PCCP

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Phys. Chem. Chem. Phys.*, 2025, **27**, 18665

Towards nature-inspired materials for adsorbing pesticides: a multi-stage computational approach

J. R. C. Santos, P. E. Abreu and J. M. C. Marques*

Received 15th April 2025,
Accepted 26th July 2025

DOI: 10.1039/d5cp01445j

rsc.li/pccp

The design of new materials that can be effective for adsorbing pesticides constitutes an important contribution for water remediation. Amino acids are the building blocks of proteins that bind pesticides in living organisms, and thus, they are expected to contribute to improve the adsorption properties of materials for water remediation. In this work, we propose a multi-stage computational strategy, based on docking, molecular dynamics (MD) simulations and electronic-structure calculations, to unveil relevant interactions between pesticides and amino acids from typical target proteins. This allows us to obtain detailed molecular-based insight about the binding complex and constitutes a straightforward procedure to select amino acids that can be effective for the adsorption of pesticides. As a case study, we applied the methodology to imidacloprid (IMI), a neonicotinoid insecticide used worldwide, and the *Aplysia californica* acetylcholine-binding protein as the target biomolecule. The most promising amino acids were then used to functionalize monomers and trimers of chitosan and the ability of the resulting model systems to adsorb IMI was assessed by MD simulations.

1. Introduction

Throughout the past century, as an attempt to optimize food production, chemical pesticides have been extensively used worldwide for the prevention, elimination, or control of pests in agriculture.[1] According to the data made available by the Food and Agriculture Organization of the United Nations (FAO), the global use of pesticides in agriculture, in 2021, reached 3.5 million tonnes.[1,2] From these applied pesticides, only a minimal portion reach the final target, with the remaining dispersing and accumulating in the environment. On top of that, many of these pesticides also exhibit great stability over time, allowing them to be transported by water and air to areas distant from their origin, causing severe soil and water pollution.[3,4] As a result, these compounds can directly or indirectly impact a wide range of organisms, including birds, wildlife, domestic animals, fish, and livestock, and some pose a considerable risk to human health.[5,6] For instance, imidacloprid (IMI) was the first commercially produced neonicotinoid[7] and, since then, it has been extensively employed in agriculture to combat insect pests, while it appears to be also harmful for several other animals.[8–12]

To mitigate this challenging environmental problem, advanced methodologies for the detection of contaminants[13] and water remediation[14] based on new adsorption materials need to be developed. A first step in such an effort is to identify molecules that are potentially promising for building up a material with the desired adsorption properties. This requires deep knowledge about the interactions between pesticides and the candidate molecules, which can be achieved by employing theoretical methods. Actually, computational chemistry has a panoply of techniques that may be applied to calculate relevant properties and provide molecular-level insight to the interaction between pollutants and molecules used as templates for adsorption.[15,16] Specifically, electronic-structure calculations (either in the gas phase or in aqueous solution) as well as molecular dynamics simulations are among the most relevant methodologies that can be employed to understand the underlying adsorption processes and to help in the design of more efficient adsorbent materials.

A natural source of information regarding pesticides arises from the study of their interactions with the corresponding target proteins. Indeed, the way a pesticide acts toward a living organism is based on the interaction it can establish with a specific binding site of a target protein that regulates a vital function and, thus, acquiring detailed knowledge of these pesticide–protein binding processes will contribute to guide the functionalization with chemical species that may improve the adsorption properties of the materials. In particular, the mechanism of action of neonicotinoids, such as IMI (Fig. 1a), is based on their interaction with the nicotinic acetylcholine receptor (nAChR) in several insects, which leads to the disruption of nerve transmission. Due to structural and chemical similarities, *Aplysia californica* acetylcholine-binding protein (*Ac*-AChBP, Fig. 1b) has been consensually used as a functional

CQC-IMS, Department of Chemistry, University of Coimbra, 3004-535 Coimbra, Portugal. E-mail: qtmarque@ci.uc.pt

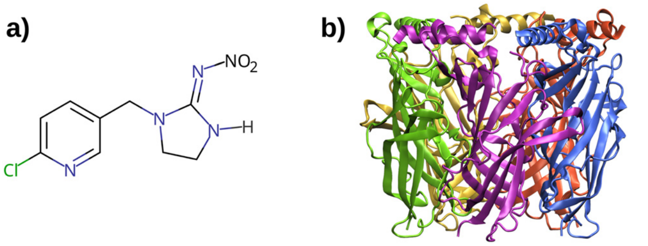

Fig. 1 Representation of the (a) imidacloprid molecule and (b) crystal structure of *Aplysia californica* acetylcholine-binding protein (*Ac*-AChBP, PDB ID: 3C79), with each chain of the pentamer displayed in a different color.

and structural surrogate to model the interaction between neonicotinoids and the extracellular ligand-binding domain of the insect nAChR receptor.[17] This surrogate is particularly suitable when membrane effects are not the main focus and it offers a significant reduction in computational cost compared to the simulation of the full membrane-embedded receptor. Moreover, high-resolution crystal structures of the full nAChR are extremely challenging to obtain, further justifying the use of *Ac*-AChBP in ligand-binding studies. Furthermore, *Ac*-AChBP appears to be very sensitive to both neonicotinoids and nicotinoids, while *Lymnaea stagnalis* acetylcholine binding protein (*Ls*-AChBP), another possible surrogate protein, is less sensitive to neonicotinoids.[18–20] The crystal structures of these IMI–AChBPs complexes (PDB IDs: 3C79 and 2ZJU) reveal some differences in the residues of the binding pocket of the two proteins, such as the greater amount of Tyr residues in *Ac*-AChBP as well as the presence of Ile106 and Val108 (that are missing in *Ls*-AChBP).

Ac-AChBP is a homopentamer, consisting of five identical subunits arranged symmetrically around a central axis (see Fig. 1b), with the agonist binding site located between two adjacent chains. Imidacloprid-AChBP binding has been subject to both experimental and theoretical studies.[19,21] Talley *et al.*[19] reported for the first time the high-resolution crystal structures of *Ac*-AChBP complexed with imidacloprid and thiacloprid (PDB IDs: 3C79 and 3C84), being the only neonicotidois with available crystallographic data that accurately describe their specific interactions with *Ac*-AChBP at the atomic level.

Several studies have explored molecular docking within the binding pocket of the IMI–AChBP complex, demonstrating a significant alignment between the predicted poses and the crystal structure.[22–25] Although MD simulations can offer detailed and dynamic insights into the interactions between proteins and small molecules, only a few elementary investigations using this technique can be found in the literature regarding the IMI–AChBP complexes.[17,26,27] In particular, a recent study used molecular docking and MD simulations for the selection of peptides for the recognition of IMI, starting from the structure of *Lymnaea stagnalis* acetylcholine-binding protein Q55R mutant receptor-imidacloprid complex. The results showed that the WQA34 peptide with 34 amino acids is a viable candidate for the selective recognition of imidacloprid (among other neonicotinoids), revealing a high potential for application in biosensors.[27] A few articles have also used the crystal structure of *Ac*-AChBP as a template to build several homology models of nAChRs, such as the cockroach Pameα7 nAChR, and further employed MD simulations,[17,28–30] but not directly investigate the interactions within the crystal structure of *Ac*-AChBP.

The interaction of IMI with target proteins has been also studied by performing electronic-structure calculations.[17,31–33] Selvam *et al.*[17] highlighted the role of the aromatic residues in the stabilization of complexes of imidacloprid or thiacloprid with nAChRs of cockroaches and honeybees. Wang *et al.*[31] investigated the interaction between models of key residues (Trp and Arg) and neonicotinoid-derivative analogues. At the MP2/6-311++G** level of theory, hydrogen bonding (H-bond) of neonicotinoids with Arg/Lys and π–π interactions with Trp were identified as crucial interactions. A two-layer ONIOM approach[32] used a *Ac*-AChBP binding pocket model, where the atomic positions of the polypeptide chain were kept frozen during optimization, and pinpointed Trp and Cys as key residues (CH-π, van der Waals, and H-bond interactions). A recent strategy,[33] employing local energy decomposition analysis at the DLPNO-CCSD(T)[34,35] level, has been developed and applied to IMI–nAChR binding, allowing the identification, quantification, and analysis of key ligand–residue interactions for molecular recognition in the active protein site.

Despite the great number of studies on the IMI–AChBP complex, it is relevant to establish a standard protocol that can be applied to any pesticide–protein system in order to assess the most important interactions and further incorporate this knowledge in the design of new adsorption materials. In fact, such an abundance of available information makes the IMI–AChBP complex a good system to test a new protocol. In this work, we propose a multi-stage computational approach to assess the interactions that contribute most to the stability of the pesticide–protein complexes (here applied to the IMI–AChBP complex), with the aim of selecting a relevant set of amino acids to be incorporated into pesticide-adsorbing materials. As a proof of concept, the selected amino acids were used to incorporate chitosan-based material models, which were then tested by MD simulations for their ability to promote IMI adsorption. In fact, chitosan has gained significant attention as an effective material for pesticide adsorption,[36] because it can be derived from chitin, which is an abundant and biodegradable natural resource.[37,38]

2. Systems and methodology

We employ a multi-stage computational methodology that uses freely available software to identify the amino acids from the *Ac*-AChBP protein that should be preferentially incorporated in a chitosan-based material to promote the adsorption of IMI. First, the crystal structure of the IMI–*Ac*-AChBP complex was downloaded from the Protein Data Bank (PDB ID: 3C79[39]); the structures of both IMI and *Ac*-AChBP are displayed in Fig. 1.

Calculations were performed for a reduced model of the *Ac*-AChBP protein to maintain an affordable computational burden. In this work, we considered an *Ac*-AChBP model composed by chains A and B of the protein (hereafter designated as AChBP_AB), which is the smallest model that describes the entire binding pocket.

The next stage of our strategy consists of applying blind and targeted molecular docking to identify promising configurations that are then tested by running long molecular dynamics (MD) simulations within the GROMACS framework.[40,41] The MD trajectories of the most stable configuration encountered for the complex are analyzed in detail to obtain the ligand interaction-network with the protein residues by using the ProLIF code.[42] In addition, the effective free energies were calculated with both MM-PBSA and MM-GBSA methods[43–45] by using the gmxMMPBSA program.[46,47] With these data in consideration, it is possible to choose the most relevant amino acids, whose interactions with the pesticide are then further studied at higher levels of theory by employing electronic-structure calculations.

Finally, we verified whether the functionalization of chitosan with such promising amino acids is relevant for promoting the IMI adsorption in this type of materials. The tests were carried out using long MD simulations for one IMI molecule and several models of chitosan-based systems in aqueous solution. Specifically, the adsorbing model systems employed in the simulations are formed by one or three monomers of chitosan that can be then functionalized in their $-NH_2$ groups, respectively, with one or three amino acids.

2.1. Molecular docking

In order to identify the most promising poses of imidacloprid in protein structure, docking was performed with Autodock Vina,[48,49] using its built-in scoring function. We conducted semi-flexible docking with the protein molecule treated as a rigid structure during the docking process. Before docking, chains A and B of the protein crystal structure were extracted from the PDB, and then hydrogens were added to the polar atoms. The analysis and visualization of the results, including binding affinity values, were performed, and the relevant structures were saved for later use in MD. Using the AChBP_AB protein model, three variations of docking were performed. Blind docking, where the entire protein was placed within a defined grid box with dimensions of 80 Å × 80 Å × 80 Å. Afterwards, we conducted a target docking with the grid box centered on the interface of the two chains, with the following dimensions 24 Å × 22 Å × 48 Å. After identifying the preferential binding region of the protein, we carried out a final docking with a grid box centered at the selected binding site (16 Å × 16 Å × 16 Å).

2.2. Molecular dynamics simulations

2.2.1. Topology construction and systems preparation. For the generation of the protein topology, the pdb2gmx tool of GROMACS 2019.6[40,41] was used, and hydrogens were added according to protonation states automatically determined by the pdb2gmx module. For both protein and ions, the Amber99SB-ILDN[50] force field was used. This is an updated version of the Amber99 force field[51] with Stony Brook modifications (for better protein backbone performance)[52] and additional side-chain torsion potential refinements, known as ILDN, for the amino acids isoleucine, leucine, aspartate, and asparagine. To construct the topologies of the IMI and chitosan-based models, we follow the standard procedure for the Amber force field using GAFF2 (general Amber force field) parameters.[53] GAFF2 extends Amber capabilities to model a wide range of organic molecules and is designed for compatibility across different Amber-based force fields, such as Amber99SB-ILDN. This force field has been extensively validated against experimental data for various organic molecules across multiple studies,[54,55] including protein–ligand simulations.[56,57] The construction of these topologies included the following steps: (i) molecules were optimized at the RHF/6-31G* level of theory, using GAMESS-US package;[58] (ii) using the RESP fitting protocol implemented in the R. E. D. package,[59] the partial charges were derived; (iii) finally, the ACPYPE script[60] was used to convert the geometry and generate the topologies to the GROMACS input format.

All molecular dynamics simulations were performed using GROMACS 2019.6,[40,41] which has support for NVIDIA GPUs, that are available in modern computing laboratories. The protein–ligand complex was placed in a cubic simulation box, ensuring a minimum distance of 10 Å between the complex and the box boundaries. Each system was then hydrated with TIP4P-EW[61] water molecules and 19 sodium ions were added to ensure overall charge neutrality of the system. The box dimensions were approximately 107 Å × 107 Å × 107 Å for the simulations with AChBP_AB.

Similar procedures were also applied in the preparation of the simulations with chitosan-based model systems, except that each trajectory now begins with the IMI well separated from the adsorbent molecule inside a box of dimensions 50 Å × 50 Å × 50 Å (see Fig. S1 of the SI).

2.2.2. Minimization, equilibration and production. After the preparation of the simulation box, an energy minimization step was carried out to relax the system and reduce molecular repulsions, followed by two consecutive equilibrium simulations of 10 ns each (NVT followed by NPT). The temperature was maintained at 298.18 K using the v-rescale[62,63] temperature coupling scheme. The pressure was maintained at 1 bar using a Parrinello–Rahman barostat[64] with a coupling constant of 0.2 ps and an isothermal compressibility of 4.5×10^{-5} bar^{-1}. The protein and ligand were coupled separately from water and ions, with a relaxation time constant of 0.1 ps, while chitosan models and IMI were coupled separately from water. The production phase of the MD simulation was run for 500 ns. A total of 3 (10) trajectories were run for MD simulations of protein–IMI complexes (chitosan-based model systems with IMI). The classical equations of motion were integrated using the leapfrog method, with a time step of 2 fs. Bond constraints during trajectory integration were employed through the LINCS (*i.e.*, linear constraint solver) algorithm.[65] Periodic boundary conditions were used in all simulations. The long-range electrostatic energy was evaluated using the particle mesh Ewald method,[66,67] with a 10 Å cut-off applied to both the Coulombic and van der Waals interactions.

2.2.3. Trajectory analysis. For the evaluation of stability of the protein–ligand complex and characterization of the inherent intermolecular interactions, we calculated several physical quantities with the software tools available in GROMACS,[40,41] such as root-mean-square deviation (RMSD) and number of hydrogen bonds over the simulation. We calculated the RMSD of the protein backbone for each system (see Fig. S2 of SI) and no significant conformational modifications were detected, either in the simulations of the protein alone or in the presence of the ligand. In the context of stability assessments of protein–ligand complexes, we have considered for each trajectory the RMSD of the ligand in relation to the protein backbone. The VMD package[68] was used for visual inspection of the trajectories and image generation. Schematic diagrams of the ligand interaction-network with protein residues during the simulations, and the respective frequency at which the interactions occur, were generated using the ProLIF code 2.0.3.[42] ProLIF generates interaction fingerprints by detecting and classifying non-covalent interactions based on predefined geometric rules from structural chemistry literature. These include specific distance and angle thresholds for each interaction type, such as hydrogen bonds, π–π stacking, cation–π, and hydrophobic contacts, applied across molecular dynamics trajectories. In this study, we used the default geometric criteria provided by ProLIF version 2.0.3 without any modification.[42]

To assess the ability of chitosan-based model systems to adsorb IMI, we calculated radial distribution functions (RDFs) using the default settings of GROMACS.[40,41] We also performed a clustering analysis on the chitosan trajectories to identify the most significant chitosan-IMI configurations observed over the entire simulation time. The occurrence of specific configurations was assessed using the default GROMACS "single linkage" method, with an RMSD cut-off of 0.15 nm (0.20 nm) for monomer-chitosan (trimer-chitosan) model simulations. These cut-off values were optimized through a trial-and-error approach to ensure that only a limited number of dominant structures emerged.

2.2.4. MM/PB(GB)SA methods. The calculation of binding free energy was carried out by utilizing the gmxMMPBSA program,[46,47] after removing the periodic boundary conditions of the MD trajectory obtained with GROMACS.[40,41] For this calculation, we have used both molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) and molecular mechanics generalized Born surface area (MM/GBSA) methods[43–45] that are available in the gmxMMPBSA program.[46,47]

The MM/PBSA and MM/GBSA methods are considered as end-point approaches for the calculation of binding free energy, since only the ending structures of a reaction are taken into account for estimating both the ligand binding free energy and relative free energy differences among distinct conformational states.[46,47] Accordingly, these methods are computationally affordable for calculating binding free energies at a reasonable level of accuracy. Because of this, MM/PB(GB)SA approaches have been used in the investigation of protein–ligand systems,[69,70] and are also applied in the present work to calculate average values of the binding free energy of the IMI–*Ac*-AChBP complex.

Basically, assuming that the bound and unbound states share similar conformational characteristics, the dynamics of both protein and pesticide may be derived from the same trajectory of the complex. The average binding free energy of the complex can be calculated by using the equation[47,71]

$$\Delta G_{\text{bind}} = \langle G_{\text{comp}} \rangle - \langle G_{\text{prot}} \rangle - \langle G_{\text{pest}} \rangle \quad (1)$$

where $\langle G_{\text{comp}} \rangle$ is the total free-energy of the complex, and the terms $\langle G_{\text{prot}} \rangle$ and $\langle G_{\text{pest}} \rangle$ refer, respectively, to the protein and pesticide, when isolated in water. We further note that the total free-energy terms are estimated, as an average over a specified time interval of the MD trajectory, by employing the following formula:

$$\langle G_x \rangle = \langle E_{\text{MM}} \rangle + \langle G_{\text{solv}} \rangle - \langle TS \rangle \quad (2)$$

where $\langle E_{\text{MM}} \rangle$ represents the average of the molecular mechanics potential energy in vacuum (*i.e.*, given by the corresponding force field), $\langle G_{\text{solv}} \rangle$ is the free energy of solvation, and $\langle TS \rangle$ corresponds to the contribution of entropy (S) to the free energy in vacuum at the simulation temperature (T); x represents the label for each free-energy term appearing in eqn (1). In the calculation of $\langle G_{\text{solv}} \rangle$, the MM/PBSA method employs the Poisson-Boltzman equation, while the MM/GBSA approach uses a generalized Born model (GB^{OBC2} was used in this study),[72] as implemented in the gmxMMPBSA program.[46,47] In turn, the entropic term in eqn (2) is not expected to be relevant when comparing relative binding free energies for similar systems[47] and, usually, it may be neglected (which happens in the present work). In this case, the second member of eqn (2) reduces to the first two terms and ΔG_{bind} is then designated as effective binding free energy.

We should also emphasize that, within the MM-GBSA framework, it is possible to decompose the binding free energy into the contributions of each residue of the protein. This allows us to identify the amino acids that contribute the most to the binding free energy of the complex, which is particularly relevant in the context of the present work. For this calculation, we have considered only amino acids that are within 6 Å from IMI.

2.3. Electronic-structure calculations

We have performed electronic-structure calculations to obtain further insight into the interaction of IMI with the amino acids revealed as the most important by the MD simulations starting from the IMI–AChBP_AB complex extracted from the crystal structure. All the electronic-structure calculations were carried out with the ORCA 5.0.4 package[73,74] and most of them employed the conductor-like polarizable continuum solvation model (CPCM),[75] using values of 80.4 and 1.33, respectively, for the dielectric constant and refractive index of water. For comparison purposes, single-point calculations in vacuum were also performed. The calculations were conducted according to the following procedure: (i) the geometry of the selected amino acids or complexes involving IMI was extracted from the last frame of the MD simulations. The structure of the amino acids must be completed as they are no longer involved in the peptide bond, which was done by adding a hydrogen atom to the amine group and a hydroxyl to the carboxylic acid group with the wxMacMolPlt program.[76] (ii) The geometries of the individual amino acids, IMI, and IMI complexes with one, two, and three amino acids were optimized at the HF-3c level of theory,[77] which we showed to be adequate for the present purpose (see Section 3.2); for the complexes, the optimizations were carried out by constraining the internal coordinates of each individual molecule, but the Cartesian coordinates of any atom were not restricted to their original values in the protein. (iii) The linear scaling domain-based local pair natural orbital CCSD(T) method, *i.e.*, DLPNO-CCSD(T),[34,35] was employed for the calculation of accurate energies, using the def2-TZVPP basis set[78] with matching def2/C auxiliary basis set;[79,80] this type of single-point calculation, both in vacuum and using the CPCM approach, is now available in ORCA.[81] (iv) Finally, the DLPNO-CCSD(T) energies were corrected for the basis set superposition error (BSSE) with the counterpoise method,[82] according to the following equation:

$$E_{\text{bind}} = E_{\text{AB}}^{\text{AB}}(\text{AB}) - E_{\text{A}}^{\text{AB}}(\text{AB}) - E_{\text{B}}^{\text{AB}}(\text{AB}) \tag{3}$$

where $E_{\text{X}}^{\text{Y}}(\text{Z})$ means the energy of the specie X calculated at the optimized geometry of fragment Z with the basis set of fragment Y, *i.e.*, $E_{\text{AB}}^{\text{AB}}(\text{AB})$ stands for the energy calculated to the AB species with the AB basis set at the geometry of the AB complex, and $E_{\text{A}}^{\text{AB}}(\text{AB})$ and $E_{\text{B}}^{\text{AB}}(\text{AB})$ correspond to the energy of the species A and B, respectively, with the AB basis set at the geometry of the complex AB.[82,83] Note that eqn (3) refers to the particular case where only two monomer molecules (A and B) were considered, but the correction is straightforward when more monomers are involved.

3. Results and discussion

We applied docking techniques to discover promising IMI–*Ac*-AChBP binding configurations. Although docking studies (and especially blind docking) are particularly relevant for discovering stable configurations of the pesticide–protein complex when little information is available in the literature, we should emphasize that they may still provide complementary details at the molecular level even for systems with resolved crystal structure (like IMI–*Ac*-AChBP). Indeed, the molecular-docking stage can be applied for two purposes, *i.e.*, blind docking is employed to localize the preferential binding region (usually designated as the binding pocket) for the specific protein model, while targeted docking allows discovery of the most promising pesticide configurations within the binding pocket.

The stability of the configurations obtained from the docking procedure have to be then evaluated by MD simulations. Once the most stable configurations are identified, we further look into the dynamical details of the corresponding pesticide–protein complex system. From the analysis of the MD trajectories with ProLIF, we identified and classified the interactions that, based on geometrical criteria, are established between atoms of the protein and pesticide. In addition, pesticide–protein binding energies are calculated by employing the MM/GBSA and MM/PBSA methods; the MM/GBSA method also estimates the contribution of each amino acid for the pesticide–protein binding energy, which allows identification of the most relevant residues on a more quantitative basis. Thus, in Section 3.1, we present detailed docking and MD investigations for the interaction of imidacloprid with the AChBP_AB protein model, seeking to complement the results already published in the literature.

In turn, Section 3.2 presents and discusses electronic-structure calculations on interactions between IMI and the five most relevant amino acids that were selected based on the MD simulations. The study of the interactions comprise structure optimization with the HF-3c method, followed by single-point calculation with the DLPNO-CCSD(T) method and a triple-zeta basis set.

As a proof of concept, the promising amino acids regarding the interactions with IMI (*cf.* the study in Section 3.2) were used to functionalize chitosan; then, MD simulations were performed to assess how IMI adsorption is enhanced with the new chitosan-based material models. The results are presented and discussed in Section 3.3.

3.1. Imidacloprid–protein complex

To ensure the robustness of our computational protocol, we first evaluated whether our docking and MD strategy could accurately reproduce the well-established crystal structure of the IMI–*Ac*-AchBP complex. This validation is particularly important, as it provides a solid foundation for extending the same procedure to other pesticide–protein systems that lack experimentally resolved structures.

The complexes that may be formed between AChBP_AB and IMI were initially investigated using a blind-docking procedure. The IMI–AChBP_AB complexes corresponding to the six best scoring docking poses were then studied by MD simulations in order to evaluate their stability. The results for RMSD as a function of time displayed in Fig. 2A indicate that most of the starting configurations are not stable. For instance, configurations where IMI is close to the α-helix of the protein (pose 4 and pose 5) tend to break down after a very short simulation time. In turn, pose 1, pose 2, and pose 6 also appear to be generally unstable, although some trajectories show RMSD values that remain approximately constant, on average, over a long time interval. As a general trend, the formation of a stable complex similar to the crystal one occurs whenever the IMI closely approaches the binding pocket. For example, in the case of t2 of pose 6, IMI escapes from its initial position (which leads to an increase in the RMSD values in the first 150 ns) and then forms a stable complex inside the binding pocket for the remaining part of the trajectory. In contrast, the third best-scoring pose (*i.e.*, pose 3) corresponds to an IMI–AChBP_AB complex that keeps small values of the RMSD for most of the simulation time; the exception occurs only in the final part of

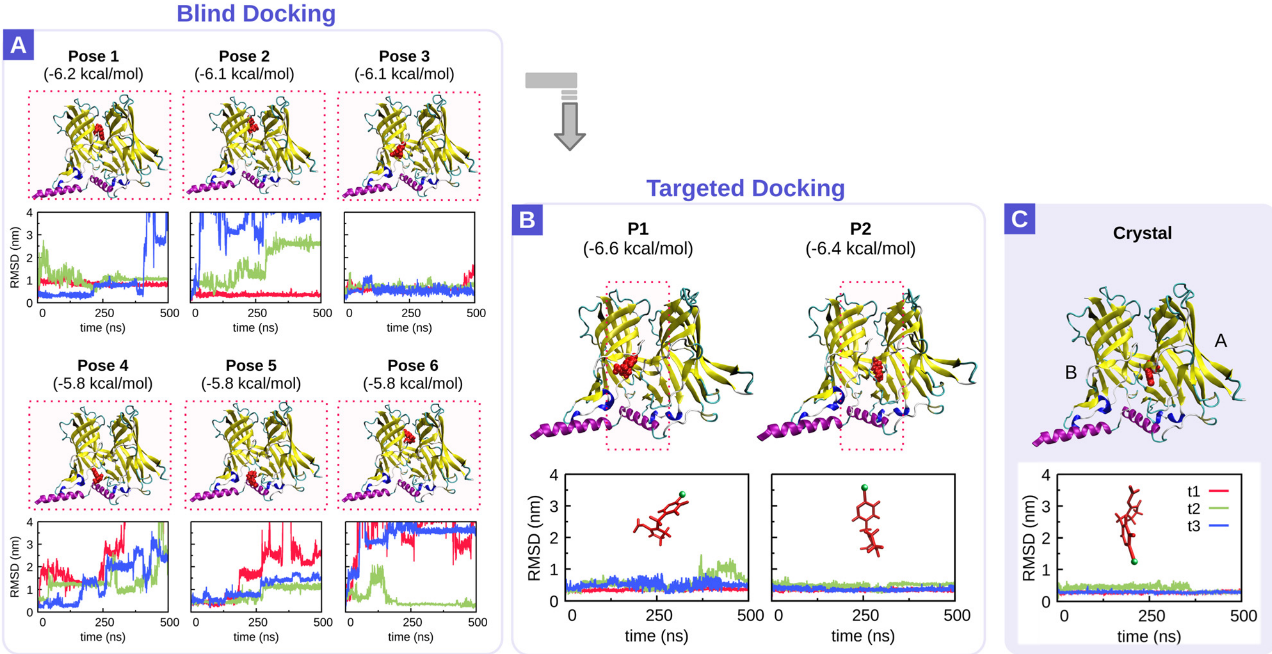

Fig. 2 Docking and MD results for the complexes of AChBP_AB with IMI (red). Panel A: Best scoring blind-docking poses (pose 1 to pose 6) used as starting structures for the MD simulations. Panel B: Best scoring targeted-docking poses (P1 and P2) used as starting structures for the MD simulations. Panel C: Starting structure extracted from the crystal (PDB ID: 3C79) used for the MD simulations. The scoring value of each pose is represented in parenthesis. The pink dashed lines correspond to schematic representations of the docking search-box (panels A and B). For each starting structure and three distinct trajectories (t1, t2, and t3), the respective RMSD of the IMI heavy-atoms in relation to the protein backbone is represented as a function of the simulation time (panels A–C). The IMI starting structures for the simulations of P1 and P2 (panel B) and crystal (panel C) are also displayed in red as inserts to the corresponding RMSD plots, with chlorine atom in green. For comparison purposes, the same scale is maintained in all RMSD plots.

t1, where the RMSD becomes larger. Actually, a complex of pose 3 seems to be essentially stable, even though IMI presents a certain degree of mobility within the binding pocket (as shown by the oscillations of RMSD).

Since these results point to an increase in the stability for complexes when IMI is located in the interface of chain A and chain B of AChBP_AB, the next step of the present approach consists of performing targeted docking within the dashed-line region represented over poses P1 and P2 in panel B. This means that targeted docking is somehow guided by the MD simulations. Actually, the main idea is to verify if the targeted docking within a small promising region is able to generate poses associated to more stable IMI–AChBP_AB complexes. Thus, we also ran three MD trajectories starting from the structure of the IMI–AChBP_AB complex corresponding to the two best-scoring poses, P1 and P2, obtained in the targeted-docking procedure (Fig. 2B). Although P1 (which is similar to pose 3 obtained in the blind docking) presents the best scoring parameter, we conclude from the MD simulations that the complex corresponding to P2 is the most stable one. Clearly, this emphasizes the need for always using MD simulations to validate docking results. It is also important to notice that moving from blind to targeted docking leads to approaching closer to the structure extracted from the crystal, which is represented in Fig. 2C. The MD simulations for P2 and crystal structures show similar trends regarding the stability of the complex. In both cases, the pesticide molecule appears to be confined in the same pocket of the protein during the whole simulation, even though IMI is oriented in a different manner in the initial configurations (as shown by the red IMI representations in the zoom insert of panel B and C of Fig. 2). Indeed, the chlorine atom is oriented upwards in both P1 and P2, while it is downwards in the crystal. Despite the above mentioned mobility of IMI within the binding pocket, we notice that a change in the chlorine orientation from upwards to downwards does not occur in the simulation (see also the snapshots in Fig. S3–S5 of the SI). Actually, a complete rotation of IMI leading to a downwards orientation of the chlorine atom was not observed within the binding pocket in any of the trajectories. In t3 of P1 (and similarly for t2 of P2), IMI attempts to perform this kind of rotation within the binding pocket, but apparently there is not enough room to be successful. Curiously, t2 of P1 ends up with the chlorine oriented downwards, but this was only achieved because IMI has previously exited the pocket and then entered again with the new orientation. These achievements on the rotation of the IMI associated with the dynamics of the binding pocket are illustrated in Fig. S6 and S7 of the SI.

In order to obtain some quantitative insight about the relative stability of the P1, P2 and crystal IMI–AChBP_AB complexes, we have employed both MM/GBSA and MM/PBSA to calculate the effective binding free energy, *i.e.*, ΔG; the calculated values and the associated errors are presented in Table 1. We note that the calculation of $\Delta G_{\mathrm{MM/GBSA}}$ and

Table 1 Effective binding free energy for the three trajectories (t1, t2, and t3) in the MD simulations starting from P1 and P2 of the targeted docking, and the crystal structure. This was estimated by employing the molecular mechanics Poisson–Boltzmann surface area (MM/PBSA) and molecular mechanics generalized Born surface area (MM/GBSA) methods. All the values were calculated over a time interval of 100 ns, which is shown for each MD trajectory in Fig. S3–S5

	P1	P2	Crystal
$\Delta G_{\mathrm{MM/GBSA}}$/kcal mol^{-1}			
t1	-23.79 ± 2.27	-32.25 ± 2.36	-34.24 ± 3.17
t2	-22.79 ± 2.18	-33.90 ± 2.53	-33.34 ± 2.65
t3	-22.46 ± 2.40	-28.63 ± 2.07	-34.06 ± 2.64
$\Delta G_{\mathrm{MM/PBSA}}$/kcal mol^{-1}			
t1	-18.41 ± 3.24	-22.24 ± 2.63	-25.44 ± 3.10
t2	-17.17 ± 2.91	-25.44 ± 3.15	-23.60 ± 2.94
t3	-17.55 ± 2.79	-18.08 ± 4.66	-26.62 ± 2.69

$\Delta G_{\mathrm{MM/PBSA}}$ was carried out over part of the trajectories where the dynamic equilibrium was assured (by the inspection of the RMSD plots shown in Fig. 2). Except for t2 of P1 which appears to be rather unstable in the last part and where the most stable complex occurs in the time interval between 225 ns and 335 ns, as a general procedure the last 100 ns of each trajectory were used for the calculation of the effective binding free energy; Fig. S3–S5 of the SI show the effective binding free energy as a function of time for the interval used in the calculations of both $\Delta G_{\mathrm{MM/GBSA}}$ and $\Delta G_{\mathrm{MM/PBSA}}$. It is apparent from Table 1 that the results obtained with the MM/GBSA approach are qualitatively similar to those of the MM/PBSA method, which is considered most accurate. For instance, Fig. S4 of the SI shows that, in contrast with MM/GBSA, the MM/PBSA method is able to account for the change of a dihedral angle, which leads to the internal reorientation of the imidazole ring of IMI in relation to the chloropyridine group and, consequently, to a decrease in the average values of the effective binding energy during the last 100 ns of t3. However, since the MM/PBSA method is more time consuming by a factor of $\sim$4.4, the results in Table 1 show that MM/GBSA is more advantageous than MM/PBSA for ranking the different complexes according to their relative stability. The values of the average binding free energy confirm that the complex from P2 of the targeted docking is more stable than the corresponding one from P1, while presenting similar stability to the complex from the crystal structure. However, as already mentioned, the relative orientation of IMI within the binding pocket in the case of P2 varies from one trajectory to the other (*cf.* snapshots of Fig. S4), which may explain some small differences among the corresponding average binding energies. By contrast, the orientation of IMI is essentially the same (despite sometimes involving the rotation of the imidazole and the chloropyridine rings) in the three trajectories of the crystal structure (see Fig. S5 of the SI). Thus, it appears that the crystal structure is dynamically more stable than the P2 complex. Given these results, we wonder whether it is possible to obtain a structure similar to the crystal one by docking. Actually, we manage to obtain this kind of structure by reducing the region of the target docking to essentially coincide with the binding pocket of the best docking pose P2 previously obtained (see Fig. S8 in the SI). This structure is very similar to the crystal one (*cf.*, Fig. S8) and it presents a lower docking score-value (-7.7 kcal mol^{-1}) in comparison to P2. Thus, it appears that for the present system a more realistic configuration of the complex is obtained when targeted docking is applied with a successive reduction of the searching region.

The amino acids of the protein that contribute the most to stabilizing the complex may be identified by performing the decomposition of the effective binding free energy for each residue using the MM/GBSA method. In Fig. 3, we show the energetic contribution of each residue for the value of $\Delta G_{\mathrm{MM/GBSA}}$ presented in Table 1 for each trajectory of the crystal structure. Only the residues with a contribution higher than 0.5 kcal mol^{-1} are represented. Analogous plots for the trajectories of P1 and P2 are shown, respectively, in Fig. S9 and S10 of the SI. Due to the similarities between P2 and the crystal structure discussed above, the residues that mostly contribute for $\Delta G_{\mathrm{MM/GBSA}}$ do not differ too much in both cases. Conversely, distinct amino acids arise in the case of P1 as the most relevant ones, which could be anticipated by the different binding position of IMI in the protein.

Additionally, it is apparent from Fig. 3 that all the values of the effective binding energies for the three trajectories agree with each other within the error bars. The main residues from chain A that contribute to the effective binding energy include

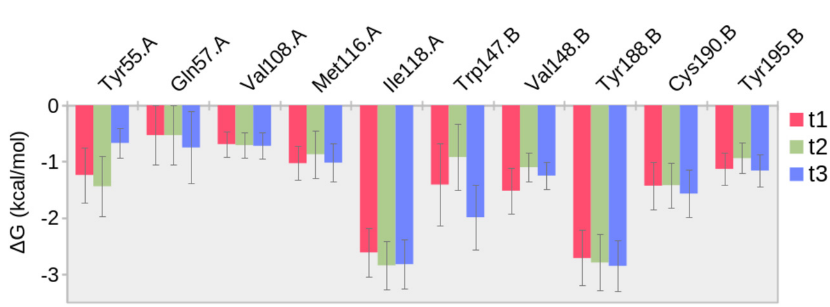

Fig. 3 Contribution of the residues to the IMI–AChBP_AB binding energy calculated by using the MM/GBSA method for the MD simulations starting from the crystal structure. Only the residues with an energy contribution higher than 0.5 kcal mol^{-1} are represented. Results from the three trajectories (t1, t2, and t3) are identified by different colors (red, green, and blue). Letters A and B label the protein chain (*e.g.*, Tyr55.A means residue Tyr55 of chain A, while Val148.B stands for residue Val148 of chain B).

Tyr55, Gln57, Val108, Met116, and Ile118. In turn, the main residues from chain B are Trp147, Val148, Tyr188, Cys190, and Tyr195. It is worth mentioning that residue Ile118 from chain A and Tyr188 from chain B present the most significant contributions to $\Delta G_{\mathrm{MM/GBSA}}$, with absolute values higher than 2 kcal mol^{-1}. On average, this accounts for $\sim$8% of the total effective binding energy of the complex, which is similar to the value ($\sim$7%) previously found for the residues Met114 and Tyr185 that are the most relevant ones for the binding complex of IMI with the *Ls*-AChBP protein.[26] Actually, Tyr is present in both protein models as a key residue for establishing the complex with IMI, but Met (which is also available in *Ac*-AChBP) appears to comprise an important role only in the case of the *Ls*-AChBP protein. We should note that Tyr, Ile and Met have a hydrophobic side chain, but the first (second) amino acid is classified as aromatic (aliphatic), while the third one contains a sulfur atom.

To complement the previous discussion, we performed an additional analysis using ProLIF, a computational tool that identifies and classifies interactions based on geometric patterns between groups of atoms in the protein and ligand. The main results for the analysis of the MD simulations starting from P1 and P2, and the crystal structure are summarized in Fig. 4. We represent in this figure the interaction network between IMI and the amino acids of AChBP_AB that are present for at least 30% of the simulation time. The average percentage of the simulation time that an interaction type is occurring is also shown at the bottom of the figure; more detailed plots that include all detected interactions as a function of the simulation time for the trajectories of P1, P2, and the crystal structure can be seen in Fig. S11–S19 of the SI. In all simulations, interactions classified as hydrophobic and van der Waals can be associated to the IMI–AChBP_AB complex during the whole time of the trajectories. In addition, for trajectories starting from P1 (P2), hydrogen bonding (π-stacking) is also present during 42% (48%) of the simulation time. In the case of P1, residues classified as either polar (Thr91, Gln105, Gln121) or aliphatic (Ile90, Val99, Leu102, Pro104) are the most relevant for the interaction with IMI. In turn, we may also observe in Fig. 4 that trajectories starting from either P2 or the crystal structure show similar interactions. In both cases, aromatic (Tyr188, Tyr93, Tyr195, Trp147, Tyr55), sulfur-containing (Cys190, Met116) and aliphatic (Ile118) residues are frequently in contact with IMI. However, two additional aliphatic residues (Val108 and Val148) appear to play an important role regarding the interaction with IMI for the simulation starting from the crystal structure. In summary, we may say that a hydrophobic environment, generally composed by both aromatic (which also allows for π-stacking) and aliphatic residues, around the pyridine ring of IMI as well as the presence of cysteine and glutamine nearby the nitro group appear to stabilize the complex in the binding pocket. Actually, both the P2 and the crystal structure tend to have IMI confined in the binding pocket, encased by loop C (formed by Gln186 up to Tyr195 residues of chain B), which allows optimal exposure of conjugated and hydrophobic regions of the ligand to aromatic side chains of the protein, as previously described by Talley *et al.*[19] Although the configuration of the crystal structure may allow the IMI to establish hydrogen bonds simultaneously with residues Cys190.B, Gln57.A and Tyr55.A, it is clear by the low values of the average number of H-bonds (*cf.* Table S1 of the SI) that such an event does not occur with high frequency during the simulation.

Conversely, π-stacking is not significant for MD trajectories with P1, being the hydrophobic environment guaranteed by aliphatic residues, while important hydrogen-bonding is now established between the Gln residue and nitrogen of the imidazole ring. Indeed, the relative relevance of hydrogen-bonding for P1 in comparison with P2 or the crystal structure has been generally confirmed by calculating the average number of H-bonds along each trajectory (see Table S1). However, since P1 is less stable than P2 (and also the crystal structure), it is expected that amino acids favoring π-stacking interactions are the most relevant for incorporating IMI-adsorbing materials.

3.2. Insights from electronic-structure calculations

Once the residues of AChBP that appear to be the most relevant for stabilizing the complex with IMI were identified in the MD simulations, further insight into the interaction between such amino acids and the pesticide is desirable for materials modeling purposes. We note that while previous studies[32,33] were focused on individual residue contributions for the AChBP–IMI complex (*i.e.*, keeping the entire structure of the binding pocket), we are mainly interested in exploring the geometric flexibility and optimal configurations of selected amino acids approaching IMI in water media, outside the protein environment, by applying electronic-structure methods.

The residues to be investigated through electronic-structure calculations were selected using the crystal structure as the reference conformation, as it is the most stable among the poses analyzed in Section 3.1. The selection of residues was guided by a combination of two key criteria: (i) top percentage of occupancy as obtained from the ProLIF analysis, and (ii) top per-residue contributions to the binding free energy as determined by MM/GBSA energy decomposition. Accordingly, the residues Ile118.A, Trp147.B, Val148.B, Tyr188.B, and Cys190.B were identified as key interaction partners with IMI; note that we maintain the numbering of the amino acids as they appear in the protein, in order to facilitate their identification and discussion. We performed electronic-structure calculations for the isolated amino acids and their complexes with IMI, which include all combinations of IMI with one, two, and three amino acids. Because Trp147.B and Val148.B are contiguous residues of the protein chain, they are treated as being bound in all complexes where both are present. In order to reduce the computational cost of the calculations, geometry optimization was carried out at the HF-3c level of theory for all systems. In an attempt to validate the application of the HF-3c approach to the present systems, we performed geometry optimization for the smallest complexes employing the Møller–Plesset perturbation theory using the resolution of identity (RI-MP2) method and the def2-TZVPP basis set[78] along with the def2-TZVPP/C auxiliary basis set.[79,80] For the optimized geometries with both HF-3c and RI-MP2, improved binding energy values were obtained by single-point calculations and BSSE correction

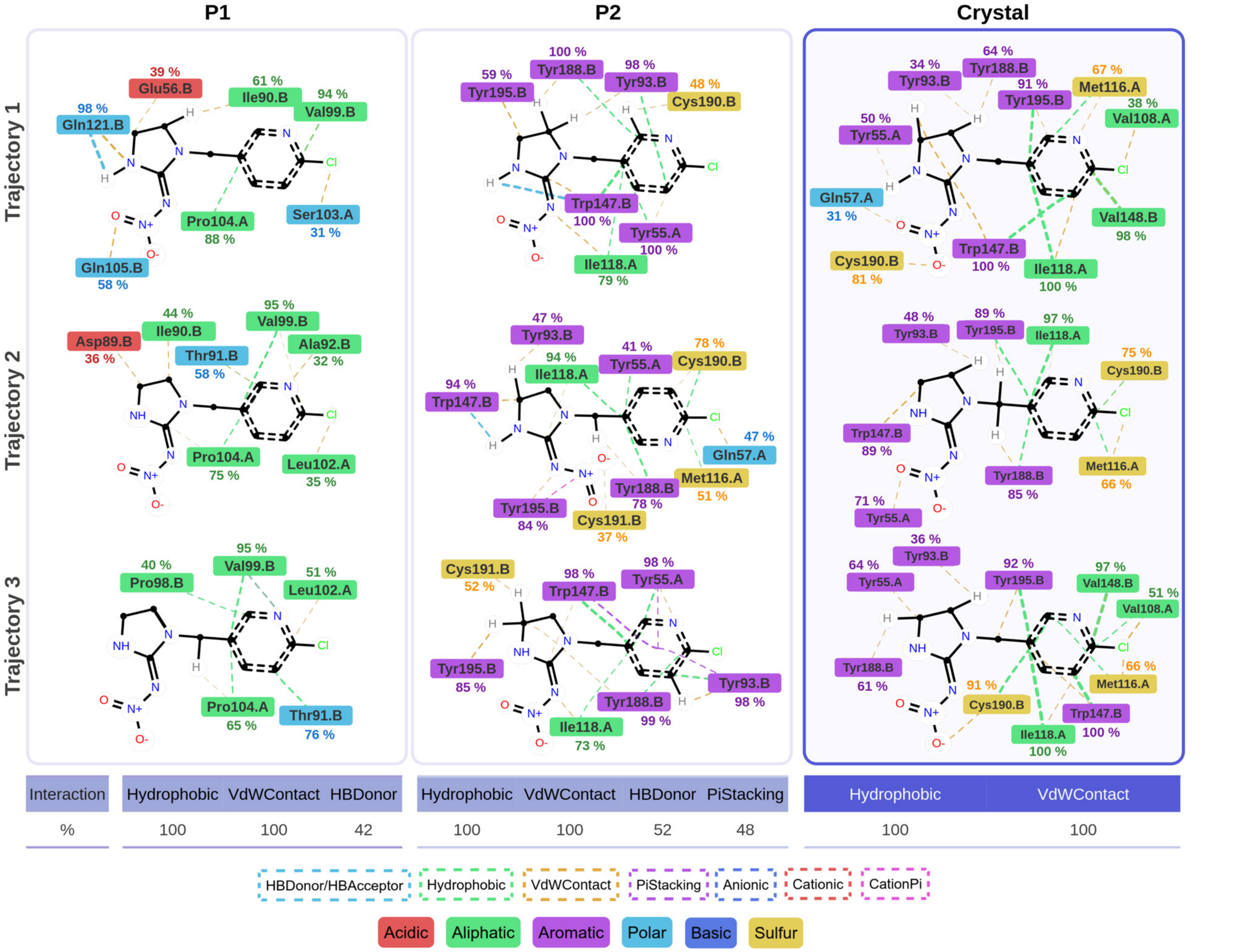

Fig. 4 Interaction-network of IMI with the residues of the AChBP_AB obtained with ProLIF for the MD simulations starting from P1, P2, and the crystal structure. The class of residues and type of interaction is identified by different colors. The percentage of the trajectory where the interaction is detected is also shown. Below the panels, we represent the percentage of the trajectory where each interaction type is present. Only the residues and interaction types present for at least 30% of the trajectory time are shown. Each value is an average over the three trajectories. Letters A and B in the name of residues indicate the chain to which they belong.

at the DLPNO-CCSD(T)/def2-TZVPP[78] (including the def2-TZVPP/C auxiliary basis set[79,80]) level of theory. All calculations were carried out using the conductor-like polarizable continuum model (CPCM).[81] A comparison of the two sets of results is shown in Table S2 of the SI. In general, the RMSD values are small, thus indicating that optimized structures obtained with the HF-3c approach do not differ significantly from the RI-MP2 ones. Furthermore, the order of the binding energies is essentially the same for both methods; the only exception arises for the complexes IMI–Trp147.B and IMI–Tyr188.B whose binding energies are close to each other. In turn, we have also verified that, at the HF-3c level of theory, the structures of the complexes obtained by the optimization with the individual molecules considered as rigid are not significantly different from those resulting from full optimization (see Fig. S20). Based on these results, we believe that the structures arising in the optimization of complexes with IMI and the selected amino acids at the HF-3c level of theory by constraining the internal coordinates of each individual molecule are reliable enough for the present purposes.

In Fig. 5–8, we represent the structures that were optimized with the HF-3c method by starting from the corresponding geometries in the IMI–AChBP_AB complex; the differences between the initial and optimized structures are shown in Fig. S21–S24 of the SI. The optimized geometries of the amino acids (Fig. 5) resemble their structures in the protein. In turn, the effect of the optimization can be better understood by looking at Fig. S21. It is apparent that the optimized IMI and amino acid structures are similar to the analogous ones in the protein, with the largest RMSD (calculated over C, N, and Cl atoms) of 0.70 Å for the Trp147.B–Val148.B peptide. In what concerns the complexes of one, two, or three amino acids with IMI (Fig. 6–8), the amino acids tend to closely approach IMI in the optimized structures. Some orientational rearrangements are also observed, especially in the case of complexes involving the Trp147.B amino acid (*e.g.*, complexes in panels a and f of

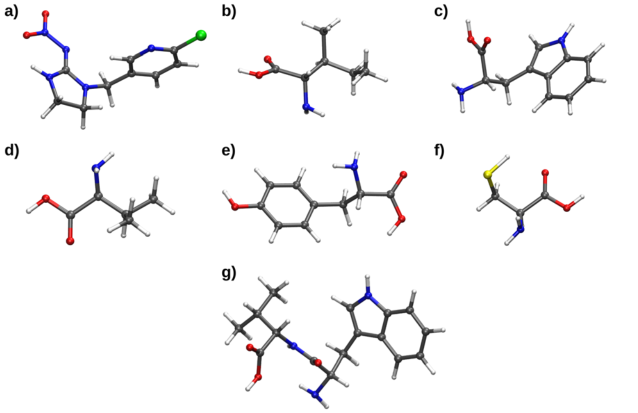

Fig. 5 Structures of imidacloprid and selected amino acids optimized at the HF-3c/CPCM-water level of theory: (a) IMI; (b) Ile118.A; (c) Trp147.B; (d) Val148.B; (e) Tyr188B; (f) Cys190.B; (g) Trp147.B–Val148.B.

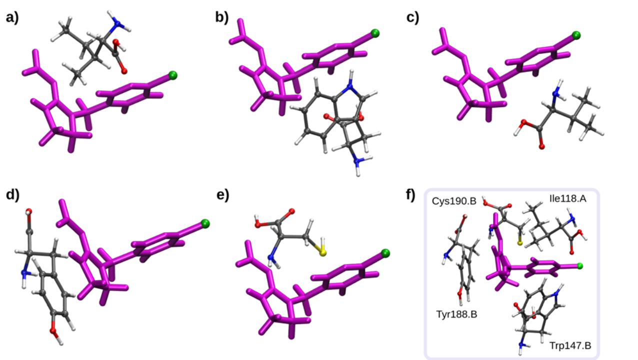

Fig. 6 Structures of the complexes of IMI with one amino acid that were optimized at the HF-3c/CPCM-water level of theory: (a) IMI + Ile118.A; (b) IMI + Trp147.B; (c) IMI + Val148.B; (d) IMI + Tyr188.A; (e) IMI + Cys190.B; (f) overlap of structures a, b, d, and e. For visualization purposes, IMI is represented in magenta and the chlorine atom is highlighted in green.

Fig. S23, and panel g of Fig. S24 of the SI). It is also interesting to observe in Fig. 6 that, in optimal configurations, Trp147.B and Val148.B occupy essentially the same regions around the IMI molecule, while the remaining residues stand in non-overlapping positions (*cf.* panel f of Fig. 6).

In Table 2, we represent the interaction energies for all the IMI–amino acids complexes shown in Fig. 6–8 that were calculated by employing the DLPNO-CCSD(T) method and the def2-TZVPP basis set[78] with the matching def2-TZVPP/C auxiliary basis set,[79,80] while the effect of water solvation was taken into account using CPCM.[81] For comparison purposes, analogous single-point DLPNO-CCSD(T) calculations were also conducted in vacuum (see Table S3 of the SI). As a general trend, the results of Table 2 and Table S3 are qualitatively equivalent, although the calculations in vacuum tend to give lower energy values than the corresponding CPCM ones. Accordingly, we only analyze the results from the CPCM calculations in the following discussion.

We may observe from Table 2 that, for complexes of one amino acid with IMI (panels a–e of Fig. 6), the weakest interaction occurs with the aliphatic residues Ile118.A and Val148.B (-4.13 kcal mol^{-1} and -3.96 kcal mol^{-1}). In turn, the sulfur-containing residue (Cys190.B) exhibits an intermediate binding energy, while the strongest interactions arise with the aromatic residues (Trp147.B and Tyr188.B) whose complexes with IMI show values of -7.91 kcal mol^{-1} and -7.16 kcal mol^{-1} for the corresponding binding energies. Comparison of these results with the contributions of the residues to the effective binding energy of IMI–AChBP_AB during MD simulations (see Fig. 3) confirms that Tyr188.B may have an important role as an adsorbent of IMI in water media. Conversely, Ile118.A does not show significant importance in binding to IMI, although it is revealed to be the residue that contributes the most to the effective binding energy of the IMI–AChBP_AB complex during the MD simulations. This discrepancy may be attributed to the confinement of IMI within the binding pocket, which ensures proximity to the Ile118.A residue during the MD simulation. As for Trp147.B, its movement is also restricted within the binding pocket during MD simulations, preventing the residue from adopting an optimal position regarding the interaction with IMI. By contrast, the HF-3c optimization leads Trp147.B to adjust its orientation, enabling a more favorable interaction with IMI (*cf.* the large value of the binding energy presented in Table 2).

The addition of a second amino acid contributes to strengthen the complex with IMI in different ways (see Table 2). The best combination of two amino acids is the pair formed by Trp147.B and Tyr188.B (*cf.* entry 7f in Table 2), closely followed by the Ile118.A/Trp147.B pair (entry 7a). We should emphasize that the first complex has two aromatic amino acids and, specifically, Trp147.B is also present in the second complex. This result appears to reinforce the general idea that the presence of aromatic amino acids contributes to the formation of stronger complexes with IMI. Nonetheless, such a simple assumption is not clear when looking at the other pairs of amino acids involving Tyr188.B (*cf.* entries 7c, 7h and 7j in Table 2) and especially in the case of the complex of IMI with the pair Trp147.B–Val148.B (entry 7e), which corresponds to the weakest interaction. We note that Trp147.B and Val148.B were kept linked through the peptide bond during the HF-3c optimization, which may have hindered the possibility of adopting the optimal positions of both amino acids in the complex with IMI. Indeed, Trp147.B and Val148.B tend to occupy the same region nearby IMI when individually optimized at the HF-3c level (see panels b and c of Fig. 6).

Furthermore, the presence of a third amino acid in the complex with IMI (Fig. 8 and entries 8a–8j in Table 2) leads to conclusions consistent with those observed above for complexes involving two residues. While in general aromatic residues tend to enhance the stability of the complex, the aliphatic residue Val148.B consistently contributes to weaker interactions with IMI, regardless of the amino acid combination. Once again, the combination Trp147.B–Val148.B is clearly unfavorable, except when the third amino acid is Tyr188.B (entry 8g). The synergistic effect of Trp147.B/Tyr188.B (already mentioned for the complex shown in Fig. 7f) is also apparent in the complexes involving these two amino acids and Ile118.A

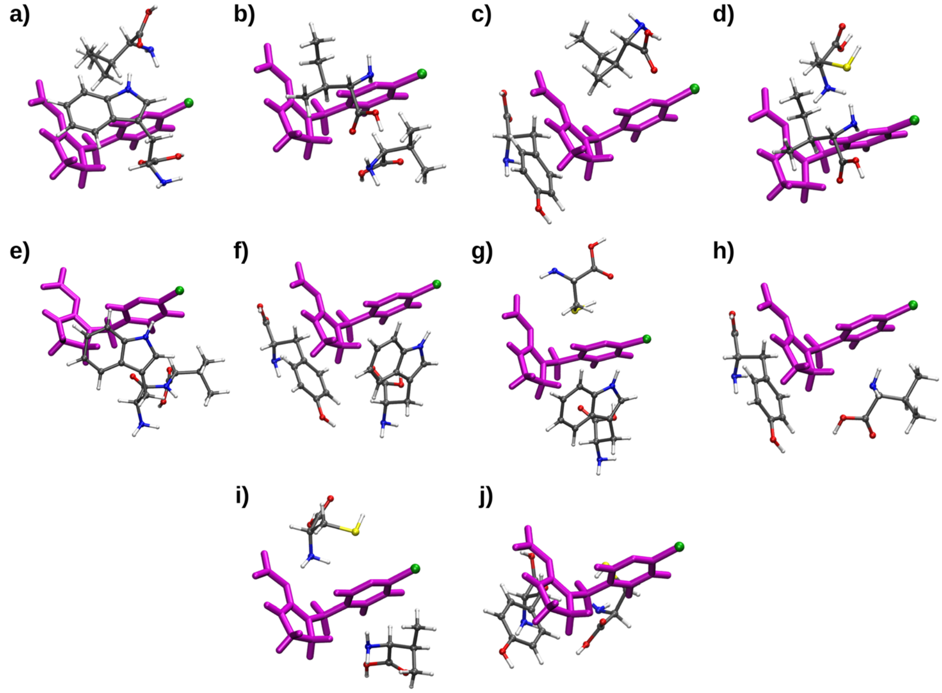

Fig. 7 Structures of the complexes of IMI with two amino acids that were optimized at the HF-3c/CPCM-water level of theory: (a) IMI + Ile118.A + Trp147.B; (b) IMI + Ile118.A + Val148.B; (c) IMI + Ile118.A + Tyr188.A; (d) IMI + Ile118.A + Cys190.B; (e) IMI + Trp147.B–Val148.B; (f) IMI + Trp147.B + Tyr188.B; (g) IMI + Trp147.B + Cys190.B; (h) IMI + Val148.B + Tyr188.B; (i) IMI + Val148.B + Cys190.B; (j) IMI + Tyr188.B + Cys190.B. For visualization purposes, IMI is represented in magenta and the chlorine atom is highlighted in green.

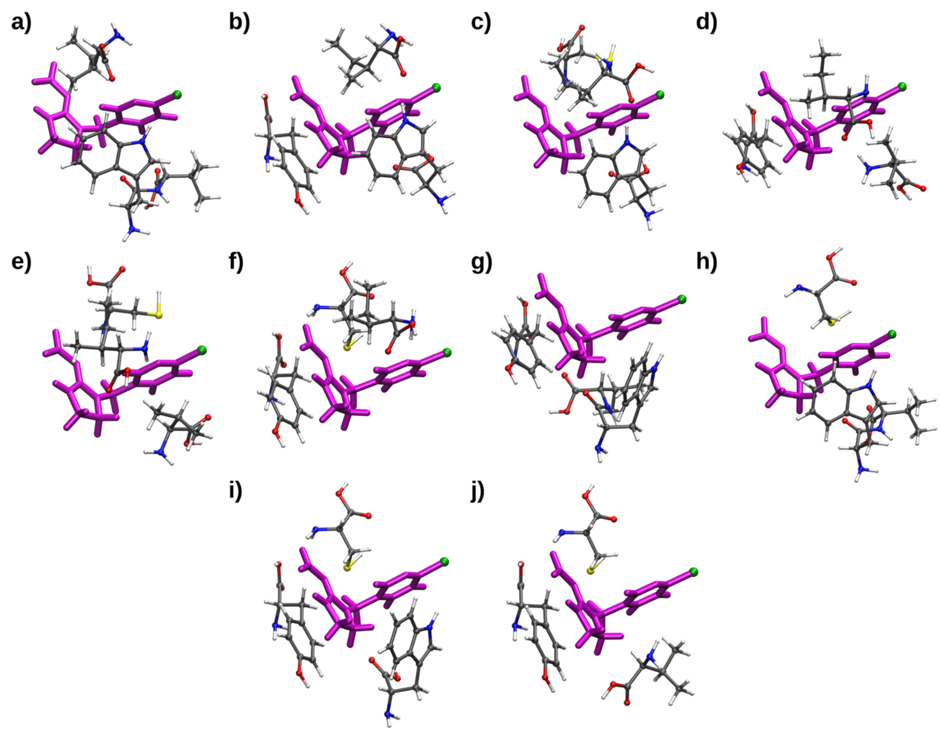

Fig. 8 Structures of the complexes of IMI with three amino acids that were optimized at the HF-3c/CPCM-water level of theory: (a) IMI + Ile118.A + Trp147.B–Val148.B; (b) IMI + Ile118.A + Trp147.B + Tyr188.B; (c) IMI + Ile118.A + Trp147.B + Cys190.B; (d) IMI + Ile118.A + Val148.B + Tyr188.B; (e) IMI + Ile118.A + Val148.B + Cys190.B; (f) IMI + Ile118.A + Tyr188.B + Cys190.B; (g) IMI + Trp147.B–Val148.B + Tyr188.B; (h) IMI + Trp147.B–Val148.B + Cys190.B; (i) IMI + Trp147.B + Tyr188.B + Cys190.B; (j) IMI + Val148.B + Tyr188.B + Cys190.B. For visualization purposes, IMI is represented in magenta and the chlorine atom is highlighted in green.

Table 2 Interaction energies, *i.e.*, ΔE (kcal mol^{-1}) for the studied IMI-amino acids systems. Calculated at the DLPNO-CCSD(T) def2-TZVPP basis with the auxiliar basis def2-TZVPP/C level of theory in water (CPCM), with BSSE counterpoise correction

Panel	System	Energy (kcal mol^{-1})
IMI + 1 amino acid		
6a	Ile118.A	−4.13
6b	Trp147.B	−7.91
6c	Val148.B	−3.96
6d	Tyr188.B	−7.16
6e	Cys190.B	−5.28
IMI + 2 amino acids		
7a	Ile118.A; Trp147.B	−15.18
7b	Ile118.A; Val148.B	−13.37
7c	Ile118.A; Tyr188.B	−11.57
7d	Ile118.A; Cys190.B	−13.64
7e	Trp147.B–Val148.B	−9.20
7f	Trp147.B; Tyr188.B	−16.38
7g	Trp147.B; Cys190.B	−12.93
7h	Val148.B; Tyr188.B	−11.04
7i	Val148.B; Cys190.B	−10.52
7j	Tyr188.B; Cys190.B	−13.10
IMI + 3 amino acids		
8a	Ile118.A; Trp147.B–Val148.B	−15.35
8b	Ile118.A; Trp147.B; Tyr188.B	−22.43
8c	Ile118.A; Trp147.B; Cys190.B	−24.88
8d	Ile118.A; Val148.B; Tyr188.B	−18.27
8e	Ile118.A; Val148.B; Cys190.B	−17.79
8f	Ile118.A; Tyr188.B; Cys190.B	−20.85
8g	Trp147.B–Val148.B; Tyr188.B	−19.62
8h	Trp147.B–Val148.B; Cys190.B	−14.67
8i	Trp147.B; Tyr188.B; Cys190.B	−23.10
8j	Val148.B; Tyr188.B; Cys190.B	−18.21

(entry 8b in Table 2) or Cys190.B (entry 8i). Nonetheless, the most favorable complex with IMI involves the amino acids Trp147.B, Ile118.A, and Cys190.B. It is worth noting that, as mentioned above (Fig. 6f), Trp147.B, Tyr188.B, Ile118.A, and Cys190.B create optimal confinement of IMI by wrapping it around and, hence, allowing different groups of the pesticide to contribute to the binding energy. During the optimization process, the three amino acids can move to positions slightly different from those of the corresponding single-residue complexes (Fig. 6 and 8) so that amino acid–amino acid interactions might be promoted and, hence, the synergistic effect will even be improved.

3.3. Imidacloprid adsorption on amino acid-modified chitosan

We investigated how the most promising amino acids incorporated in chitosan-based material models may contribute to promote the adsorption of IMI. From the insights of the electronic structure calculations, we have selected Trp, Tyr, Ile and Cys as promising amino acids.

Independent MD simulations were performed to study the adsorption of IMI by various chitosan-based model systems (*cf.* Fig. S25 of the SI). We have considered monomers of chitosan (CTS1) functionalized with each of the promising amino acids that, henceforward, will be designated as CTS1_Trp1, CTS1_Tyr1, CTS1_Ile1 and CTS1_Cys1, respectively. We note that chitosan functionalization is achieved by covalently linking the carboxyl (–COOH) group of the amino acids to the amine (–NH_2) group of chitosan, leading to the formation of an amide bond (*cf.* peptide bond).

The bottom panel of Fig. 9 displays the average RDFs for 10 independent trajectories of each of the systems mentioned above, with the corresponding standard deviation given by the shaded area; the corresponding plots for the distance between the IMI molecule and the chitosan-based monomers are shown in Fig. S26–S30 of SI. It is apparent from Fig. 9 that the probability of having the IMI near the non-functionalized monomer of chitosan is very small. This is in agreement with the limited effectiveness of chitosan in adsorbing imidacloprid reported in previous works.[36,84] In contrast, we observe that chitosan functionalized with amino acids consistently leads to a noticeable increase in the RDF peak located at approximately 0.5 nm, indicating greater efficiency of these material models to capture IMI. Specifically, the introduction of the amino acids Tyr and Trp leads to a significant enhancement in the strength of the interaction with IMI, resulting in an increase in the RDF

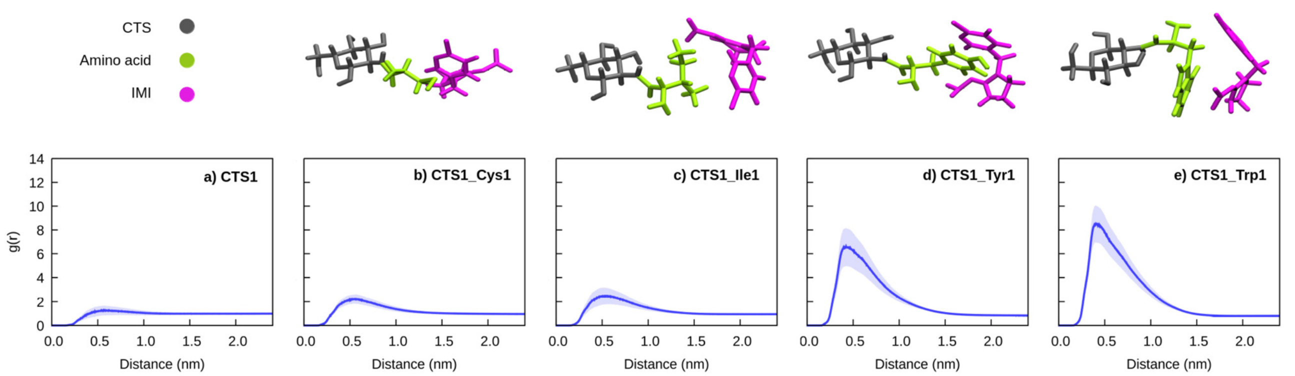

Fig. 9 (bottom panels) Average RDFs between IMI and chitosan-based monomer obtained from 10 trajectories per system: (a) CTS1; (b) CTS1_Cys1; (c) CTS1_Ile1; (d) CTS1_Tyr1; (e) CTS1_Trp1. The shaded area represents the standard deviation. (upper panels) Main structures arising from the clustering analysis of the complexes formed in the MD simulations of the above systems, with a RMSD cut-off of 0.15 nm. The average frequencies of the represented structures are about 2%, 3%, 13%, and 16% for simulations of panels (b) to (e), respectively. All simulations include one molecule of IMI and one chitosan-based monomer solvated with water.

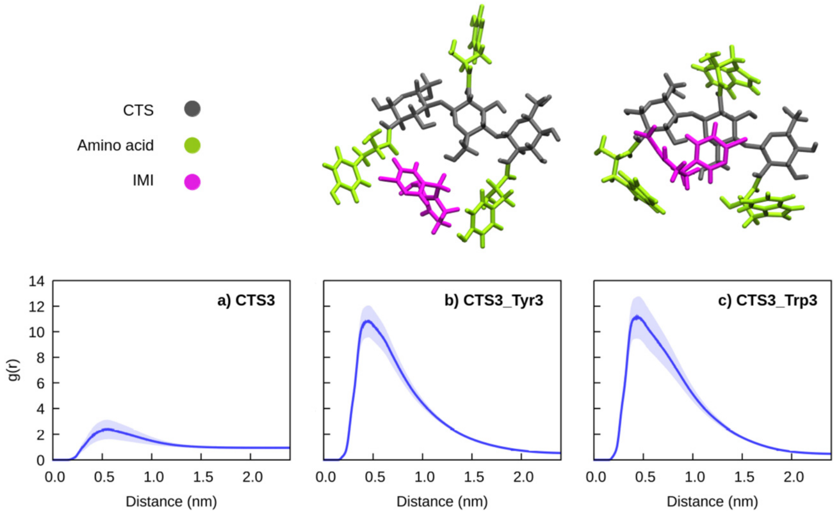

Fig. 10 (bottom panels) Average RDFs between IMI and chitosan-based trimer obtained from 10 trajectories per system: (a) CTS3; (b) CTS3_Tyr3; (c) CTS3_Trp3. The shaded area represents the standard deviation. (upper panels) Main structures arising from the clustering analysis of the complexes formed in the MD simulations of the above systems, with a RMSD cut-off of 0.20 nm. The average frequencies of the represented structures are about 46%, and 53% for simulations of panels (b) and (c), respectively. All simulations include one molecule of IMI and one chitosan-based trimer solvated with water.

peak by a factor of 5 and 7, respectively, compared to non-functionalized chitosan. The amino acids Cys and Ile are not as efficient in capturing IMI, showing a smaller increase of the RDF peak (by a factor of $\sim$2) in comparison with non-functionalized chitosan. This behavior was already anticipated from the weaker interaction energies of the IMI-Cys and IMI-Ile complexes compared to the IMI-Trp and IMI-Tyr systems, as revealed by the electronic structure calculations presented in Section 3.2.

In the upper panels of Fig. 9, we represent illustrative examples of the main structures obtained from the clustering analysis of the simulations, within a 0.15 nm RMSD cut-off. The average probability of each cluster observed during the simulations was 2%, 3%, 13% and 16% for CTS1_Cys1, CTS1_Ile1, CTS1_Tyr1 and CTS1_Trp1, respectively. The most stable structures are those with either CTS1_Tyr1 or CTS1_Trp1, which exhibit interactions primarily between the aromatic rings of the amino acids with either the imidazole group or the chloropyridine ring of IMI (see the upper panel of Fig. 9). Conversely, in the simulations with the CTS1_Ile1 system, IMI exhibits weak interactions with the hydrophobic groups of Ile. In turn, in the rare cases where CTS1_Cys1 captures IMI, the interaction mainly involves the sulfur atom of Cys and the six-membered ring of IMI (see the upper panel of Fig. 9). Then, amino acids containing aromatic groups appear to be the most relevant to incorporate in chitosan-based materials for removing IMI from water.

We performed additional MD simulations to assess the effect on IMI adsorption of using multiple amino acids in chitosan. For that, we used a trimeric chitosan model functionalized with three residues of tyrosine (CTS3_Tyr3) or three residues of tryptophan (CTS3_Trp3), as shown in Fig. S25 of the SI. We chose these amino acids to functionalize chitosan, because they yielded the best results with respect to the adsorption of IMI. For comparison purposes, a simulation with a non-functionalized chitosan trimer (CTS3) was also carried out. The analysis of the RDF plots of Fig. 10 and the distance plots (see Fig. S30–S32 of SI) reveals that the functionalization of the chitosan trimer with three amino acids (CTS3_Tyr3 or CTS3_Trp3) contributes to a further increase in IMI adsorption. In fact, the maximum values of the RDFs for CTS3_Tyr3 and CTS3_Trp3 surpass those for CTS1_Tyr1 and CTS1_Trp1, respectively. Now, the RDF peaks increase by a factor of approximately five compared to CTS3. The main structures arising from the clustering analysis for simulations with CTS3_Tyr3 or CTS3_Trp3 (within a RMSD cut-off of 0.20 nm) are shown in the upper panels of Fig. 10. These structures occur in simulations with an average probability of about 46% and 53%, respectively, and suggest that IMI is captured mainly due to the synergistic contribution of two amino acids, which capture the pesticide molecule between them (see the upper panel of Fig. 10). In summary, the presence of Trp and Tyr clearly contributes to the enhancement of the adsorption efficiency of IMI, which is sufficiently convincing to say that they should be used in future chemical modifications of materials such as chitosan.

4. Conclusions

In the present study, we devise a computational strategy that can be generally employed to unveil relevant interactions between pesticides and amino acids from a typical target protein, which could form the basis for the design of new adsorbing materials for water remediation. The proposed approach begins with the selection of the protein system that is likely to form a stable complex with the pesticide. In doing this, we aim to adopt a minimal protein model that allows us to maintain an affordable computational cost, while being sufficiently realistic to guarantee the dynamical stability of the pesticide in the binding pocket. Ultimately, the reduction strategy adopted for the protein should lead us to identify the amino acids that contribute the most to binding to the pesticide. To pursue such a goal, three consecutive main stages, corresponding to different levels of theory, are covered: (i) molecular docking, (ii) molecular dynamics simulations, and (iii) electronic-structure calculations. As a case study, we have applied this methodology to investigate the interaction between imidacloprid (IMI) and the typical target *Aplysia californica* Acetylcholine Binding Protein (*Ac*-AChBP). However, this computational strategy is general and, hence, it can be applied in studies involving other proteins and pesticides.

The blind-docking procedure is particularly necessary when information about the crystal structure of the protein–pesticide complex is not available. Conversely, whenever a crystal structure is available (as in the case of the IMI–*Ac*-AChBP complex), computational time may be saved by just applying targeted docking within the binding-pocket region and then probing the stability of the highest-scoring configurations upon MD simulations. The configurations so obtained led actually to the most stable protein–pesticide complexes, alongside the structure taken from the crystal.

Based on the ProLIF and MM/GBSA analysis carried out over MD simulations of the IMI–*Ac*-AChBP crystal structure, it appears that Ile118.A, Tyr188.B, Trp147.B, Val148.B, and Cys190.B are the amino acids that assume a major role in the stabilization of the complex. Then, we studied the interactions of these relevant amino acids with IMI by employing electronic-structure calculations and concluded that tryptophan shows the greatest binding affinity to IMI, followed by tyrosine, isoleucine, and cysteine.

Moreover, the simultaneous presence of some of such amino acids tends to produce a synergistic stabilization of the complex, which may be due to an optimal confinement of IMI. Conversely, the presence of valine leads to an opposite effect, since it preferentially occupies the same region near the IMI as that of tryptophan.

In a final step, we use the previous knowledge to build chitosan models functionalized with the amino acids that present the strongest interaction with IMI. These chitosan-based models were tested for the adsorption of IMI employing MD simulations. The calculations showed that non-functionalized chitosan exhibits negligible interaction with IMI, while incorporation of the aromatic amino acids Tyr and Trp led to a substantial increase of the IMI adsorption. Furthermore, MD simulations for trimeric chitosan models functionalized with multiple units of Tyr or Trp present a synergistic effect, which promotes IMI adsorption. The present findings suggest that the functionalization of chitosan with specific amino acids can effectively improve its performance in the adsorption of pesticides, and that the proposed computational pipeline may serve as a robust strategy to guide the design of bio-based materials for environmental remediation applications. In future work, we will explore the heterogeneous functionalization of chitosan and other materials with different amino acids, aiming at investigating potential synergistic effects on the efficiency of neonicotinoid adsorption.

Author contributions

JRCS conducted the research, performed all calculations and analyses, and prepared all tables and figures. PEA and JMCM supervised the research and edited the final version of the manuscript. JRCS, PEA, and JMCM wrote, reviewed, and edited the manuscript. All authors have edited and given their approval to the final version of the manuscript.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting the findings of this study are available in the SI of this article. All data is available without infringements or restrictions at **https://github.com/comput-chem-uc/Imidacloprid_Ac-AChBP_Chitosan.git**. The GitHub repository includes a file detailing its content, which includes all configurations, topologies, parameters, and input files required to reproduce the MD simulations.

(i) Illustration of the simulation box; (ii) RMSD plots of the protein backbone as a function of the simulation time for all MD simulations; (iii) plots of the effective binding energy as a function of time for the results of Table 1; (iv) binding pocket dynamics for P1 and P2: IMI rotation; (v) binding site docking pose; (vi) plots of the energy contribution of the residues for the binding energy of P1 and P2; (vii) IMI–AChBP_AB interaction-network over the simulation obtained with ProLIF and Table S1 with the average number of hydrogen bonds; (viii) validation of the HF-3c method with RI-MP2 calculations; (ix) full *vs.* partial optimized structures of pesticide–amino acid complexes; (x) overlap of initial and optimized structures of the amino acids; (xi) plots of the distance between IMI and the chitosan-based molecules as a function of time; (xii) GROMACS input files required to reproduce the MD calculations can be downloaded from **https://github.com/comput-chem-uc/Imidaclorid_Ac-AChBP_Chitosan.git**. See DOI: **https://doi.org/10.1039/d5cp01445j**

Acknowledgements

The authors thank the support from the Coimbra Chemistry Centre (CQC–IMS), which is financed by the Portuguese ''Fundação para a Ciência e a Tecnologia'' (FCT) through the Project UIDB/00313/2020 (DOI: 10.54499/UIDB/00313/2020), co-funded by COMPETE2020-UE. The authors are also grateful for the provision of computational time in the supercomputer resources hosted at Laboratório de Computação Avançada (reference: 2022.15842.CPCA) and INCD Cirrus HPC cluster (reference: 2023.10584.CPCA.A2) which was financed by FCT. J.R.C. Santos also acknowledges FCT for the PhD fellowship (reference: 2022.12451.BD; DOI: 10.54499/2022.15842.CPCA.A1). This publication is also based upon the work of COST Action CA21101 "Confined molecular systems: from a new generation of materials to the stars" (COSY) supported by COST (European Cooperation in Science and Technology).

Notes and references

1 *World Food and Agriculture – Statistical Yearbook 2023*, FAO, 2023.
2 *World Food and Agriculture – Statistical Yearbook 2021*, FAO, 2021.
3 G. Lofrano, G. Libralato, S. Meric, V. Vaiano, O. Sacco, V. Venditto, M. Guida and M. Carotenuto, in *Visible Light Active Structured Photocatalysts for the Removal of Emerging Contaminants*, ed. O. Sacco and V. Vaiano, Elsevier, 2020, pp. 1–25.
4 F. Carvalho, *Food Energy Secur.*, 2017, **6**, 48–60.
5 T. Parween and S. Jan, *Pesticide consumption and threats to biodiversity*, Elsevier, Amsterdam, The Netherlands, 1st edn, 2019, pp. 39–73.
6 R. W. Coppock and M. M. Dziwenka, Threats to wildlife by chemical and warfare agents, *Handbook of Toxicology of Chemical Warfare Agents*, Academic Press, Boston, 2020, ch. 63, pp. 1077–1087.
7 S. Kagabu, *J. Agric. Food Chem.*, 2011, **59**, 2887–2896.
8 S. Pang, Z. Lin, Y. Zhang, W. Zhang, N. Alansary, S. Mishra, P. Bhatt and S. Chen, *Toxics*, 2020, **8**, 65.
9 W. Han, Y. Tian and X. Shen, *Chemosphere*, 2018, **192**, 59–65.
10 J.-G. Zhang, W. Shi, D.-D. Ma, Z.-J. Lu, S.-Y. Li, X.-B. Long and G.-G. Ying, *Environ. Sci. Tech.*, 2023, **57**, 13384–13396.
11 S. R. Khalil, A. Awad, H. H. Mohammed and M. A. Nassan, *Environ. Toxicol. Pharmacol.*, 2017, **55**, 165–174.
12 C.-H. Wu, C.-L. Lin, S.-E. Wang and C.-W. Lu, *Pestic. Biochem. Physiol.*, 2020, **163**, 94–101.
13 M. Ceglowski, Y. W. Marien, S. Smeets, L. De Smet, D. R. Dhooge, G. Schroeder and R. Hoogenboom, *Chem. Mater.*, 2022, **34**, 84–96.
14 A. Valverde, G. I. Tovar, N. A. Rio-López, D. Torres, M. Rosales, S. Wuttke, A. Fidalgo-Marijuan, J. M. Porro, M. Jiménez-Ruiz, V. García Sakai, A. García, J. M. Laza, J. L. Vilas-Vilela, L. Lezama, M. I. Arriortua, G. J. Copello and R. Fernández de Luis, *Chem. Mater.*, 2022, **34**, 9666–9684.
15 A. Malloum, K. A. Adegoke, J. O. Ighalo, J. Conradie, C. R. Ohoro, J. F. Amaku, K. O. Oyedotun, N. W. Maxakato, K. G. Akpomie, E. S. Okeke and C. Olisah, *J. Mol. Liq.*, 2023, **390**, 123008.
16 P. A. Townsend, *ACS Omega*, 2024, **9**, 5142–5156.
17 B. Selvam, J. Graton, A. D. Laurent, Z. Alamiddine, M. Mathé-Allainmat, J. Lebreton, O. Coqueret, C. Olivier, S. H. Thany and J.-Y. Le Questel, *J. Mol. Graph. Model.*, 2015, **55**, 1–12.
18 M. Tomizawa, D. Maltby, T. T. Talley, K. A. Durkin, K. F. Medzihradszky, A. L. Burlingame, P. Taylor and J. E. Casida, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 1728–1732.
19 T. T. Talley, M. Harel, R. E. Hibbs, Z. Radic, M. Tomizawa, J. E. Casida and P. Taylor, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 7606–7611.
20 P. Rucktooa, A. B. Smit and T. K. Sixma, *Biochem. Pharmacol.*, 2009, **78**, 777–787.
21 M. Tomizawa, T. T. Talley, D. Maltby, K. A. Durkin, K. F. Medzihradszky, A. L. Burlingame, P. Taylor and J. E. Casida, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 9075–9080.
22 T. Sander, A. T. Bruun and T. Balle, *J. Mol. Graph. Model.*, 2010, **29**, 415–424.
23 A. Taly, C. Colas, T. Malliavin, A. Blondel, M. Nilges, P.-J. Corringer and D. Joseph, *J. Mol. Graph. Model.*, 2011, **30**, 100–109.
24 Q. Ma, H.-S. Tae, G. Wu, T. Jiang and R. Yu, *J. Chem. Inf. Model.*, 2017, **57**, 1947–1956.
25 A. Babakhani, T. T. Talley, P. Taylor and J. McCammon, *Comput. Biol. Chem.*, 2009, **33**, 160–170.
26 J. Tian, Q. Zhang, X. An, H. Liu, Y. Liu and H. Liu, *Mol. Inform*, 2019, **38**, e1800125.
27 S. Aldulaijan, *PLoS ONE*, 2023, **18**, e0295619.
28 J. Shuai, X. Wang, G. Li, Y. Kong, W. Li, Z. Li and J. Cheng, *J. Mol. Graphics Modell.*, 2022, **114**, 108177.
29 Q. Li, X. Kong, Z. Xiao, L. Zhang, F. Wang, H. Zhang, Y. Li and Y. Wang, *J. Mol. Model.*, 2012, **18**, 2279–2289.
30 A. Cartereau, E. Taillebois, J.-Y. Le Questel and S. H. Thany, *Int. J. Mol. Sci.*, 2021, **22**, 9880.
31 Y. Wang, J. Cheng, X. Qian and Z. Li, *Bioorg. Med. Chem.*, 2007, **15**, 2624–2630.
32 J. P. Cerón-Carrasco, D. Jacquemin, J. Graton, S. Thany and J.-Y. Le Questel, *J. Phys. Chem. B*, 2013, **117**, 3944–3953.
33 M. E. Beck, C. Riplinger, F. Neese and G. Bistoni, *J. Comput. Chem.*, 2021, **42**, 293–302.
34 C. Riplinger, B. Sandhoefer, A. Hansen and F. Neese, *J. Chem. Phys.*, 2013, **139**, 134101.
35 C. Riplinger, P. Pinski, U. Becker, E. F. Valeev and F. Neese, *J. Chem. Phys.*, 2016, **144**, 024109.
36 F. G. A. Estrada, J. M. C. Marques and A. J. M. Valente, *ChemistryOpen*, 2019, **8**, 438–446.
37 A. Harugade, A. P. Sherje and A. Pethe, *React. Funct. Polym.*, 2023, **191**, 105634.
38 C. Liu, G. Crini, E. Lichtfouse, L. D. Wilson, L. A. Picos-Corrales, P. Balasubramanian and F. Li, *J. Water Process Eng.*, 2025, **71**, 107327.
39 R. P. D. Bank, *RCSB PDB - 3C79: Crystal structure of Aplysia californica AChBP in complex with the neonicotinoid imidacloprid*, https://www.rcsb.org/structure/3C79.
40 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1–2**, 19–25.
41 H. Lindahl, *Abraham and van der Spoel, GROMACS 2019 Source code*, 2018.
42 C. Bouysset and S. Fiorucci, *J. Cheminf.*, 2021, **13**, 72.
43 J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman and D. A. Case, *J. Am. Chem. Soc.*, 1998, **120**, 9401–9409.

44 P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case and T. E. Cheatham, *Acc. Chem. Res.*, 2000, **33**, 889–897.
45 I. Massova and P. Kollman, *Perspect. Drug Discovery Des.*, 2000, **18**, 113–135.
46 B. R. I. Miller, T. D. J. McGee, J. M. Swails, N. Homeyer, H. Gohlke and A. E. Roitberg, *J. Chem. Theory Comput.*, 2012, **8**, 3314–3321.
47 M. S. Valdés-Tresanco, M. E. Valdés-Tresanco, P. A. Valiente and E. Moreno, *J. Chem. Theory Comput.*, 2021, **17**, 6281–6291.
48 O. Trott and A. J. Olson, *J. Comput. Chem.*, 2010, **31**, 455–461.
49 J. Eberhardt, D. Santos-Martins, A. F. Tillack and S. Forli, *J. Chem. Inf. Model.*, 2021, **61**, 3891–3898.
50 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins*, 2010, **78**, 1950–1958.
51 J. Wang, P. Cieplak and P. A. Kollman, *J. Comput. Chem.*, 2000, **21**, 1049–1074.
52 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 712–725.
53 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
54 J. Loschwitz, A. Jäckering, M. Keutmann, M. Olagunju, O. O. Olubiyi and B. Strodel, *Data Brief*, 2021, **35**, 106948.
55 I. Doytchinova, M. Atanasova, E. Salamanova, S. Ivanov and I. Dimitrov, *Biomolecules*, 2020, **10**, 1323.
56 D. F. Hahn, V. Gapsys, B. L. de Groot, D. L. Mobley and G. Tresadern, *J. Chem. Inf. Model.*, 2024, **64**, 5063–5076.
57 H. Zhang, S. Kim and W. Im, *J. Chem. Inf. Model.*, 2022, **62**, 6084–6093.
58 M. W. Schmidt, K. K. Baldridge, J. A. Boats, S. T. Elbert, M. S. Gorgon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis and J. Montgomery, Jr., *J. Comput. Chem.*, 1993, **14**, 1347.
59 F.-Y. Dupradeau, A. Pigache, T. Zaffran, C. Savineau, N. G. R. Lelong, D. Lelong, W. Rosanski and P. Cieplak, *Phys. Chem. Chem. Phys.*, 2010, **12**, 7821–7839.
60 A. W. S. Silva and W. F. Vranken, *BMC Res. Notes*, 2012, **5**, 367.
61 H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura and T. Head-Gordon, *J. Chem. Phys.*, 2004, **120**, 9665–9678.
62 H. J. C. Berendsen, in *Computer Simulations in Material Science*, ed. M. Meyer and V. Pontikis, Kluwer Academic Publishers, Dordrecht, 1991, pp. 139–155.
63 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.
64 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
65 B. Hess, H. Bekker, H. Berendsen and J. Fraaije, *J. Comput. Chem.*, 1998, **18**, 1463–1472.
66 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
67 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
68 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.
69 S. Genheden and U. Ryde, *Expert Opin. Drug Discovery*, 2015, **10**, 449–461.
70 E. Wang, H. Sun, J. Wang, Z. Wang, H. Liu, J. Z. H. Zhang and T. Hou, *Chem. Rev.*, 2019, **119**, 9478–9508.
71 R. Kumari, R. Kumar and A. Lynn, *J. Chem. Inf. Model.*, 2014, **54**, 1951–1962.
72 A. Onufriev, D. Bashford and D. Case, *Proteins: Struct., Funct., Bioinf.*, 2004, **55**, 383–394.
73 F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 73–78.
74 F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2022, **12**, e1606.
75 V. Barone and M. Cossi, *J. Phys. Chem. A*, 1998, **102**, 1995–2001.
76 B. M. Bode and M. S. Gordon, *J. Mol. Graphics Modell.*, 1998, **16**, 133–138.
77 R. Sure and S. Grimme, *J. Comput. Chem.*, 2013, **34**, 1672–1685.
78 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
79 A. Hellweg, C. Haettig, S. Hoefener and W. Klopper, *Theor. Chem. Acc.*, 2007, **117**, 587–597.
80 K. Giri, B. K. Mishra and N. Sathyamurthy, *J. Indian Chem. Soc.*, 2021, **98**, 100101.
81 M. Garcia-Ratés, U. Becker and F. Neese, *J. Comput. Chem.*, 2021, **42**, 1959–1973.
82 S. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553–566.
83 F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt and J. H. van Lenthe, *Chem. Rev.*, 1994, **94**, 1873–1885.
84 M. Moustafa, M. Abu-Saied, T. Taha, M. Elnouby, M. El-shafeey, A. G. Alshehri, S. Alamri, A. Shati, S. Alrumman, H. Alghamdii and M. Al-Khatani, *Int. J. Biol. Macromol.*, 2021, **168**, 116–123.

