structural bioinformatics and quantum mechanical calculations

Authors: Ernesto Andrade-Collantes[1], Bruno Landeros-Rivera[2], Yudibeth Sixto-López[3], Ciresthel Bello-Rios[4] Julia Contreras-García[5], José Antonio Garzón Tiznado[1], Abraham Pedroza-Torres[6], Beni Camacho-Pérez[7] and Sarita Montaño[1].

Affiliation:

[1] Laboratorio de Modelado Molecular y Bioinformática, *Facultad de Ciencias Químico-Biológicas, Universidad Autónoma de Sinaloa. Ciudad Universitaria s/n, CP 80010, Culiacán, Sinaloa, México*

[2] Facultad de Química, Departamento de Química Inorgánica y Nuclear, Universidad Nacional Autónoma de México, Circuito exterior S/N, Ciudad Universitaria, Ciudad de México, Mexico

[3] Departamento de Química Farmacéutica y Orgánica, Facultad de Farmacia, Universidad de Granada, Granada España.

[4] Molecular Biomedicine Laboratory, Faculty of Chemical-Biological Sciences, Autonomous University of Guerrero, Chilpancingo, Mexico. CP 39087

[5] CNRS, Laboratoire de Chimie Théorique, LCT, Sorbonne Université, Paris, France

[6] Cátedra CONACyT-Clínica de Cáncer Hereditario, Instituto Nacional de Cancerología, Mexico City, Mexico

[7] *Instituto Tecnológico y de Estudios Superiores de Occidente, Periférico Sur Manuel Gómez Morín CP. 45604 Tlaquepaque, Jalisco, México*

mmontano@uas.edu.mx

Laboratorio de Modelado Molecular y Bioinformática, *Facultad de Ciencias Químico-Biológicas, Universidad Autónoma de Sinaloa. Ciudad Universitaria s/n, CP 80010, Culiacán, Sinaloa, México*

Funding:

This work was supported by PRO_A2_018 to SM.

Disclosure statement: The funders had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript

ABSTRACT

Endosulfan is an organochlorine insecticide widely used for agricultural pest control. Many nations worldwide have restricted or completely banned it due to its extreme toxicity to fish and aquatic invertebrates. *Arthrobacter sp*. strain KW has the ability to degrade α, β endosulfan and its intermediate metabolite endosulfate; this degradation is associated with Ese protein, a two-component flavin-dependent monooxygenase (TC-FDM). Employing *in silico* tools, we obtained the 3D model of Ese protein, and our results suggest that it belongs to the Luciferase Like Monooxygenase family (LLM). Docking studies showed that the residues V59, V315, D316, T335 interact with α-endosulfan. The residues: V59, T60, V315, D316, and T335 are implicated in the interacting site with β-endosulfan, and the residues: H17, V315, D316, T335, N364, and Q363 participate in the interaction with endosulfate. Topological analysis of the electron density by means of the Quantum Theory of Atoms in Molecules (QTAIM) and the Non-Covalent Interaction (NCI) index reveals that the Ese-ligands complexes are formed manly by dispersive forces, where Cl atoms have a predominant role. As Ese is a monooxygenase member, we predict the homodimer formation. However, enzymatic studies must be developed to investigate the Ese protein's enzymatic and catalytic activity.

Keywords: Ese, Arthrobacter, Docking, Molecular dynamics simulation, α-endosulfan, β-endosulfan, endosulfate.

1. Introduction

Endosulfan (6,7,8,9,10,10-Hexachloro-1,5,5a,6,9,9a-hexahydro- 6,9-methano-2,4,3-benzodioxathiepine-3-oxide) is a cyclodiene organochlorine pesticides (OCPs) legacy of persistent organic pollutants (POPs) used extensively in crops, pest, and mite vectors of animal and human diseases, besides of public health control (Kataoka and Takagi, 2013; Khuman and Chakraborty, 2019). POPs share four basic characteristics: persistence, bioaccumulation, transport, and adverse effects (Rosenfeld and Feng, 2011; Menezes et al., 2017).

Endosulfan is semi-volatile and persistent in the environment; it is deposited in crops, soil, or water due to its long-way transport. It may remain in the soil for several years before its degradation. It is highly toxic to fish and a lesser degree, to aquatic invertebrates (Rosenfeld and Feng, 2011). Mammals are susceptible to endosulfan, which causes neurological, renal, liver, respiratory, and reproductive toxicity (Naqvi and Vaishnavi, 1993). It can produce mutations and DNA damage in mice, fish, birds, and humans (Rosenfeld and Feng, 2011). Due to its environmental persistence and risk for wildlife and human, exploring new alternatives to eliminate endosulfan from the environment is necessary.

Different methods have been reported for treating contaminated endosulfan sites, such as photocatalytic degradation (Begum and Gautam, 2011; Thomas et al., 2011), ozone oxidation (Yazgan et al., 2003), adsorption, advanced oxidation processes, and biological methods, among those, are phytoremediation, fungi, microalgae and bacterial degradation (Kataoka and Takagi, 2013; Supreeth and Raju, 2017; SureshKumar et al., 2018).

Bacterial degradation is an efficient way to remove contaminants in the most environmentally friendly way (Bhatt et al., 2019; Kataoka and Takagi, 2013; Supreeth and Raju, 2017); however, only a few enzymes have been identified participating in this biodegradation process. Ese-protein from *Arthrobacter sp.* KW, a two-component flavin-dependent monooxygenase (TC-FDM), can degrade both endosulfan isomers, α-endosulfan and β-endosulfan, and their intermediate metabolite endosulfate, which is even more persistent (Weir et al., 2006). The Esd protein from *Mycobacterium sp,* is a TC-FDM member, which degrades β-endosulfan (Sutherland et al., 2002). Both enzymes, Esd and Ese, are TC-FDM members and require reduced FMN supplied by a NAD(P)H-dependent flavin reductase (Sutherland et al., 2002; Weir et al., 2006).

Ese-protein belongs to the LLM family, which is TC-FDM; these enzymes possess a TIM-barrel domain and utilize flavin or the flavin analogue F420 as a coenzyme. Besides, dimerization has been reported on the LLM members, which suggests that Ese may also form homodimers (Cao et al., 2018).

Here, we presented the *in silico* characterization of Ese-protein and investigated Ese-protein's interacting site with α-endosulfan, β-endosulfan, and endosulfate. In addition, we explored the association of the homodimer formation of Ese-protein. The interactions responsible for the Ese-protein-endosulfan complex formation were studied by means of electronic structure calculations and quantum chemical topology analysis. Results obtained evidenced the active participation of Ese in the organochlorine insecticide degradation process.

2. Results and Discussion

2.1. Ese-protein possesses the IR regions of monooxygenases

An iterative search of Ese-like proteins was done using BLASTp (Blanco-Míguez et al., 2018), where the following crystal structures were found: LadA (A4IU28), DmoA (E9JFX9), MoxC (O34974), RcaE (A0A3B6UEK8), BdsA (Q8GRC7), and NTA-MO (Q3JIP8), all the enzymes are flavin-dependent, belonging to the LLM family (Moore and James, 1995; Fisher et al., 1996; Shima et al., 2000). The results shows the TIM-Barrel domain, and the five insertion regions (IR) characteristics of the LLM family are conserved in Ese-protein (Fig. 1) (Li et al., 2008; Okai et al., 2017; Cao et al., 2018). Interestingly, the residues F10, F55, T104, N133, S230, F245 are conserved and placed in the FMN binding site as reported in other LLM (Li et al., 2008; Okai et al., 2017; Cao et al., 2018) (Fig. 1).

2.2. The 3D model of Ese-protein conserved its structure as monooxygenase

The 3D homology model was predicted in the I-TASSER server using the following crystal structures from PDB as templates 6AK1: 3B9N, 5W48, 5TLC, 3SDO. Ese-protein possesses a TIM-barrel domain; it consists of eight α-helices and eight β strands. These structures form a solenoid, where the eight antiparallel β strands are inside the barrel, and the α-helices are exposed (Fig. 2A). Ese presents five IR, (IR1 to IR5) and a long carboxy-terminus upon the TIM-barrel, as reported in other monooxygenases (Fig. 2B) (Li et al., 2008; Okai et al., 2017; Cao et al., 2018). IR1 consists of 24 residues (M12 to D35) located between the β1 and α1; IR2 has 27 (A57 to D84) located between the β2 and α2; IR3 comprises 18 residues (T136 to

D153) located between β4 and α4; IR4 consists of 48 residues from G174 to P222, with two antiparallel hairpins between the α4 and β5; and IR5 with 84 residues from G286 to T370, located between β7 and α7 (Fig. 2B and 2C). The carboxy-terminus (450 to 475) is similar to DmoA, which is elongated (Li et al., 2008; Cao et al., 2018).

To verify that the 3D model obtained from I-TASSER has a good quality, we performed a 3D structural alignment with the crystal structures of other monooxygenases, DmoA (PDB: 6AK1), LadA (PDB: 3B9N), BdsA (PDB: 5TLC), MoxC (PDB: 1TVL), NTA-MO (PDB: 3SDO), and RcaE (PDB: 5W48) (Fig. S1). The results show that the 3D-structure of Ese-protein is conserved as monooxygenase. The double hairpin on the IR4 is characteristic in all the analyzed enzymes and participates in the dimer and tetramer formation among the monooxygenase family (Li et al., 2008; Cao et al., 2018).

The Ese-FMN was obtained by positioning the FMN in Ese-protein using the COFACTOR tool from I-TASSER (Zhang et al., 2017). The FMN is horizontally positioned on the TIM-barrel between the β4 y β5, interacting with the residues F10, F55, T104, N133, S230 and F245 (Fig. S2).

2.3. Homodimer structure analysis

LLM family enzymes are predisposed to form dimers, tetramers, and other quaternary structures (Eichhorn et al., 1999; Li et al., 2008; Cao et al., 2018), which suggests that Ese-protein could form homodimers. Therefore, protein-protein docking with the Ese-FMN was carried out. The homodimer obtained by docking presented a ΔG= -4055.4 (Fig. S3), bound by four salt-bridges and 19 hydrogen-

bonds with an interatomic distance ranging from 2.66 Å to 3.12 Å (Table S1). Ese-protein's homodimer formation is given through the IR1, IR2, IR3, and IR4, which coincides with the reported for LadA, DmoA and BdsA (Li et al., 2008; Okai et al., 2017; Cao et al., 2018). The two hairpins present in the elongated IR4 seem decisive for the homodimer formation. The first hairpin lies over the IR1, projecting towards the groove formed by the α1 and α8 of the TIM-barrel in both monomers. Contrastingly, the second hairpin projects opposite the first hairpin, interacting with the IR2 and IR3 of the second monomer (Fig. S3).

2.4. Molecular dynamics simulation of Ese-protein

To refine the 3D-structure of Ese-protein, a Molecular dynamics simulation (MDs) of 200 ns was performed for the apoenzyme (Ese) and the holoenzyme (Ese-FMN), while the homodimer (Ese-Dimer) was refined through 150 ns. Trajectories were analyzed by calculating the root mean square deviation (RMSD), root mean square fluctuation (RMSF), and the radius of gyration (Rg).

The RMSD indicates that Ese and Ese-FMN systems reach equilibrium after 50 ns of MDs. The trajectory has ±1 Å of difference, which suggests that the FMN absence does not significantly affect the stability of the mature enzyme (Fig. 3A).

2.4.1. The IR regions exhibited the principal fluctuations

The Rg shows the Ese system exhibits a lower compaction rate than the Ese-FMN (Fig. 3B); this could be due to the FMN presence on the Ese-FMN that promotes the opening of the enzymatic pocket where the cofactor is positioned and functions as a catalytic site (Lobanov et al., 2008).

Both systems exhibit five flexible zones in RMSF with a similar fluctuation pattern, corresponding to the IR of the enzyme. The first peak is located at H20-Q30, belonging to IR1, the second peak from T60-F70 belonging to IR2; the third fluctuation, from D140-Q150 inside the IR3; the fourth peak, from A180–K195 in IR4, the last one at A320-G355 inside IR5 (Fig. 3C). These regions have an essential role in the catalysis and dimer formation, as reported for other LLM family proteins (Li et al., 2008; Cao et al., 2018). The C-terminus has significant fluctuations; this could be due to its long-loop nature.

In the Ese-Dimer system, the RMSD analysis of the protein reached equilibrium at 80 ns (Fig. 3D). The Rg exhibited a similar compaction rate as the Ese and Ese-FMN (Fig. 3E). Interestingly, the RMSF showed the same five flexible regions with higher fluctuation than the Ese system (Fig. 3F).

Previous studies of the enzymes LadA, BdsA and DmoA confirmed that the IRs participate in forming the active site (Okai et al., 2017; Cao et al., 2018). In LadA IR1, IR2, IR3, and IR4 mainly constitute the wall of this pocket, where FMN is embedded in a protein scaffold formed by several residues, particularly those on IR2, IR3, and IR4 (Li et al., 2008). In Ese-protein, the IRs play an important role in forming the catalytic site due to the structurally conserved regions in the enzyme and the residues that participate in the FMN cofactor binding (Fig.S2).

Once the MDs were done, the conformation of the FMN did not change significantly. FMN interacted with F10, F55, T104, N133, and F245 residues, mainly by the isoalloxazine ring. Meanwhile, S230 interacted with the phosphate group (Fig S2). These residues seem crucial to the FMN binding with the protein,

190 which coincides with previous reports from other proteins of the LLM family (LadA,
191 DmoA, MoxC, RcaE, BdsA, NTA-MO) (Li et al., 2008; Okai et al., 2017; Cao et al.,
192 2018). Furthermore, the MDs showed that in Ese-protein, S230 seems to function
193 as an anchoring site to the phosphate group of the FMN, being vital for the ligand
194 stability within the enzyme.

195 The Ramachandran plots exhibited that the 3D model of Ese-protein before the
196 MDs presents 421 residues (89 %) inside the favoured region, 33 (7.0 %) in the
197 allowed region, and 19 (7 %) outside the allowed region (Fig. S4A). The refined
198 model (200 ns) presents 364 residues (82.9%) in the favoured region, 63 (14%) in
199 the allowed region, and 12 (2.7%) outside the allowed region (Fig. S4B). A
200 considerable decrease in the amino acids in the outlier region and an increase in
201 the stereochemical quality of the models can be noticed, indicating that certain
202 residues were refined through the MDs.

203 **2.5. Docking analysis with endosulfan and metabolites.**

204 Once we obtained the refined structures by MDs, molecular dockings calculations
205 were carried out into the active site of the Ese-FMN with the ligands: α-endosulfan,
206 β-endosulfan, and endosulfate using AutoDock software (version 4.2). The
207 resulting ligand structures were further used as initial coordinates for 50 ns MDs.
208 Ese-FMN complexes showed the following binding energies: -7.37 kcal/mol; -7.14
209 kcal/mol, and -7.28 kcal/mol, for α-endosulfan (Ese-alpha), β-endosulfan (Ese-
210 beta), and endosulfate (Ese-endosulfate), respectively (Table 1, Fig. 4). The
211 ligands were bound near FMN into the enzyme catalytic cavity. The residues

involved in protein-ligand interaction are V59, V315, D316, T335 with α-endosulfan (Fig. 4A); V59, T60, V315, D316, T335 with β-endosulfan (Fig. 4B); and H17, V315, D316, T335, N364, Q363 with endosulfate (Fig. 4C). H17 is conserved in DmoA and LadA, reporting a vital role in the substrates catalysis in LadA, and where mutagenesis caused the loss of enzymatic function (Li et al., 2008; Cao et al., 2018). Thus H17 could be crucial for the catalytic activity in Ese-protein.

2.6. Molecular dynamics simulation of Ese-protein with endosulfan and metabolites

To understand the behaviour of the protein-ligands complexes, we performed a 50 ns of MDs. The RMSD showed all systems converged at 10 ns with a difference of ± 1 Å among them, showing the ligands stabilize the Ese-FMN under the same conditions (Fig. 5A). The Rg of the three complexes did not show significant differences over time, which indicates high stability during the simulation time (Fig. 5B).

The RMSF of Ese-alpha and Ese-beta complexes showed the same five fluctuating regions as the Ese-FMN (Fig. 5C). Nevertheless, a decrease was observed in the IR regions of the Ese-ensodulfate complex; IR3, IR4 and IR5 show ± 4 Å, ± 6 Å and ± 4 Å of decrease, respectively. This behaviour could be due to the structural arrangement needed for the ligand accommodation, which interacts with several protein residues stabilizing the protein structure (Fig. 5C).

2.7. Binding free energy calculations of complex protein-ligand

Binding free energies (ΔG_{bind}) of the protein-ligand complexes were calculated from the last 20 ns of the MDs once the system reached the equilibrium using the Molecular Mechanics Poisson–Boltzmann surface area (MM/PBSA) approach (Table 2). There were a few differences between the Ese-alpha and the Ese-beta complexes' binding free energies. However, the Ese-alpha complex is energetically more favourable than the Ese-beta (-20.6407 vs -19.454 kcal/mol). The binding free energy of the Ese- endosulfate complex becomes more positive, indicating a lesser favoured complex (-14.0214 kcal/mol).

A more detailed inspection showed that non-polar contributions (ΔE_{np}= ΔE_{vdw} + ΔG_{np}) contributed primarily to the binding free energy, where van der Waals term was the most significant contribution in comparison to the non-polar solvation term (ΔG_{np}). Instead, the polar solvation term (ΔG_{PB}) was stronger unfavourable, while the electrostatic term (ΔE_{ele}) was favorable, but its contribution was minor. These findings suggest that the studied compounds mainly interact by hydrophobic interactions.

2.7.1. Most populated cluster conformation

The most-populated cluster conformation of each complex was obtained from the MDs by clustering analysis. Regarding, Ese-alpha complex did not show a higher difference against the initial conformation (RMSD= 1.262 Å), while Ese-beta (RMSD= 1.412 Å) and Ese-endosulfate showed higher differences (RMSD= 1.437 Å) at a structural level (Fig. S5). These findings coincide with the geometrical parameters RMSD and Rg. The three complexes were slightly displaced from their

initial positions concerning the binding modes. In Ese-alpha, Ese-beta, in the beginning, the sulfite portion was closer to the hydroxyl groups of the FMN, but in the most populated cluster conformation, the molecules were moved toward the flavin moiety aromatic region of the FMN. A similar phenomenon occurs to Ese-endosulfate, which initially oriented the Cl closer to the FMN towards the flavin moiety aromatic region. However, the most populated cluster conformation moves slightly away from the FMN. The quantum chemical calculations give more insight about the specific interactions that stabilize the Ese-protein-ligand complexes.

2.7.2. Topological analysis

In QTAIM, the bond critical points (BCP) and their corresponding bond paths (BP) indicate the presence of covalent and non-covalent (*e.g.*., hydrogen-bonds) interactions. The strength of each interaction can be determined quantitatively by the value of the electron density at the BCP (ρ_b), *i.e*., the larger ρ_b, the stronger the interaction. Besides, the interaction energy associated with each BCP can be approximated by the Espinosa-Molins-Lecomte equation, $E_{int} = ½ \cdot V_b$, where V_b is the potential energy density evaluated at the BCP. For the three complexes, the Laplacian value of the electron density evaluated at the BCP is positive in all cases, implying that all the intermolecular contacts are non-covalent in nature.

All the intermolecular BCP and BP formed between the ligands and the surrounding molecules in the Ese-alpha, Ese-beta and Ese-endosulfate complexes are depicted as orange dots and orange lines, respectively, in Fig. 6. The two most abundant interactions in all complexes are the non-canonical C-H···Cl and C-H···O hydrogen-bonds, which indicate the Cl atoms and the SO_3 (SO_4 in the case of Ese-

endosulfate) groups have a predominant role in the stabilization of the Ese-protein-ligand complexes. All complexes present the formation of C-H···Cl interactions between the ligands and I392 and I18, and C-H···O hydrogen-bonds between the ligands and V315. Additionally, Ese-alpha and Ese-endosulfate show more common specific interactions with identical neighbouring residues. For instance, both ligands form halogen bonds (Cl···S) between chlorine and sulphur atoms of M12. Interestingly, hydrogen-hydrogen-bonds (H···H) (Matta et al., 2003) are also found in all complexes, although in less quantity than the other type of interactions, pointing out that the contribution of H atoms in the protein-ligand complex formation is non-negligible. Even though there are charged COO^- groups near of the three complexes (D58 in Ese-alpha and Ese-beta, and D316 in Ese-endosulfate), these do not interact directly with the ligands. Instead, they show affinity to form hydrogen-bonds with more polar H atoms (a N-H···O with FMN in Ese-alpha and Ese-beta, and a O-H···O with T335 in Ese-endosulfate) (Fig. S6).

From the values of ρ_b and E_{int} (Table S2), it is observed that the strongest interactions in each complex are a N-H···O hydrogen bond in Ese-alpha (-5.0 kcal/mol), a C-H···O contact in Ese-beta (-2.6 kcal/mol), and a H···H interaction in Ese-endosulfate (-2.5 kcal/mol). These are manifested as blue or blue-green thick disks in the NCI isosurfaces depicted in Fig. S7, which are typical of strong or medium-strength localized interactions (contacts between two atoms), respectively. Most of the remaining contacts have an E_{int} lower than -2.0 kcal/mol. Moreover, they are embedded in green extended surfaces (Fig. S7) that are more abundant than the localized interactions and which are identified with regions of weak

301 delocalized interactions, *i.e*., zones of van der Waals forces. Thus, these results
302 suggest that the three ligands show a preference for hydrophobic residues, and the
303 complexes are stabilized mainly by dispersive interactions, which is consistent with
304 the fact that the van der Waals terms are predominant in the free energies of
305 binding (Table 2).

306 Regarding FMN, the three ligands interact via dispersion forces between Cl, O and
307 H with the aromatic and non-aromatic rings of the FMN, which are shown as green
308 concave surfaces in the NCI graphical index (Fig. S7). In agreement, BCPs
309 connecting the same ligand atoms with the C and N atoms of FMN were also found
310 (Fig. 6). Additionally, to the dispersive component described by the NCI analysis,
311 the ligand-FMN interactions also have an electrostatic contribution, as can be seen
312 from the electrostatic potential (ESP) of the three complexes shown in Fig. 7. The
313 positive and negative regions around the H and O atoms of the ligands,
314 respectively, interact with zones of contrary sign in the ESP of the FMN rings. Due
315 to polarization effects, the Cl atoms possess regions of positive and negative ESP
316 and, therefore, are able to interact with different parts of the FMN rings. Finally, the
317 PO_4^{-2} group of FMN, which has a very negative ESP, never interacts directly with
318 the ligands, corroborating the latter's affinity for non-polar molecules.

319 **3. Conclusions**

320 Through *in silico* analysis, we explored the 3D-structure of the Ese-protein to
321 evaluate the binding site to α-endosulfan, β-endosulfan and endosulfate. Docking,
322 MDs and quantum chemical studies showed that the residues involved in the

interaction of Ese-protein with endosulfan and its metabolites fall into the catalytic site of Ese-protein, suggesting that Ese could degrade endosulfan and its metabolites. The topological analysis revealed that the three ligands prefer hydrophobic residues, and the protein-ligand complexes are mainly stabilised by dispersive interactions. Besides, Ese-protein possesses conserved motifs that contribute to the degradation of organochlorine compounds present in other monooxygenase enzymes. These results allow establishing the molecular basis for the interaction of this protein with this compound; however, further experimental studies are needed to demonstrate the endosulfate and metabolites degradation by Ese protein.

4. Methods

4.1. Multiple Alignments of Ese-protein

Multiple alignments were done in BLASTp of The National Center for Biotechnology Information (NCBI) against the PDB database with the PSI-BLAST algorithm (Yoshinaga et al., 2014; Blanco-Míguez et al., 2018) to find crystal structures similar to Ese from *Arthrobacter*; after that, multiple sequence alignments were carried out in Clustal Omega, using default parameters (Madeira et al., 2019).

4.2. 3D-structure of Ese-protein and homodimer structure

The 3D-structure of Ese-protein from *Arthrobacter sp.* KW was predicted from their amino acid sequence downloaded from UniProtKB, ID Q45N64 (UniProt, 2019), and submitted to the I-TASSER server (Zhang, 2009), where the model with the higher C-

345 score was selected (Zhang, 2009; Roy et al., 2010; Yang et al., 2015). The following
346 crystal structures were used as templates from I-TASSER to obtain the 3D model of
347 Ese-protein: 6AK1 (DmoA from *Hyphomicrobium sulfonivorans*, 3B9N (LadA from
348 *Geobacillus thermodenitrificans*, 5W48 (RcaE from *Herbiconiux*), 5TLC (BdsA from
349 *Bacillus subtilis* WU-S2B), 3SDO (*Burkholderia pseudomallei*), 1TVL (YtnJ from
350 *Bacillus subtilis*), 5DQP (EmoA from *Chelativorans sp*. BNC1), 5WAN (RutA
351 *Escherichia coli)*, 1NQK (Alkanesulfonate monooxygenase from *Escherichia coli)* and
352 2WGK (3,6- diketocamphane 1,6 monooxygenase from *Pseudomonas putida*). To
353 obtain the correct position of the FMN cofactor in Ese-protein, the COFACTOR tool
354 from I-TASSER was used, where model with the higest C-score was selected.
355 The structural alignment of Ese-protein 3D model and the crystal structures from
356 monooxygenase proteins was performed in Chimera using the following crystal
357 structures PDB: 6AK1, 3B9N, 5TLC, 1TVL, 3SDO and 5W48 (Pettersen et al., 2004).
358 After that, an MDs of 200 ns was run for the Ese and Ese-FMN. The refined structure
359 of the Ese-FMN was used to predict the homodimer formation on ClusPro webserver
360 (Comeau et al., 2004; Kozakov et al., 2017). The predicted homodimer was refined
361 through 150 ns of MDs. Further, the Ese-FMN was used for docking studies with α-
362 endosulfan, β-endosulfan, and endosulfate in Autodock. All 3D-structure visualizations
363 were done employing VMD (Visual Molecular Dynamics) software (Humphrey et al.,
364 1996).

365 **4.3. Molecular Dynamics Simulations**

366 Molecular dynamics simulations (MDs) were carried out using NAMD 2.8 (Phillips
367 et al., 2005) using GPU-CUDA with video card graphics NVIDIA Tesla
368 C2070/Tesla C2075. The force fields CHARMM22 and CHARMM27 (MacKerell et

al., 1998) were used to create the topologies for protein and lipids, respectively. The TIP3 model was used for water molecules. The system was solvated using the *psfgen* software in the VMD program (Humphrey et al., 1996), adding 13828, 16946 water molecules and 17,19 sodium ions to Ese and Ese-FMN systems were added respectively. For Ese-dimer, 27048 water molecules and 38 sodium ions were added. The systems were minimized for 1000 steps, followed by equilibrium under constant temperature and pressure (NPT) conditions for one ns with protein and lipid atoms restrained. Afterwards, 150 ns MDs were run, considering Ese-protein as soluble, without position restraints under PBC (periodic boundary conditions) and using an (Isothermal-Isobaric) NPT ensemble at 301.15 K, since it is the temperature where *Arthorbacter* KW, reported a successful growth (Weir et al., 2006).

4.4. Trajectory analysis of MDs

The stability and conformational changes of the system's trajectory were evaluated by analyzing RMSD, RMSF, and the Rg, calculated with the Carma program (Glykos, 2006), and the representative conformation used for docking purposes (Koukos and Glykos, 2013). Molecular graphics were performed in SigmaPlot 12.0. We evaluated the quality of the MDs's last frame compared to the I-TASSER obtained model, using Ramachandran's plot on the RAMPAGE server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php).

4.5. Docking analysis

The ligand structures were retrieved from the Zinc15 database with ID: ZINC2557938 for α-endosulfan, ZINC18069227 for β-endosulfan, and ZINC102213075 for endosulfate (Sterling and Irwin, 2015). Ligands were optimized

with Gaussian 9 (Frisch et al., 2016). Molecular docking of α-endosulfan, β-endosulfan, and endosulfate was performed on Ese-protein using AutoDock 4.2 (Morris et al., 2009). Both receptors and ligands were prepared with Autodock Tools 1.5.7 (Morris et al., 2009). Kollman and Gasteiger charges were assigned for proteins and ligands, respectively. The docking was performed with a grid box with a grid dimension of 40 Å and grid spacing of 0.35 Å, using the Lamarckian Genetic Algorithm as scoring sampling with an initial randomized population of 100 individuals and a maximum number of energy evaluation of 1×10^7 and 100 runs. The lowest energy binding position predicted was considered for the analysis. The visualization and analysis were done in AutodockTools 1.5.7 (Morris et al., 2009). We obtained the ligands' initial coordinates from docking calculation and performed 50 ns of MDs of each protein-ligand complex, carried out in NAMD 2.8 software (Phillips et al., 2005) as described above. The topologies and parameters for the substrates were generated with the SwissParam server (Zoete et al., 2011). The MDs were performed in the Laboratory of Molecular Modeling and Bioinformatics at Facultad de Ciencias Químico Biológicas de la Universidad Autónoma de Sinaloa, and in the Hybrid Cluster Xiuhcoatl (http://clusterhibrido.cinvestav.mx) of the CINVESTAV-IPN, México.

4.6. Binding free energy calculation of complex protein-ligand

The MM/PBSA approach was employed to estimate the binding free energy of the protein-ligand complexes. The calculations were performed with the CAFE 1.0 software (Liu and Hou, 2016). For the calculations, the last 20 ns of the MDs of each system were sampled, analyzing 200 frames. MM/PBSA is a method that

combines molecular mechanics and continuum solvent, where the binding free energy (ΔG_{bind}) is estimated as follows:

$$\Delta G_{bind} = G_{complex} - (G_{protein} - G_{ligand}) \quad (1)$$

Where, $G_{complex}$ is the energy of the protein-ligand complex, and $G_{protein}$ and Gl_{igand} correspond to the unbound protein and free ligand in the complex, respectively. The individual free energies of each one of the components (1), are estimated as follows:

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{solv} - T\Delta S \quad (2)$$

Where, ΔE_{MM} is the vacuum molecular mechanics energy due to the interaction of the ligand and protein in vacuum, ΔG_{solv} is the solvation energy, and $T\Delta S$ corresponds to the entropic contribution, which is not calculated in the present work; because it involves a higher computational cost and the current methods remain inaccurate, and it has been suggested that this item can be omitted because it does not improve the results in large sets (Homeyer and Gohlke, 2012). Therefore, in the present work, only effective energy is calculated. In addition, MMPBSA methods render better, more accurate results if effective free energy is determined (Genheden and Ryde, 2015), therefore, only the effective energy was calculated. Further, ΔE_{MM} and ΔG_{solv} are calculated as follows:

$$\Delta E_{MM} = \Delta E_{vdW} + \Delta E_{elec}$$

$$\Delta G_{solv} = \Delta G_{PB} + \Delta G_{np}$$

436 ΔE_{vdW} and ΔE_{elec} are the van der Waals and electrostatic, respectively; these terms
437 are contributions due to the vacuum binding between ligand and receptor. In
438 contrast, ΔG_{PB} and ΔG_{np} are the electrostatic solvation free energy and non-polar
439 contributions to the solvation free energy (Kollman et al., 2000; Genheden and
440 Ryde, 2015).

441 **4.7 Clustering analysis**

442 Aiming the obtention of a representative conformation from the MD simulations of
443 the ligand-protein complexes to study the structural changes occurred and the
444 interactions established between the protein and the ligands, a method to reduce
445 the dimension of the conformation was used. Since MD simulations generate
446 thousands of conformations of the systems studied, one of the major challenges is
447 reducing dimensionality of the conformational space and extract accurate
448 information for further calculations (De Paris et al., 2015). Currently, there are
449 available several statistical methods to reduce the conformational space and allow
450 to select representative conformations that captures the average properties of the
451 system, such as principal component analysis (PCA), clusterin analysis or optimal
452 wavelet signal compression algorithm (OWSCA), where this last one was
453 successfully used to extract conformations for further carry out quantum mechanics
454 (QM) calculations (Wolf and Kirschner, 2013; Gonçalves et al., 2016). Here,
455 clustering analysis was selected as method of reduction because it groups together
456 similar conformations into clusters. Therefore, the most populated cluster of
457 conformations can be identified and use this for further studies (Abramyan et al.,
458 2016). Specifically, the GROMOS algorithm was employed in the current study,

459 which allow the obtention of representative ensembles by geometrical similarity
460 using an RMSD cutoff of 2.0 Å (Lindahl et al., 2001). Further, the most populated
461 cluster conformation of each complex was used for quantum calculations.

462 **4.8 Quantum chemical calculations**

463 For the quantum chemical calculations, a model was constructed for Ese alpha,
464 Ese-beta and Ese-endosulfate from a representative structure of the clustering
465 analysis, considering all residues whose atoms lay within a 5 Å radius around the
466 ligand. By integrating the electron density in the NCI isosurfaces at different radii
467 for the 3 systems (Figure S8), it is observed that beyond a cutoff radius of 4 Å no
468 relevant change is observed. It has been shown in benchmark studies that the
469 values of these integrations show a strong correlation with interaction energies
470 computed at the quantum-mechanical golden standard CCSD(T)/CBS (Coupled
471 Cluster Single-Double and perturbative Triple/Complete Basis Set) level of theory
472 (Boto et al., 2020). Also, the NCI integrals have shown to be useful for a
473 quantitative assessment of ligand-protein interactions (Peccati, 2020). Thus, it is
474 expected that all Ese-ligand interactions beyond 5 Å will be negligible for the
475 purpose of the topological analysis.

476 The terminal CO and NH functional groups of the cut residues were completed with
477 OH and H groups, respectively. Single point calculations (on the model
478 geometries) were performed with a B3LYP/def2-SVP level of theory with the
479 Gaussian 16 program (Frisch et al., 2016). A topological study of the DFT-
480 computed electron density was carried out by means of the Non-Covalent

Interaction index (NCI) (Johnson et al., 2010), the Quantum Theory of Atoms in Molecules (QTAIM) framework (Matta et al., 2003), and the electrostatic potential, in order to characterize the intermolecular interactions that stabilize the protein-ligand complexes. The B3LYP functional was selected because it has been observed that hybrid-GGA functionals are needed for a correct description of the electron density (and its derivatives), which is the fundamental property studied in topological analysis (Medvedev et al., 2017). Besides, the use of this functional along with double z basis sets has been successful in ligand-protein interaction studies (Kollar and Frecer, 2017; Frontera and Bauza, 2021; Gervasoni et al., 2022). Since no interaction energies were computed with the DFT method (basis set superposition error is irrelevant), the def2-SVP basis set provides an optimal cost-benefit relationship. The Multiwfn (Lu and Chen, 2012) and NCIPLOT4 (Boto et al., 2020) software were employed for the QTAIM and NCI calculations, respectively. The results were visualized with the VMD (Humphrey et al., 1996) molecular graphics program.

References

Abramyan, T.M., Snyder, J.A., Thyparambil, A.A., Stuart, S.J., Latour, R.A., 2016. Cluster analysis of molecular simulation trajectories for systems where both conformation and orientation of the sampled states are important. Journal of computational chemistry 37, 1973-1982.

Begum, A., Gautam, S.K., 2011. Application of photocatalysis for the degradation of endocrine disrupting chemicals: Endosulphan and lindane. Indian J Environ Protect 31, 658-665.

Blanco-Míguez, A., Fdez-Riverola, F., Sánchez, B., Lourenço, A., 2018. BlasterJS: A novel interactive JavaScript visualisation component for BLAST alignment results. PLoS One 13, e0205286-e0205286.

Boto, R.A., Peccati, F., Laplaza, R., Quan, C., Carbone, A., Piquemal, J.-P., Maday, Y., Contreras-García, J., 2020. NCIPLOT4: Fast, Robust, and Quantitative Analysis of Noncovalent Interactions. Journal of Chemical Theory and Computation 16, 4150-4158.

510 Cao, H.Y., Wang, P., Peng, M., Shao, X., Chen, X.L., Li, C.Y., 2018. Crystal
511 structure of the dimethylsulfide monooxygenase DmoA from Hyphomicrobium
512 sulfonivorans. Acta Crystallogr F Struct Biol Commun 74, 781-786.
513 Comeau, S.R., Gatchell, D.W., Vajda, S., Camacho, C.J., 2004. ClusPro: an
514 automated docking and discrimination method for the prediction of protein
515 complexes. Bioinformatics 20, 45-50.
516 De Paris, R., Quevedo, C.V., Ruiz, D.D., Norberto de Souza, O., 2015. An
517 Effective Approach for Clustering InhA Molecular Dynamics Trajectory Using
518 Substrate-Binding Cavity Features. PLoS One 10, e0133172.
519 Eichhorn, E., van der Ploeg, J.R., Leisinger, T., 1999. Characterization of a Two-
520 component Alkanesulfonate Monooxygenase from Escherichia coli*.
521 Journal of Biological Chemistry 274, 26639-26646.
522 Fisher, A.J., Thompson, T.B., Thoden, J.B., Baldwin, T.O., Rayment, I., 1996. The
523 1.5-A resolution crystal structure of bacterial luciferase in low salt conditions. J Biol
524 Chem 271, 21956-21968.
525 Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A.,
526 Cheeseman, J.R., Scalmani, G., Barone, V., Petersson, G.A., Nakatsuji, H., Li, X.,
527 Caricato, M., Marenich, A.V., Bloino, J., Janesko, B.G., Gomperts, R., Mennucci,
528 B., Hratchian, H.P., Ortiz, J.V., Izmaylov, A.F., Sonnenberg, J.L., Williams, Ding,
529 F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T.,
530 Ranasinghe, D., Zakrzewski, V.G., Gao, J., Rega, N., Zheng, G., Liang, W., Hada,
531 M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T.,
532 Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery Jr., J.A.,
533 Peralta, J.E., Ogliaro, F., Bearpark, M.J., Heyd, J.J., Brothers, E.N., Kudin, K.N.,
534 Staroverov, V.N., Keith, T.A., Kobayashi, R., Normand, J., Raghavachari, K.,
535 Rendell, A.P., Burant, J.C., Iyengar, S.S., Tomasi, J., Cossi, M., Millam, J.M.,
536 Klene, M., Adamo, C., Cammi, R., Ochterski, J.W., Martin, R.L., Morokuma, K.,
537 Farkas, O., Foresman, J.B., Fox, D.J., 2016. Gaussian 16 Rev. C.01. Wallingford,
538 CT.
539 Frontera, A., Bauza, A., 2021. Biological halogen bonds in protein-ligand
540 complexes: a combined QTAIM and NCIPlot study in four representative cases.
541 Org Biomol Chem 19, 6858-6864.
542 Genheden, S., Ryde, U., 2015. The MM/PBSA and MM/GBSA methods to estimate
543 ligand-binding affinities. Expert Opinion on Drug Discovery 10, 449-461.
544 Gervasoni, S., Spencer, J., Hinchliffe, P., Pedretti, A., Vairoletti, F., Mahler, G.,
545 Mulholland, A.J., 2022. A multiscale approach to predict the binding mode of
546 metallo beta-lactamase inhibitors. Proteins 90, 372-384.
547 Glykos, N.M., 2006. Software news and updates. Carma: a molecular dynamics
548 analysis program. J Comput Chem 27, 1765-1768.
549 Gonçalves, M.A., Santos, L.S., Prata, D.M., Peixoto, F.C., da Cunha, E.F.F.,
550 Ramalho, T.C., 2016. Optimal wavelet signal compression as an efficient
551 alternative to investigate molecular dynamics simulations: application to thermal
552 and solvent effects of MRI probes. Theoretical Chemistry Accounts 136.
553 Homeyer, N., Gohlke, H., 2012. Free Energy Calculations by the Molecular
554 Mechanics Poisson-Boltzmann Surface Area Method. Mol Inform 31, 114-122.
555 Humphrey, W., Dalke, A., Schulten, K., 1996. VMD: Visual molecular dynamics.
556 Journal of Molecular Graphics 14, 33-38.

Johnson, E.R., Keinan, S., Mori-Sánchez, P., Contreras-García, J., Cohen, A.J., Yang, W., 2010. Revealing Noncovalent Interactions. Journal of the American Chemical Society 132, 6498-6506.
Kataoka, R., Takagi, K., 2013. Biodegradability and biodegradation pathways of endosulfan and endosulfan sulfate. Appl Microbiol Biotechnol 97, 3285-3292.
Khuman, S.N., Chakraborty, P., 2019. Air-water exchange of pesticidal persistent organic pollutants in the lower stretch of the transboundary river Ganga, India. Chemosphere 233, 966-974.
Kollar, J., Frecer, V., 2017. How accurate is the description of ligand-protein interactions by a hybrid QM/MM approach? J Mol Model 24, 11.
Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D.A., Cheatham, T.E., 2000. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. Accounts of Chemical Research 33, 889-897.
Koukos, P.I., Glykos, N.M., 2013. Grcarma: A fully automated task-oriented interface for the analysis of molecular dynamics trajectories. J Comput Chem 34, 2310-2312.
Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padhorny, D., Yueh, C., Beglov, D., Vajda, S., 2017. The ClusPro web server for protein-protein docking. Nat Protoc 12, 255-278.
Li, L., Liu, X., Yang, W., Xu, F., Wang, W., Feng, L., Bartlam, M., Wang, L., Rao, Z., 2008. Crystal Structure of Long-Chain Alkane Monooxygenase (LadA) in Complex with Coenzyme FMN: Unveiling the Long-Chain Alkane Hydroxylase. Journal of Molecular Biology 376, 453-465.
Lindahl, E., Hess, B., van der Spoel, D., 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. Journal of Molecular Modeling 7, 306-317.
Liu, H., Hou, T., 2016. CaFE: a tool for binding affinity prediction using end-point free energy methods. Bioinformatics 32, 2216-2218.
Lobanov, M.Y., Bogatyreva, N.S., Galzitskaya, O.V., 2008. Radius of gyration as an indicator of protein structure compactness. Molecular Biology 42, 623-628.
Lu, T., Chen, F., 2012. Multiwfn: A multifunctional wavefunction analyzer. Journal of Computational Chemistry 33, 580-592.
MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, W.E., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., Karplus, M., 1998. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. The Journal of Physical Chemistry B 102, 3586-3616.
Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., Lopez, R., 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 47, W636-W641.
Matta, C.F., Hernandez-Trujillo, J., Tang, T.H., Bader, R.F., 2003. Hydrogen-hydrogen bonding: a stabilizing interaction in molecules and crystals. Chemistry 9, 1940-1951.

Medvedev, M.G., Bushmarinov, I.S., Sun, J., Perdew, J.P., Lyssenko, K.A., 2017. Density functional theory is straying from the path toward the exact functional. Science 355, 49-52.
Menezes, R.G., Qadir, T.F., Moin, A., Fatima, H., Hussain, S.A., Madadin, M., Pasha, S.B., Al Rubaish, F.A., Senthilkumaran, S., 2017. Endosulfan poisoning: An overview. J Forensic Leg Med 51, 27-33.
Moore, S.A., James, M.N.G., 1995. Structural Refinement of the Non-fluorescent Flavoprotein fromPhotobacterium leiognathiat 1.60 Å Resolution. Journal of Molecular Biology 249, 195-214.
Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J., 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of computational chemistry 30, 2785-2791.
Naqvi, S.M., Vaishnavi, C., 1993. Bioaccumulative potential and toxicity of endosulfan insecticide to non-target animals. Comparative Biochemistry and Physiology Part C: Comparative Pharmacology 105, 347-361.
Okai, M., Lee, W.C., Guan, L.J., Ohshiro, T., Izumi, Y., Tanokura, M., 2017. Crystal structure of dibenzothiophene sulfone monooxygenase BdsA from Bacillus subtilis WU-S2B. Proteins 85, 1171-1177.
Peccati, F., 2020. NCIPLOT4 Guide for Biomolecules: An Analysis Tool for Noncovalent Interactions. J Chem Inf Model 60, 6-10.
Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry 25, 1605-1612.
Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., Schulten, K., 2005. Scalable molecular dynamics with NAMD. J Comput Chem 26, 1781-1802.
Rosenfeld, P.E., Feng, L.G.H., 2011. 1 - Definition of Hazardous Waste. in: Rosenfeld, P.E., Feng, L.G.H. (Eds.). Risks of Hazardous Wastes. William Andrew Publishing, Boston, pp. 1-10.
Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5, 725-738.
Shima, S., Warkentin, E., Grabarse, W., Sordel, M., Wicke, M., Thauer, R.K., Ermler, U., 2000. Structure of coenzyme F(420) dependent methylenetetrahydromethanopterin reductase from two methanogenic archaea. J Mol Biol 300, 935-950.
Sterling, T., Irwin, J.J., 2015. ZINC 15--Ligand Discovery for Everyone. J Chem Inf Model 55, 2324-2337.
Supreeth, M., Raju, N.S., 2017. Biotransformation of chlorpyrifos and endosulfan by bacteria and fungi. Appl Microbiol Biotechnol 101, 5961-5971.
SureshKumar, P., Thomas, J., Poornima, V., 2018. Structural insights on bioremediation of polycyclic aromatic hydrocarbons using microalgae: a modelling-based computational study. Environ Monit Assess 190, 92.
Sutherland, T.D., Horne, I., Russell, R.J., Oakeshott, J.G., 2002. Gene cloning and molecular characterization of a two-enzyme system catalyzing the oxidative detoxification of beta-endosulfan. Appl Environ Microbiol 68, 6237-6245.

Thomas, J., Kumar, K.P., Chitra, K.R., 2011. Synthesis of Ag Doped Nano TiO2 as Efficient Solar Photocatalyst for the Degradation of Endosulfan. Advanced Science Letters 4, 108-114.
UniProt, C., 2019. UniProt: a worldwide hub of protein knowledge. Nucleic acids research 47, D506-D515.
Weir, K.M., Sutherland, T.D., Horne, I., Russell, R.J., Oakeshott, J.G., 2006. A single monooxygenase, ese, is involved in the metabolism of the organochlorides endosulfan and endosulfate in an Arthrobacter sp. Appl Environ Microbiol 72, 3524-3530.
Wolf, A., Kirschner, K.N., 2013. Principal component and clustering analysis on molecular dynamics data of the ribosomal L11.23S subdomain. J Mol Model 19, 539-549.
Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER Suite: protein structure and function prediction. Nat Methods 12, 7-8.
Yazgan, M.S., Kinaci, C., Toroz, I., 2003. Ozonation of aqueous solution of alpha endosulfan. J Environ Sci Health B 38, 843-853.
Yoshinaga, T., Sekijima, Y., Koyama, S., Maruyama, K., Yoshida, T., Kato, T., Ikeda, S., 2014. Clinical and radiological findings of a cerebrotendinous xanthomatosis patient with a novel p.A335V mutation in the CYP27A1 gene. Intern Med 53, 2725-2729.
Zhang, C., Freddolino, P.L., Zhang, Y., 2017. COFACTOR: improved protein function prediction by combining structure, sequence and protein-protein interaction information. Nucleic Acids Res 45, W291-W299.
Zhang, Y., 2009. I-TASSER: fully automated protein structure prediction in CASP8. Proteins 77 Suppl 9, 100-113.
Zoete, V., Cuendet, M.A., Grosdidier, A., Michielin, O., 2011. SwissParam: a fast force field generation tool for small organic molecules. J Comput Chem 32, 2359-2368.

Author Contributions: This work was carried out with the collaboration of all the authors. S.M., E.A.C., B.L.R., Y.S.L., J.C.G., C.B.R., J.A.G.T., A.P.T.,and B.C.P. Analyzed the data and wrote the manuscript. S.M., E.A.C., B.L.R., Y.S.L., and J.C.G, carried out theoretical studies results analysis and wrote the manuscript. S.M. wrote the final version of the manuscript and project design. All authors had read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Acknowledgements

686 The author wants to express their gratitude to LANCAD for the supercomputer time
687 support, and we also are grateful to Hector Oliver-Hernandez for providing us
688 valuable technical support. YSL and BLR thanks to SECTEI (secretaria de
689 Educación, Ciencia, Tecnología e Innovación de la Ciudad de México) for the
690 financial support. This work was supported by project PROFAPI 2015/182 (SM).

691

Figures

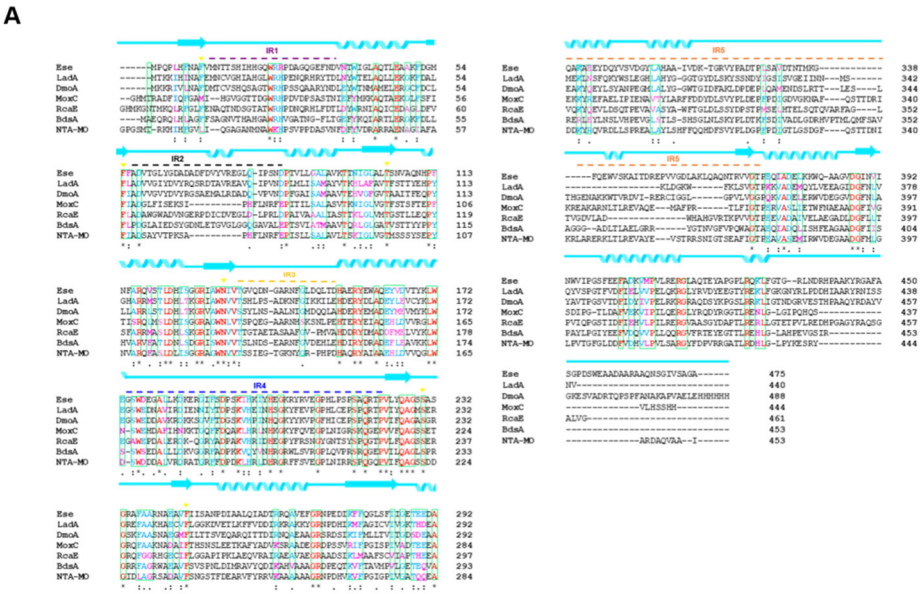
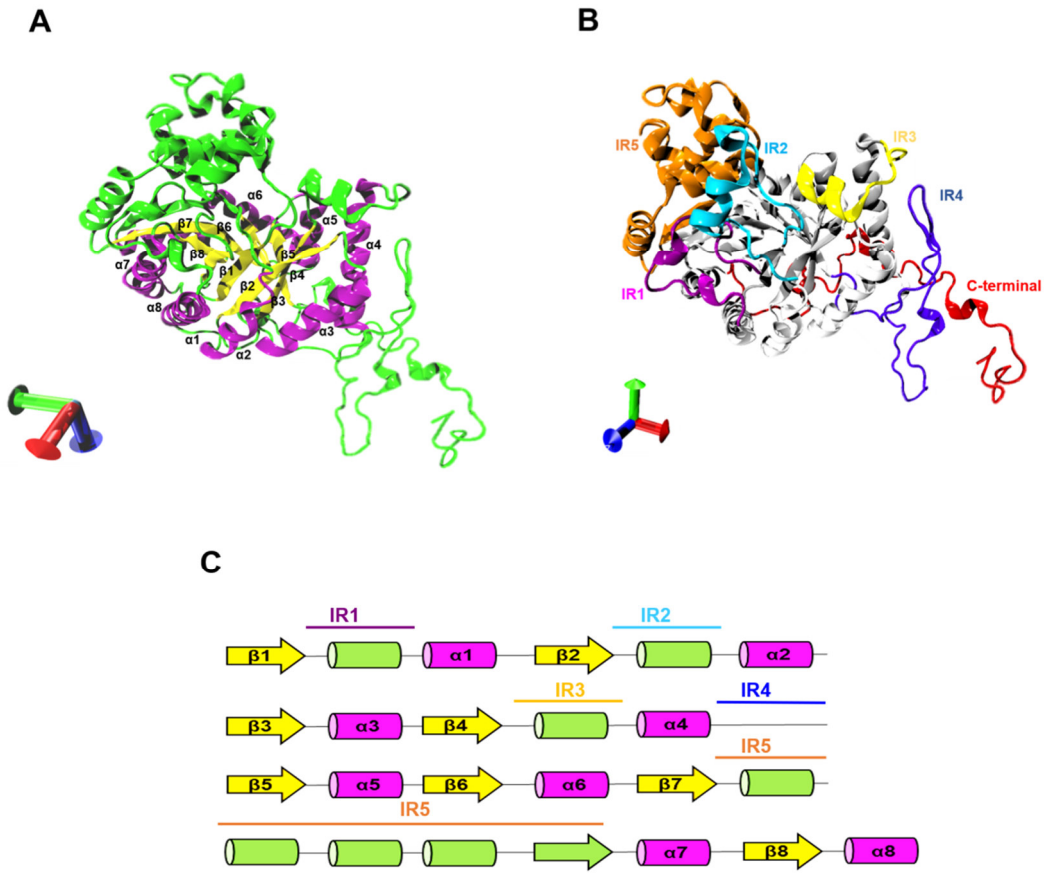


Figure 1. Multiple Sequence alignment of Ese with other bacterial luciferase monooxygenases family. Highly conserved regions are shown in green boxes. The secondary structure is above the alignment in blue. The head of the arrows (yellow) indicates residues reported in FMN binding. The IR regions are indicated as a dashed line.



705

706 **Figure 2. Three-dimensional model of the Ese protein**. A) 3D model of Ese
707 protein. The TIM-barrel is represented by α-helices (purple) and β-sheets (yellow);
708 the rest of the structure is shown in green. B) Stereo representation of Ese focused
709 on the insertion regions' conformation, where the IR1 to IR5 are coloured in
710 magenta, cyan, yellow, purple, and orange, respectively; the C-terminal extension
711 is coloured red, and the TIM-barrel is shown in grey. C) Cartoon diagram of Ese
712 protein and the corresponding insertion regions: yellow arrows: β-sheets; cylinders:
713 α-helix. Axes: X: red Y: green Z: blue.

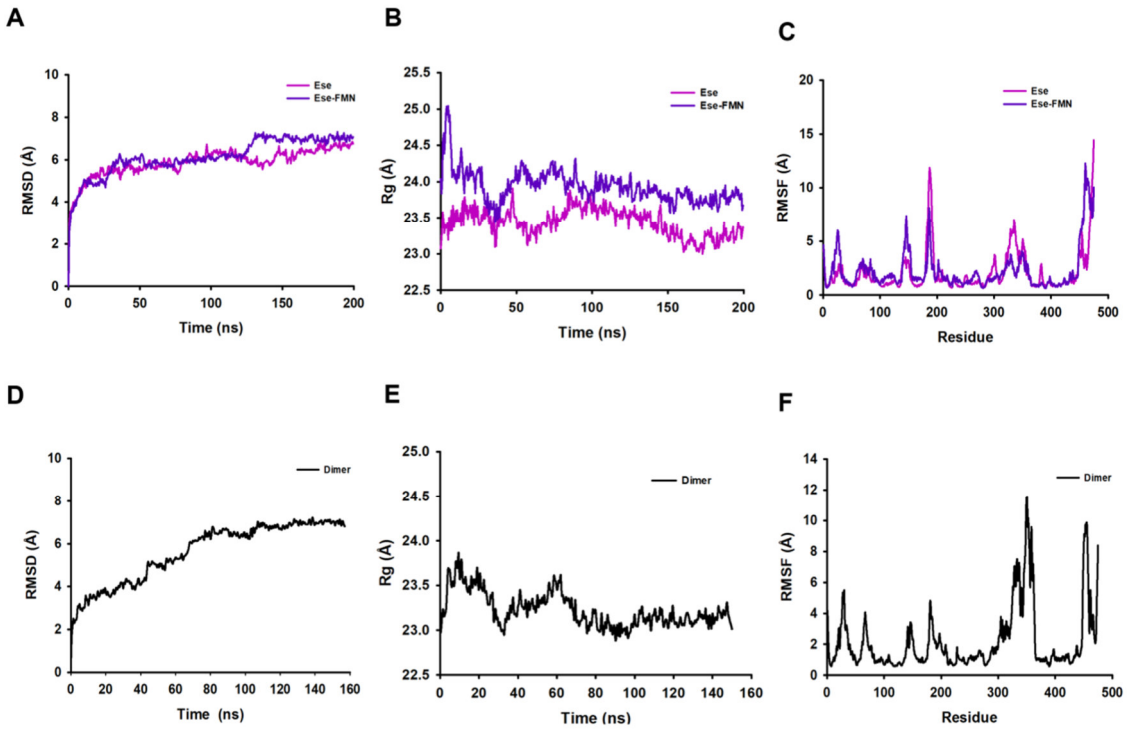


Figure 3. Trajectory analysis of molecular dynamics simulation of Ese and Dimer systems. A) RMSD. B) Rg. C) RMSF. The calculations for this system were done at 200 ns simulation. Ese trajectory (violet), Ese-FMN trajectory (purple). D) RMSD of the Dimer. B) RMSF of the Dimer. C) Rg of the Dimer. All calculations were performed after a simulation time of 150 ns.

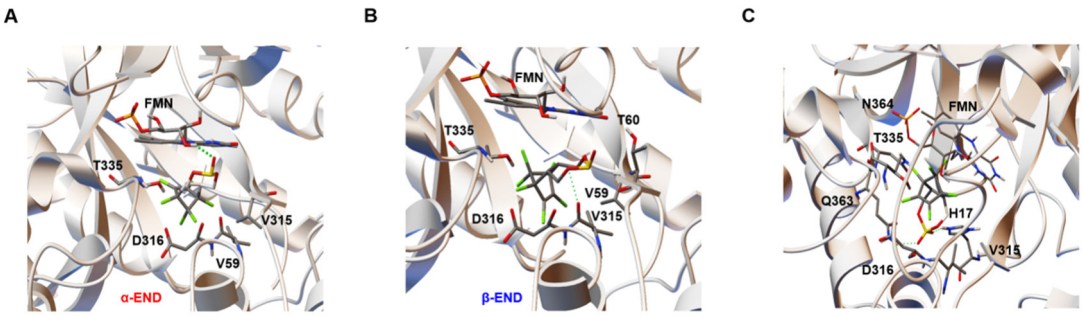
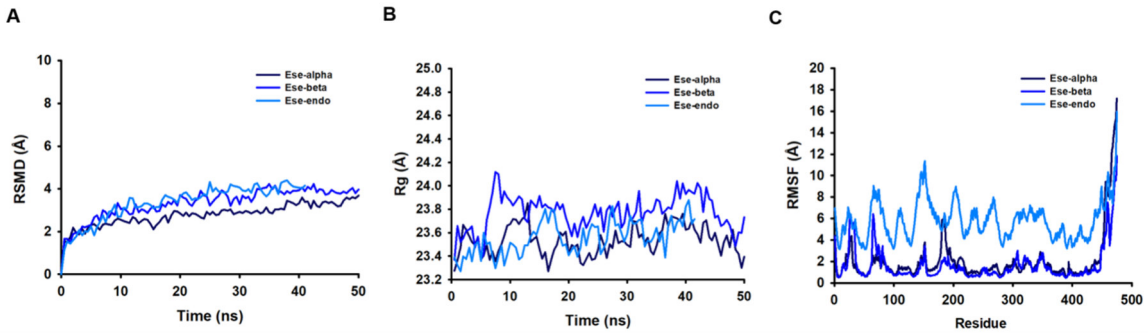


Figure 4. Docking analysis of Ese-FMN with endosulfan and its metabolites. A) Ese-FMN docked with α-endosulfan. B) Ese-FMN docked with β-endosulfan. C) Ese-FMN docked with endosulfan sulfate. Hydrogen bonds are represented as green dot lines.

Table 1. Docking results and interacting residues of Ese enzyme getting by Autodock

Model	Substrate	Binding energy (kcal/mol)	Residue
ESE	α-endosulfan	-7.37	V59, V315, D316, T335, FMN*
	β-endosulfan	-7.14	V59, T60, V315*, D316, T335
	Endosulfate	-7.28	H17*, V315, D316, T335, N364, Q363*



736

737 **Figure 5. Trajectory analysis of molecular dynamics simulation of Ese**
738 **protein with endosulfan and metabolites**. A) RMSD. B) RMSF. C) Rg. All
739 calculations were performed after 50 ns of MD simulation. Ese-alpha (black), Ese-
740 beta (blue), and Ese-endo-complexes (light blue).

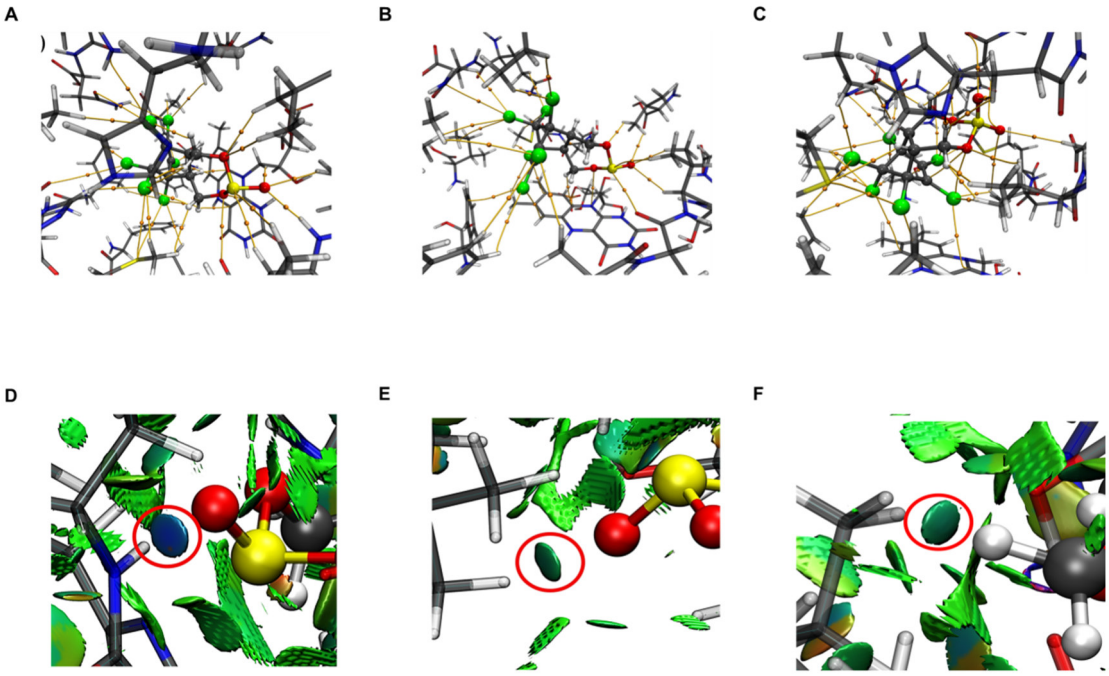
741

742 **Table 2**. Binding free energy components of complexes between ligands and Ese-
743 FMN protein (in units of kcal/mol).

Complex	ΔE_{vdw}	ΔE_{ele}	ΔG_{PB}	ΔG_{np}	ΔG_{bind}
Ese-Alpha	-43.7806	-9.0074	36.8318	-4.6845	-20.6407
Ese-Beta	-35.7416	-2.9267	23.7527	-4.5384	-19.454
Ese-Endo	-43.5032	-9.1812	43.291	-4.628	-14.0214

744 All energies are averaged over 200 snapshots at time intervals of 100 ps from the last 20
745 ns-long MD simulations and are in kcal/mol.
746

747



748

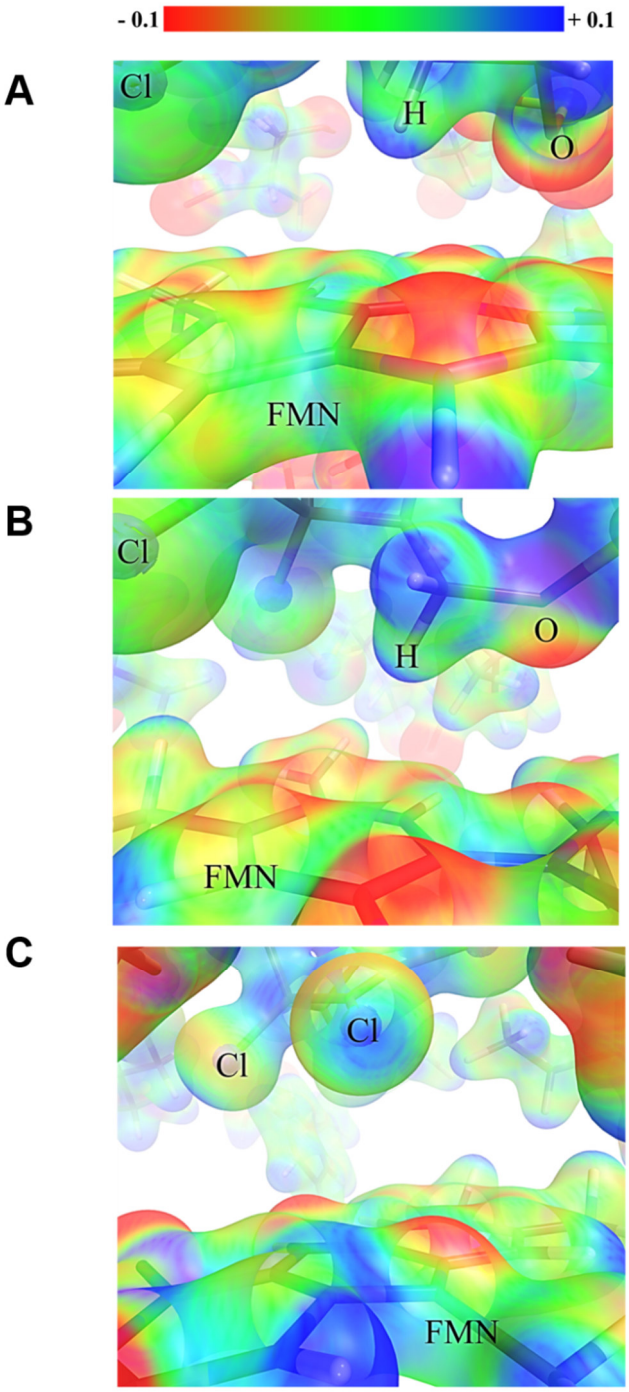
749 **Figure 6. Intermolecular specific interactions.** A) Ese-alpha B) Ese-beta, and C)

750 Ese-endo. Interactions are indicated by the BCPs and BPs (orange dots and lines,

751 respectively).

752

753



754

755 **Figure 7.** Electrostatic potential of the complexes (drawn at 0.05 a. u.). A) Ese-
756 alpha. B) Ese-beta. C) Ese-endo. The scale is indicated at the top of the figure.