RSC Advances

PAPER

View Article Online
View Journal | View Issue

Cite this: *RSC Adv.*, 2025, **15**, 4262

Discovery of novel acetylcholinesterase inhibitors through AI-powered structure prediction and high-performance computing-enhanced virtual screening†

Beatriz Chafer-Dolz, *[a] José M. Cecilia,*[b] Baldomero Imbernón,[c] Estrella Núñez-Delicado,[c] Victor Casaña-Giner [a] and José P. Cerón-Carrasco *[d]

Received 8th November 2024
Accepted 29th January 2025

DOI: 10.1039/d4ra07951e

rsc.li/rsc-advances

Virtual screening (VS) methodologies have become key in the drug discovery process but are also applicable to other fields including catalysis, material design, and, more recently, insecticide solutions. Indeed, the search for effective pest control agents is a critical industrial objective, driven by the need to meet stringent regulations and address public health concerns. Cockroaches, known vectors of numerous diseases, represent a major challenge due to the toxicity of existing control measures to humans. In this article, we leverage an Artificial Intelligence (AI)-based screening of the Drug Bank (DB) database to identify novel acetylcholinesterase (AChE) inhibitors, a previously uncharacterized target in the American cockroach (*Periplaneta americana*). Our AI-based VS pipeline starts with the deep-learning-based AlphaFold to predict the previously unknown 3D structure of AChE based on its amino acid sequence. This first step enables the subsequent ligand–receptor VS of potential inhibitors, the development of which is performed using a consensus VS protocol based on two different tools: Glide, an industry-leading solution, and METADOCK 2, a metaheuristic-based tool that takes advantage of GPU acceleration. The proposed VS pipeline is further refined through rescoring to pinpoint the most promising biocide compounds against cockroaches. We show the search space explored by different metaheuristics generated by METADOCK 2 and how this search is more exhaustive, but complementary, than the one offered by Glide. Finally, we applied Molecular Mechanics Generalized Born Surface Area (MMGBSA) to list the most promising compounds to inhibit the AChE enzyme.

1 Introduction

Insecticide resistance is a major health issue in those insect vectors of diseases.[1–3] Most of the marked insecticides disrupt the biological activity of the acetylcholinesterase enzyme (AChE), a crucial neurotransmission in insects and other major pests.[4] The inhibition of AChE results in nerve signal disruption, leading to pets elimination without harming non-target species. However, a significant drawback in that pursuit has been the lack of detailed structural information about the AChE enzyme in specific species, which is crucial for designing targeted inhibitors.

Recent advancements in artificial intelligence (AI), particularly the development of AlphaFold by DeepMind,[5,6] have revolutionized the ability to predict protein structures with accuracy rivaling traditional experimental methods. This breakthrough has opened new frontiers in bioinformatics and drug discovery, allowing for the exploration of biological targets that were previously intractable due to the absence of structural data.[7,8] The convergence of high-performance computing (HPC) and AI has further amplified the potential of such structural predictions. HPC provides the required facilities to process the vast datasets and complex algorithms essential for training and running sophisticated AI models, including AlphaFold.[9] This synergy enables the rapid and accurate prediction of protein structures, accelerating the discovery process by enabling the VS of vast libraries of compounds against newly characterized targets, thereby facilitating the assessment of potential ligands against an expanded array of protein targets.

[a]*Bio Logic Crop Science, S.L. Amadeo de Saboya 1-4, Valencia, 46010, Spain. E-mail: bchafer@biologiccropscience.com*

[b]*Universitat Politécnica de Valencia (UPV), Camino de Vera S/N, Valencia, 46022, Spain. E-mail: jmcecilia@disca.upv.es*

[c]*Universidad Católica de Murcia (UCAM), Campus de los Jerónimos, Murcia, 30107, Spain*

[d]*Centro Universitario de la Defensa, Academia General del Aire, Universidad Politécnica de Cartagena, C/Coronel López Peña s/n 30729, Santiago de la Ribera, Murcia, Spain. E-mail: jose.ceron@cud.upct.es*

† Electronic supplementary information (ESI) available: (i) Parameters used for docking simulations; (ii) configurations of the selected METADOCK 2 schemes, labeled from M1 to M4; (iii) comparison of modeled cockroach AChE structure with the crystal structures of other AChE from different species; (iv) protein structure alignment results. See DOI: https://doi.org/10.1039/d4ra07951e

VS has been extensively implemented to accelerate the identification of potential therapeutics by exploring through extensive libraries of chemical entities, referred to as ligands, which may exhibit strong affinities to specific biological macromolecules, including receptors and enzymes.[10–12] An array of computational platforms, including AutoDock,[13] Glide,[14] BUDE,[15] and DOCK,[16] has been developed to facilitate the docking simulations. These platforms typically operate by positioning a ligand within a predefined active site on a target protein, either by referencing a co-crystallized ligand or the bare crystallographic structure of the protein. In contrast, alternative docking tools such as BINDSURF[17] and METADOCK[18,19] employ a "blind-docking" approach. This strategy avoids the limitation of a single binding site, instead conducting a comprehensive search across the entire protein surface, partitioned into discrete, independently evaluated regions. This holistic exploration might uncover potential activity sites that may not have been previously considered. However, it demands greater computational resources due to the simultaneous simulation of interactions at all feasible protein surface pockets. The efficacy of VS is tied to the precision of scoring functions, which are algorithmic estimations of the binding affinity following docking. Indeed, these scoring functions are critical for accurately ranking ligand candidates,[20] as they provide a quantifiable measure of the strength of non-covalent interactions between the ligand and the target protein.[21] It should be underlined that scoring functions are formed by a set of terms that simulate physical and chemical behaviors present in the molecular interaction. Four contributions are usually accounted for: (i) the electrostatic potential, (ii) the Lennard-Jones potential which mimics van der Waals forces, (iii) the hydrogen bonding potential and (iv) the desolvation potential. Although other applications introduce weights on several of these potentials based on their own experimental studies (*i.e.* Autodock and VINA), METADOCK has been coded without assigning weights. In the continuum of the drug discovery process, only the most promising leads, as determined by these scoring functions, progress from *in vitro* validation to *in vivo* testing and potentially to clinical trials.[10]

Indeed, the application of VS, molecular docking, and AI-driven predictive models such as AlphaFold, have revolutionized the identification and optimization of pharmaceuticals. However, that technology has been less exploited in other fields such as the development of novel insecticides. The challenges posed by insect pests, especially those resistant to conventional treatments, underscore the urgent need for innovative strategies that leverage the power of these methods.[22] They could enhance the specificity and efficacy of insecticides, targeting key biological pathways unique to pest species while minimizing environmental and non-target impacts.[23] For instance, leveraging structure-based drug design to target specific insect enzymes or receptors could lead to the development of compounds with novel modes of action, a critical need given the rapid evolution of resistance mechanisms.[24] Additionally, the adoption of AI models could accelerate the discovery process, efficiently screening chemical libraries to identify candidates with optimal properties for insect control.

Herein, we report a comprehensive study that implements the computationally predicted structure of the AChE enzyme as the starting material for conducting an enhanced VS pipeline. Leveraging the computational prowess of HPC, we have performed an in-depth *in silico* screening of the Drug Bank (DB) database,[25] aiming to identify novel inhibitors of the AChE. Our approach integrates advanced ligand-based methods with a consensus-based protocol, further refined by MMGBSA rescoring to pinpoint compounds with high potential for *in vivo* efficacy. The implemented protocol merges cutting-edge AI and HPC technologies, embodying their transformative impact on both; bioinformatics and pest control research. By validating the predicted AChE structure through traditional similarity search procedures, we have confirmed the reliability of AI-generated protein models and underscored the potential of these technologies to uncover novel bioactive compounds. The main contributions of this work are as follows:

1. To the best of our knowledge, this is the first time that AlphaFold has been leveraged to predict the structure of AChE for the American cockroach, which was further validated against traditional homology modeling techniques. This methodological innovation underscores the utility of AI in enhancing the accuracy and feasibility of protein structure predictions in species where empirical data are lacking.

2. The DB library of compounds has been screened against the AI-based AChE structure. An AI-based pipeline has been defined that performs an exhaustive and blind search over the entire surface of the predicted structure to guarantee a representative screening of the found compounds.

3. The application and assessment of metaheuristic-based VS alongside the commercial Schrödinger workflow provided insights into the performance and quality trade-offs critical in computational drug discovery. This comparison demonstrates the efficacy of metaheuristics in handling complex screening tasks and emphasized the importance of selecting appropriate computational strategies based on the specific requirements of the screening problem.

4. Our analysis reveals that metaheuristics offer an efficient/fast throughput and a broader interaction with the target protein, although they must be finely tuned to balance the computational speed and the accuracy of the results. This balance is crucial for optimizing the effectiveness of VS in identifying viable insecticidal candidates.

2 Related works

Even if the literature is still scarce, a few works have applied computational methods to find new insecticides. One recent example is the work conducted by Da Silva *et al.* with a focus on combating *Aedes aegypti*, a mosquito that acts as a vector for transmitting diseases such as dengue fever and Zika.[26] These authors search novel insecticides by using pyriproxyfen as a template structure to identify structurally similar molecules from the Zinc Natural Stock (ZNSt) database, employing tools like ROCS and EON for molecular comparison. That early screening included pharmacokinetic and toxicological

assessments. This work leads to two promising molecules with large binding affinities to AChE and the juvenile hormone.

Other excellent work has been published by Gang *et al.*, who addressed the pressing issue of pesticide resistance by employing a computer-aided drug design (CADD) methodology to synthesize novel sulfonamide pesticides.[27] Their study implements homology modeling and VS to identify lead compounds, which are then chemically synthesized through a multi-step process starting from *p*-chlorocresol. The binding interactions of the top-ranked compounds were further explored through docking simulations and structure–activity relationship analysis. Their efficacy was tested using the leaf-dipping method against *Mythimna separata*. Seven of these derivatives exhibited superior insecticidal activities. One can find in the literature applications of machine learning (ML) as an advanced tool for the discovery of insecticides. Ding *et al.* explore eco-friendly and multitarget inhibitors targeting insect chitinolytic enzymes, a novel approach in the field of pesticide development.[28] Their research integrates ML with molecular docking by merging data from prior high-throughput screening to navigate a vast library of 17 600 natural compounds. This strategy successfully identified potent inhibitors, such as 3,5-di-*O*-caffeoylquinic acid and γ-mangostin, which demonstrated inhibitory action against multiple chitinolytic enzymes from *Ostrinia furnacalis* (a species of moth). The transcriptomic analysis of treated insects further elucidated the biochemical pathways affected by these compounds, underscoring the synergistic potential of ML in pesticide discovery.

Regarding the prediction of protein structures related to insecticide activities, some authors recently applied AlphaFold to find these novel targets. For instance, Zhorov and Dong assessed the challenge of pyrethroid resistance in arthropod pests, a significant hurdle in effective pest management and disease vector control.[29] Their study is specifically focused on mutations within the voltage-dependent sodium channels (VGSC), which is the primary targets of pyrethroid action. By leveraging the AlphaFold2 neural network, they generated refined models of mosquito and cockroach sodium channels, enabling a detailed investigation of pyrethroid interactions in various channel states through computational docking. Their findings hint how mutations, including those distant from known receptor sites, influence pyrethroid efficacy. It was also remarkable that molecular models were able to identify allosteric modulation of receptor site with a large impacts in pesticide binding.[29] Tahir *et al.* focus their research on combating *Ostrinia furnacalis*, a major worldwide pest for maize and other corn crops.[30] Their study employs a structure-guided computational approach to discover environmentally friendly insecticides by targeting the β-*N*-acetyl-D-hexosaminidase, an essential enzyme for the pest's survival but distinct in its substrate binding from its plant and human counterparts. These authors used Glide by Schrödinger to perform all VS predictions. Their simulations were based on the crystal structures of OfHex1 and its homologues in humans and the Alphafold model of β-*N*-acetyl-D-hexosaminidase from *Trichogramma pretiosum*, a beneficial parasitoid. Kong *et al.* also used AI methodologies to streamline the discovery of novel insecticides with unique chemical structures.[31] Their study efficiently integrates deep learning models to predict and generate potential neonicotinoid insecticides, which are subsequently validated by both VS and experimental assays.

Our groups have also contributed to the use of computational methods in the search of better insecticides. We have successfully used VS for targeting known biological pathways in pest control, *e.g.*, VGSC in the American cockroach.[32] The performed VS simulations allowed us to identify miglitol as a potent synergist, as confirmed by animal model assays. The present work aims to extend the scope of VS by integrating emerging AI technologies by predicting protein structures previously uncharacterized. The predictive power of AlphaFold is exploited to explore protein targets without prior structural data, offering a novel pathway to insecticide development that is not limited by the current knowledge of target structures. As discussed below, our strategy combines AI-driven protein structure predictions with enhanced VS and rescoring for broadening the target base in insecticide development.

3 Materials and methods

3.1 Biomolecular target discovery

The most classical computational strategy for discovering bioactive molecules requires the crystal structure of a known target as starting material, an approach usually called structure-based design.[33] However, the resolution of complex biological targets is not always possible and consequently, the standard VS procedure cannot be directly implemented. In these cases, where the crystal structure of the target is unresolved, homology modeling emerges as an alternative. This technique exploits the evolutionary conservation of protein folds, using the resolved structures of similar proteins as templates to predict the three-dimensional structure of the protein of interest. This process is based on alienating the target sequence with sequences of homologous proteins of known structure while completing missing parts.[34] The production of homology models enables the subsequent application of structure-based VS protocols. This is not a trivial task. The robustness of homology modeling is contingent upon the availability of a suitable template and the degree of conservation between the target and the template's active site. When a template with high sequence similarity is available, the resulting model is often of enough quality to allow for meaningful VS procedure of potential ligands.[35] These models can be further refined using computational methods such as loop modeling, side-chain optimization, and molecular dynamics simulations, enhancing the accuracy of the predicted binding sites.[36,37]

In the present study, our bioactive target, AChE from the American cockroach (*Periplaneta americana*), lacks crystallographic data. The construction of a structural model is, therefore, a prerequisite to facilitate our computational screening efforts. In absence of X-ray structures, modeling was initiated by searching for available nucleotide sequences in the well-known GenBank, a public database that contains genetic sequences for *ca.* 500 000 different species.[38] The sequence for *P. americana* AChE was identified by accession number ALJ10969.1. The

homology calculations were performed with the standard protocol implemented in Schrödinger, which resulted in a structural homologue deposited in the Protein Data Bank (PDB) with code 5X61. This structure, which corresponds to the *Anopheles gambiae*, was previously reported with a focus on malaria vectors.[39] For the records, the identity of 73.18%, sequence positivity of 82.95%, and minimal gaps accounting for only 0.22% of the sequence alignment. It is also worth noting that the use of that specific template matches to an earlier model for *P. americana* AChE suggested by Rajashekar and co-workers.[40] These authors identified the same template by using an alternative homology code, *e.g.*, SWISS-MODEL program.[41] That agreement further confirms the accuracy of the proposed homology strategy.

An alternative method for constructing a reliable AChE structure for the American cockroach (*Periplaneta americana*) was the application of AlphaFold, the next-generation artificial intelligence (AI) system developed by DeepMind.[6] AlphaFold has revolutionized computational biology by accurately predicting protein structures solely from their amino acid sequences. The system leverages a deep learning model extensively trained on a diverse dataset of known protein structures, enabling it to capture intricate patterns in sequence–structure relationships. AlphaFold's innovative neural network architecture integrates evolutionary information, physical, and geometric constraints, predicting the three-dimensional spatial arrangement of amino acids with remarkable precision.[42] Of course, homology models and AlphaFold structures are complementary rather than competitive solutions. The homology modeling performs more reliably than AlphaFold if the high-resolution homologous template is available for the homology modeling, while AlphaFold might be optimal for more critical systems.

To enhance accessibility and computational efficiency, the ColabFold implementation of AlphaFold2 was selected in this study.[43] This version integrates MMseqs2 for generating sequence alignments and HHsearch for identifying structural templates, streamlining the prediction process in a cloud-based environment. The ColabFold pipeline, configured for this task, included options for modeling both monomeric proteins and complexes, along with the capacity for Amber-based relaxation, further refining the predicted structures.

AlphaFold's methodology, which estimates inter-residue distances and angles to construct detailed 3D models, has been rigorously validated, notably in the Critical Assessment of Protein Structure Prediction (CASP) competitions, where its predictive accuracy was found to be on par with experimental methods.[7] For the AChE of *P. americana*, where no experimental crystallographic data were available, AlphaFold enabled the derivation of a high-fidelity structural model. This model provided the necessary foundation for subsequent virtual screening (VS) processes, facilitating precise simulations of potential inhibitor interactions at the enzyme's active site. Thus, the predictive capabilities of AlphaFold (*i.e.* ColabFold framework) were instrumental in accelerating our biocide discovery efforts, highlighting the wide impact of AI in advancing the field of drug discovery.[44]

3.2 VS pipeline

The interaction of novel insecticides with the AChE target was assessed using two VS strategies. The first approach was based on the standard protocol implemented in Glide, which performs three independent docking stages, including an initial high-throughput virtual screening (HTVS) with a subsequent single-precision and extra-precision docking (SP and XP, respectively).[45,46] This systematic method effectively reduces the initial set of over 10 000 potential candidates deposited in the DB database to around 1000 candidates during the initial HTVS stage. We stress that both HTVS and SP employ identical scoring mechanisms, though HTVS is characterized by a less demanding approach to torsional refinement and sampling. This allows for the swift elimination of ligands poorly suited to bind the target. The SP phase refines this selection further by redocking the top 1000 candidates, which is then narrowed down significantly in the XP stage. Only the top 100 scoring ligands undergo redocking with enhanced accuracy.[47] For the records, all settings were used as default in Glide, where 10% of the top-ranked compounds are retained at each step. We only imposed one change in that default parameter, by increasing the number of poses up to 10 at the XP step. Aiming to explore the whole AChE structure binding sites were determined by using the SiteMap tool also implemented in the Schröringer suite of programs. SiteMap has been designed for scanning the protein surfaces and consequently predicting target druggability.[48]

The VS workflow implemented in Glide allows for a quick VS of large libraries of compounds. However, that code is optimized for targeting specific binding sites by using crystallographic ligands and/or druggable sites predicted with SiteMap. It also restricts the generation of ligand poses in the default parameter to a maximum of 3 per ligand. The increase of computed poses (10 in our study) leads to a critical limitation to maintain computational efficiency as Glide relies exclusively on CPU resources.

As an alternative to the standard solution offered by Glide, METADOCK 2 was used for performing more systematic VS experiments. METADOCK 2 is an optimization procedure based on structured metaheuristics.[18,19] In short, metaheuristics effectively address complex optimization challenges, often classified as NP-hard problems, due to their ability to quickly converge on satisfactory solutions in scenarios with limited computational resources or incomplete data.[49] These algorithms prioritize the exploration of the most promising candidates rather than exhaustively searching all possible solutions, which allows for near-optimal solutions in a reasonable time frame. There are several metaheuristic algorithms,[50] including Scatter Search, Genetic Algorithms, Ant Colony Optimization, Tabu Search, Hill Climbing, or Simulated Annealing.[51] It should be underlined that METADOCK 2 is not based on a single specific metaheuristic but able to generate different metaheuristics by setting its input parameters.

The implemented schema integrates multiple functions commonly found across various metaheuristics, as discussed in ref. 49. Each function within the algorithm requires specific parameters that are listed in the ESI.† Upon setting these

parameters, METADOCK 2 executes its optimization routine to select a collection of poses that optimally reduce the scoring function. Indeed, METADOCK 2 is an iterative procedure with a given termination criteria (defined by the End_condition(S, ParamEnd)) that can be either: the maximum number of iterations or the number of consecutive iterations without improvement. On the one hand, the Select(S, Ssel, ParamSel) function then filters a subset of the initial population (S), selecting both; the best and worst conformations to promote diversity based on their scoring function values. On the other hand, the Combine(Ssel, Scom, ParamCom) function pairs these selected conformations for mixing, targeting combinations of the best, the worst, and cross-pairs of both to generate new variants. The Mutation(Scom, ParamMut) function then introduces variations to these combinations to maintain diversity, altering aspects such as spatial coordinates or orientations. The Improve(Scom, ParamImp) function conducts a local search to refine these new conformations after mutation. The loop finishes with the Include(Scom, S, ParamInc) function, which includes a selection of these refined conformations back into the population for the next iteration, ensuring a continuous evolution of potential solutions.

METADOCK 2 implements two distinct scoring functions, both grounded in conventional force field models that consider various molecular interactions including dispersion–repulsion, hydrogen bonds, electrostatics, and desolvation. We refer the reader to[19] for further details. Each scoring function incorporates unique modifications to its formulae to adapt to specific computational environments. Withing that framework, METADOCK 2 offers an alternative strategy to expand the scope of the search, providing a broader and more exhaustive exploration of the docking landscape. This comparison highlights the complementary roles of both systems in the VS process, with METADOCK 2 enhancing the exploration capabilities beyond the targeted approach of Glide.

The inherent differences in the scoring functions of Glide and METADOCK 2 impede a direct comparison or combination of their respective numeric outputs. To standardize these values, the binding energies of the highest-ranked hits from both methods are refined by using MMGBSA method as implemented in Prime module.[52,53] This refined computational approach calculates the binding-free energies as:

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

where ΔG_{bind} signifies the net binding-free energy, and G_{complex}, G_{target}, and G_{ligand} represent the free energies of the complex, target AChE, and ligand respectively. The equation may be expanded into:

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

where ΔE_{MM} includes all gas phase energies, ΔG_{GB} accounts for polar solvation energy calculated using the generalized Born method, ΔG_{SA} measures the nonpolar solvation component, and $T\Delta S$ denotes the change in conformational entropy.[54] These components collectively yield the binding-free energy of each ligand as estimated by the MMGBSA approach and allow for a fair comparison of the binding energies predicted by Glide and METADOCK 2.

4 Results

This section summarized how the use of AI methodologies discussed in Section 3 leads to novel candidates against the AChE of the American cockroach. We first describe the benchmarking environment where simulations have been conducted. The main configuration parameters, targeted database of compounds, and hardware environment are defined. We next discuss the use of the different computational tools used to carry out VS for AChE-inhibiting compounds. To this end, the prediction and validation of the AChE protein structure are discussed. The results derived from the proposed VS strategies are finally assessed and discussed with a more practical applications industry.

4.1 Benchmarking environment

The experiments underpinning these results were conducted on a computational node with an Intel(R) Xeon(R) Silver 4216 CPU@2.10 GHz, boasting sixteen physical cores augmented to thirty-two threads through hyperthreading. Complementing the CPU, the node's computational prowess is bolstered by 384 GB of DDR5 RAM and an NVIDIA A100 GPU. This Tensor core-equipped GPU has a capacity of up to 80 GB of memory at over 2 terabytes per second (TB per s), to run the largest AI models on large datasets.

As in any VS experiment, the selection of a compound repository is a key factor. For this study, we selected a well-established database known for its utility in drug design, albeit its potential for insecticide development has been comparatively underexplored. The DB database has been systematically screened,[55] a vast array of compounds exceeding 10 000 entries, encompassing approved drugs, nutraceuticals, experimental agents, as well as those that have been withdrawn from the market.[56] Until now, DB has been mostly used in drug design by targeting human biomolecules. However, we have recently demonstrated that library of compounds might be also used as a source for other applications, including the discovery of novel insecticides.[32] Herein, the molecular structures deposited in DB were cured by using the LigPrep tool code by Schrörindger's, which assigns bond orders, computes the predominant protonation states at neutral pH, and performs structural optimization with the OPLS4 force field.[57]

The deployment of METADOCK 2 in our study requires the strategic configuration of its parameters. These parameters govern the operational breadth of the fundamental metaheuristic algorithms for exploring the varied topography of search spaces with differential computational effort. These configurations, coded as M1, M2, M3, and M4, have been systematically designed to progressively extend the computational rigor and expand the potential for probing the conformational complexities of ligand–receptor interaction (see ESI†). These configurations are:

• M1: this setup is the baseline model, characterized by minimal parameter intensification. M1 is specifically configured for rapid, initial screenings, making it ideal for preliminary assessments where computational efficiency is prioritized over exhaustive search depth. This configuration allows for a swift traversal of the search space to quickly identify potential hits with reasonable computational resources.

• M2: building on the M1 configuration, M2 introduces moderate enhancements to the flexibility and the parameters governing initial improvement functions. This configuration is designed to strike a balance between computational speed and the depth of exploration. It adjusts the model's parameters to allow for a more detailed exploration of the docking space without a substantial increase in computational demand, making it suitable for more refined screenings.

• M3: the third configuration marks a significant escalation in the complexity of the search process. It enhances the selection and combination parameters extensively, supporting a more thorough exploration of the docking space. M3 employs a large number of initial conformations and emphasizes the strategic combination of the best and worst poses. This approach is intended to explore a broader range of potential interactions between the ligand and the receptor, potentially uncovering unique and efficacious binding configurations.

• M4: the most complex of the configurations, maximizes both computational intensity and search depth. It uses the largest number of initial conformations and intensifies each parameter to push the limits of metaheuristic search capabilities. M4 is oriented towards exhaustive computational experiments where the goal is to explore the entire search space for possible protein–ligand conformations that may have optimal interactions, albeit at a higher computational cost.

Each configuration incrementally builds on the previous, offering a spectrum of strategies from rapid screening to deep, exhaustive searches. This systematic escalation in complexity allows researchers to select a configuration that best suits the specific demands and resources of their study, aligning computational effort with the desired thoroughness of the screening process.

4.2 Chemical models

As discussed, the AChE of the American cockroach remains unresolved, which in turn impedes the use of the more classical protein–ligand VS. Fortunately, we do know the sequence for *P. americana* AChE, denoted by the GenBank sequence code ALJ10969.1. As discussed, AlphaFold takes advantage of such biological information to build up the target structure, a required step to enable structure-based VS simulations. To ensure that AI leads to meaningful structure, the generated target by AlphaFold is compared to the classical Homology strategy, which is mainly based on the search of similar sequences that are subsequently used as structural templates.

Fig. 1 shows the overlapping structures of AChE predicted by homology models with Schrödinger (green) and AlphaFold (white) both given in cartoons. This figure illustrates that both approaches exhibit a remarkable structural alignment with an

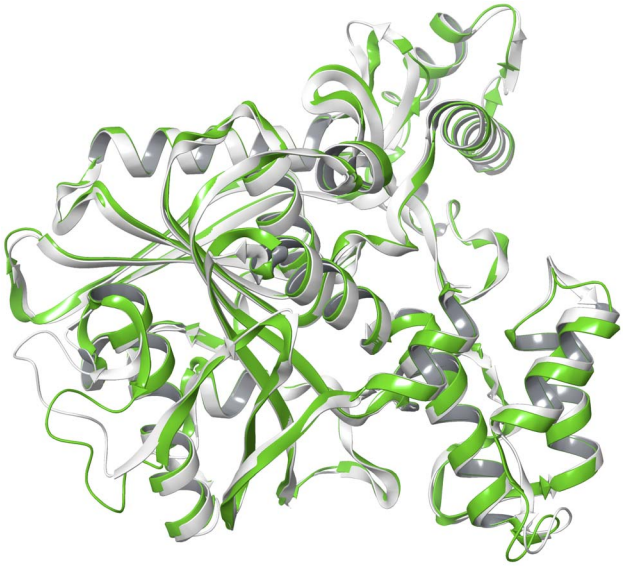

Fig. 1 Overlays for the generated AChE target models displayed as cartoons. Color scheme: homology model in green and Alphafold structure in white.

Root Mean Square Deviation (RMSD) value of 0.89 Å only, an agreement that confirms the predictive accuracy of the AlphaFold algorithm about the established homology modeling approach. That match is also observed when superposing the developed AChE models for cockroach with the experimental data available for other insect as the fruit fly *Drosophila melanogaster* (see ESI†). This structural agreement match backs up the structural consistency between the computationally derived models and validates the use of AlphaFold for generating reliable protein structures without crystallographic data.

Visual inspection and RMSD values suggest that both methods lead to very similar structures so that both can be used as targets for VS simulations. To confirm this hypothesis, we have performed a comparative analysis to compare the interaction energies between the DB compounds and each of both targets. Fig. 2 shows a box-and-whisker plot that conveys the distribution of interaction energies achieved with the less demanding M1 metaheuristics of METADOCK 2. On the *x*-axis, it can be shown the two categories of scores: those obtained using the AlphaFold-predicted AChE structure (ScoringAlphaFold) and those derived from the homology-based AChE model (ScoringHomology). Each box in the plot encapsulates the interquartile range (IQR) of scores, representing the middle 50% of data for each structure, with the horizontal line inside the box denoting the median score. The whiskers extend to cover the total range of data, excluding outliers, denoted by individual points beyond the whiskers. These outliers represent scores, significantly higher or lower than the bulk of the data and could be indicative of exceptionally strong or weak binding affinities. The scoring distribution also suggests that both the AlphaFold and homology models have a similar range of binding affinities across the screened compound library, as evidenced by the similar span of the whiskers and the IQR. It is also important to note that the score values are negative; in this

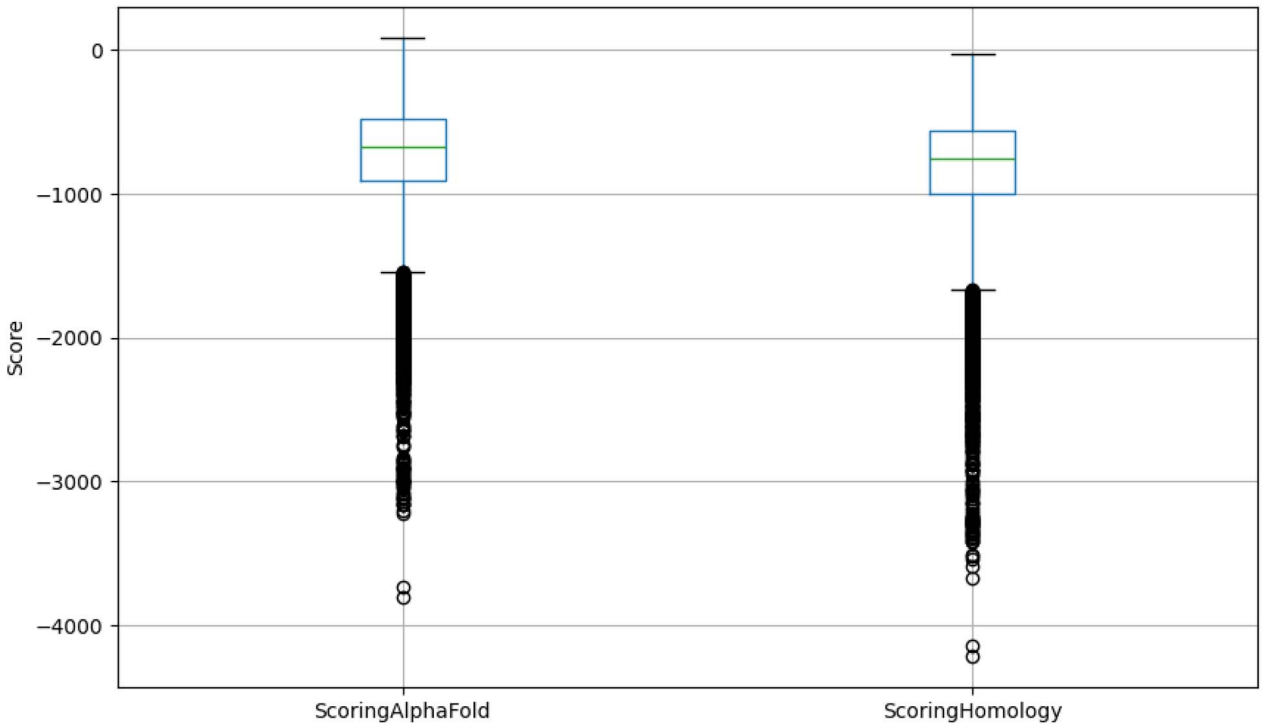

Fig. 2 This scatter plot compares the binding affinity scores obtained from blind docking of DB compounds against two receptor structures: one derived through homology modeling (blue) and the other predicted by AlphaFold (red). Each point represents the docking score for a single compound, showcasing the overall distribution and highlighting the relative binding affinities predicted for each receptor. The data indicate a similar trend.

framework, a lower (more negative) score stands for stronger predicted binding affinity. The similarity in the spread of scores between the two models could suggest that despite the methodological differences in structure prediction, the overall binding landscapes that they predict are comparable. This congruence is critical as it lends credence to the reliability of VS results irrespective of the structural model used for modeling AChE.

4.3 Computational performance

Fig. 3 shows the performance of the four metaheuristics generated by METADOCK 2. Blind docking simulation implements the AChE receptor generated by AlphaFold and all compounds included in the DB database. As discussed above, the nomenclature of these metaheuristics adopts a sequential designation, where M1 is attributed to the most computationally agile algorithm, ascending in resource intensity up to M4. The abscissa quantifies the improvement in computational performance, expressed as a speedup factor when juxtaposing the pairwise metaheuristics. The ordinate captures the variance in the score function, a metric reflecting the estimate of binding affinity. This numerical value can be expected to decrease with a more refined search, which, of course, increases the computational effort. Close observation reveals that metaheuristic M1, when compared to M2 (blue circles), shows accelerated computational performance, albeit with wide variance in the score function differential. The data sets suggest a heterogeneous response to efficiency improvements, implying a multifaceted trade-off between computational speed and fidelity of affinity predictions. Progressing to the M2 *versus* M3 comparison (orange crosses), a convergence of data points intimates a lesser magnitude of speed-up, aligned with a narrower dispersion in scoring function differentials. This pattern insinuates a progression towards balance—a moderate sacrifice in computational speed in favor of scoring precision. The M3–M4 (red squares) presents the most intriguing paradigm; the computational acceleration is minimal, implying a near-equivalence in execution time. Nonetheless, there is a pronounced leap in scoring function differentials in favor of M4, suggesting that M4, while computationally intensive, potentially achieves a marked improvement in the prediction of binding affinities. Together, these data points illuminate the intrinsic trade-off between computational convenience and the accuracy of ligand–protein interaction predictions. This graphical representation underscores the sophistication inherent in developing computational strategies for drug discovery, where algorithmic advances are meticulously calibrated to optimize both execution speed and predictive modeling insight.

Fig. 4 provides a visual comparison between the search space coverage of traditional Schröndinger workflow using GLIDE and the metaheuristics developed by METADOCK 2, as applied to a molecular docking study. Rendered in a ribbon diagram, the protein is shown as a gray structure, with the potential ligand binding poses indicated by colored points: METADOCK 2 configurations M1 through M4 are represented by red, blue,

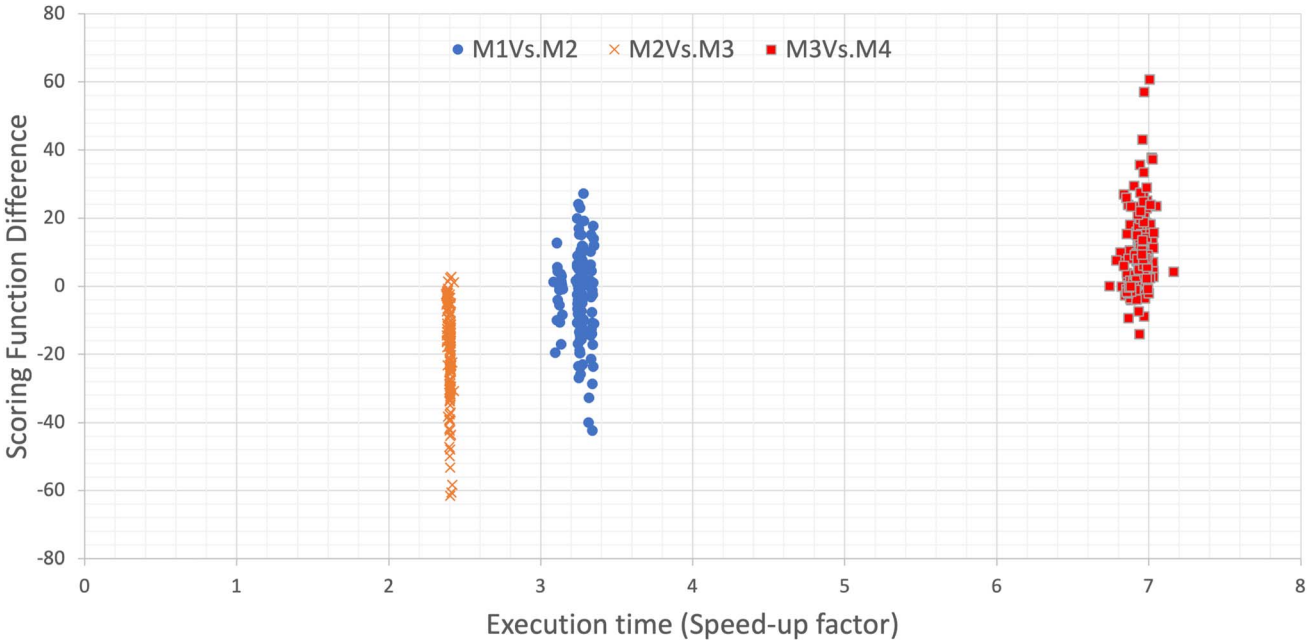

Fig. 3 Comparative performance analysis of metaheuristic algorithms in METADOCK 2: this figure illustrates the trade-offs between execution time and scoring function accuracy for different metaheuristics (M1, M2, M3, and M4) utilized in the screening of the top 50 compounds from the DrugBank database. Each point represents a pair of metaheuristics compared, with the *x*-axis showing the speed-up factor in execution time and the *y*-axis indicating the scoring function difference. The data suggest a balance between computational efficiency and the precision of binding affinity predictions across the metaheuristics evaluated.

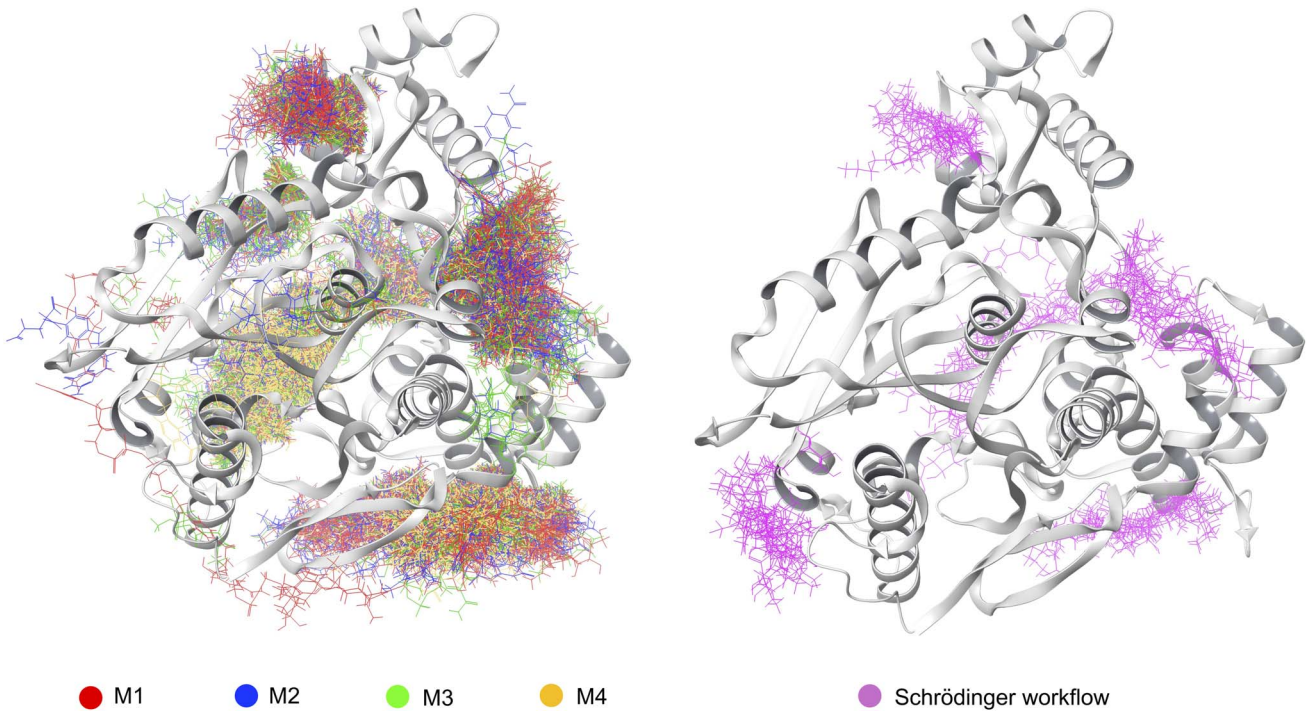

Fig. 4 Comparative analysis of search space coverage by METADOCK 2 metaheuristics (left panel) and Schröndinger Glide workflow (righ panel). The protein structure is depicted in gray ribbon representation. Docking poses are visualized as color-coded points: METADOCK 2 configurations M1 (red), M2 (blue), M3 (green), and M4 (yellow) illustrate the extensive search space covered through blind docking, while the standard Schröndinger Glide workflow is shown in purple, indicating a more targeted approach. The image highlights the comprehensive search performed by METADOCK 2 across the protein surface, contrasting with the focused search areas of the Glide workflow.

green, and yellow points, respectively, while the Schröndinger workflow is denoted in purple. Distribution based on the METADOCK 2 blind docking approach yields a more exhaustive search across the protein surface, resulting in a multitude of poses that dwarf the number generated by the Schröndinger workflow. This comprehensive search confirms the ability of METADOCK 2 to perform an extensive conformational analysis. As discussed, the standard VS workflow implemented in Glide leads to a lower number of poses. However, one can observe a match in the areas explored by both methodologies. That outcome hints that while the approaches may differ in thoroughness, they converge on similar regions of the protein that are likely of high binding relevance. In addition, the density and spread of the METADOCK 2 poses suggest an intense, stochastic sampling of the ligand space, capturing a wide array of potential binding modes that can then be further evaluated for their biochemical viability. In contrast, the Schrödinger workflow, with its targeted docking strategy, searches on predefined regions of interest that might overlook novel interaction sites that METADOCK 2 can uncover. These numeric outcomes further validate METADOCK 2 in identifying a rich landscape of binding configurations, offering a valuable complement to the targeted docking approach traditionally employed in computational drug discovery.

Fig. 5 summarizes the impact of the post-refinement through the MMGBSA. That level of theory is used to increase the accuracy of the binding energies generated by the metaheuristics M1 through M4, which are arranged in ascending order of their inherent computational complexity. Represented in this graphical analysis are the differential rescoring values between successive pairs of metaheuristics: M1 *versus* M2 (blue circles), M2 *versus* M3 (red triangles), and M3 *versus* M4 (green crosses). The abscissa enumerates the DB compounds subject to this analysis, while the ordinate captures the resultant variance in the rescoring values post-MMGBSA application. Negative values on the ordinate suggest an enhancement in the predicted binding affinity post-rescoring for poses generated by the more computationally complex metaheuristic. Positive values, however, indicate a superior rescoring result for poses obtained by the comparatively less complex metaheuristic within the compared pair. The distribution of data points, particularly the prevalence of points along and above the zero line, suggests that the MMGBSA rescoring results in similar or improved scoring outcomes for poses generated by less complex metaheuristics. This observation infers that the optimization inherent in the MMGBSA procedure can, to some extent, mitigate the disparities in the initial scoring generated by different levels of metaheuristic complexity. Notably, the frequency of negative values is diminished relative to previous comparisons in Fig. 3, which may be attributed to the intrinsic refinement capability of MMGBSA. This rescoring method independently evaluates and potentially improves the predicted binding affinities by accounting for factors such as solvation effects and entropic contributions, thus providing a more nuanced view of the thermodynamics of ligand-receptor interaction. All that accumulated numeric results demonstrate that while the adoption of more sophisticated metaheuristics is generally associated with an initial predictive advantage, the subsequent application of MMGBSA rescoring can effectively balance and enhance the binding affinity predictions, which is specially true for those poses generated by less computationally demanding algorithms, *e.g.*, M1.

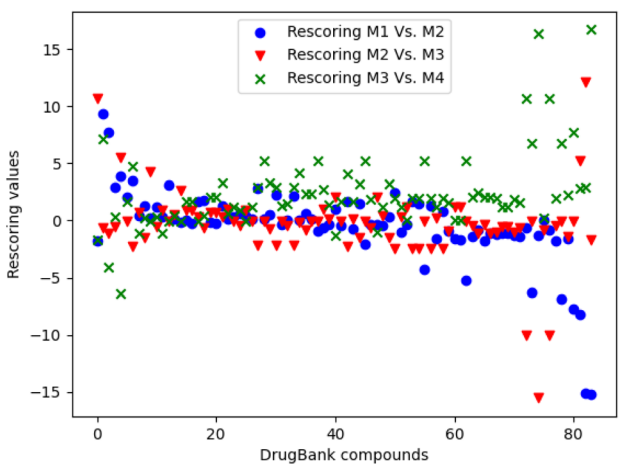

Fig. 5 Differential impact of MMGBSA rescoring on metaheuristic-generated poses. This scatter plot contrasts the differences in MMGBSA rescoring outcomes for docking poses derived from a spectrum of metaheuristic algorithms (M1–M4) implemented in METADOCK 2. Each point represents the rescoring differential between successive metaheuristics for DB compounds, indicating the ability of MMGBSA optimization to converge the predictive accuracy of various computational strategies towards a cohesive standard.

4.4 Novel insecticides

Let us finally collect all the results reported by using the standard Glide workflow and the enhanced sampling by METADOCK 2. Table 1 depicts the top-ranked compounds with the potential to target the AChE of the American cockroach. The criteria for selection was based on their refined binding-free energy at the MMGBSA level of theory, with a threshold set for values lower (more negative) than −50 kcal mol^{-1}, indicative of strong potential interactions.[32] On the left panel, molecules are listed by their DB identifiers or generic names when available, alongside their respective Glide scores and refined binding-free energy interactions with the main target, namely, AChE (ΔG_{AChE}). Aiming at providing a more general view, the interaction with the voltage-activated sodium channel (Na_v) has been also computed, which is labeled as ΔG_{Na_v}. This dual-target evaluation addresses the possibility of cumulative efficacy, which is particularly pertinent in the context of insecticidal action. The right panel of Table 1 includes numeric outcomes from the enhanced sampling by using METADOCK 2 with both targets.

It is remarkable that among the ∼10 000 compounds deposited in the DB database, both screenings match Nadide in their top-ranked list. Nadide is a coenzyme involved in numerous enzymatic reactions, which indicates the robustness of the screening across different search algorithms. We also stress that a Nadide analogue, *e.g.*, DB02732, is present in the

Table 1 Top-ranked compounds predicted by the standard Schröndinger protocol (Glide score, in kcal mol^{-1}) and by sampling with blind-docking METADOCK 2. Interaction energies are refined binding-free energy at MMGBSA level of theory (ΔG_{AChE}, in kcal mol^{-1}). Top-ranked compounds are also used to target the voltage-activated sodium channel (ΔG_{Na_v}, in kcal mol^{-1}) to assess cumulative efficacy

Standard workflow				Enhanced sampling			
Molecule[a]	Glide	ΔG_{AChE}	ΔG_{Na_v}	Molecule[a]	METADOCK 2	ΔG_{AChE}	ΔG_{Na_v}
DB02226	−12.60	−98.99	−50.59	Capivasertib	−100.43	−69.07	−50.01
RPR128515	−10.75	−65.25	−46.03	Netilmicin	−88.16	−66.81	−53.78
DB02732	−12.54	−63.75	−29.76	Acarbose	−112.81	−65.92	−39.73
DB06858	−12.97	−62.57	−50.24	Pf 04995274	−117.96	−65.01	−50.81
Zorubicin	−9.86	−55.79	−38.21	Domperidone	−99.55	−57.99	−56.78
Hexoprenaline	−10.53	−54.38	−63.01	Nadide	−100.43	−54.49	−40.09
DB06942	−11.10	−52.74	−49.11	Quinine	−95.89	−53.55	−30.78
DB07054	−9.21	−52.67	−51.43	Miglustat	−78.51	−53.36	−35.18
Nadide	−9.82	−51.03	−40.09	Sotalol	−79.25	−51.10	−39.41
				Tetracaine	−94.25	−50.13	−37.20

[a] Generic names are given when available, otherwise molecules are identified by DB codes.

standard VS workflow list, suggesting the algorithm's effectiveness in identifying biochemically relevant molecules. It is also worth mentioning that Nadide is present in commercial products, such as the referenced shampoo, which underscores its wide applicability and market availability. Finally, one can also note that the associated energies to the binding of the AChE (ΔG_{AChE}) are larger (more negative) than the values obtained for the sodium channel (ΔG_{Na_v}). This is an expected result, as the selection list was specifically based on the interaction to AChE. However, some hits fulfill the prerequisite of the −50 kcal mol^{-1} in both targets: DB02226, DB06858, hexoprenaline and DB07054 in the Glide-based standard workflow; capivasertib, netilmicin, Pf 04995274 and domperidone for the enhanced sampling with METADOCK 2. These compounds might consequently offer a more lethal insecticide action if simultaneously blocking both targets.

All accumulated data demonstrated that the enhanced sampling by METADOCK 2 increases the multiplicity and dispersion of predicted poses, and consequently conducts a more expansive search over the surface of the target. This blind docking method reveals a diverse array of candidate molecules with significant predicted interactions, as evidenced by their MMGBSA scores. Table 1 eventually provides an illustrative picture of the VS most promising findings. The reported lists (standard and enhanced samplings) might be directly implemented in experimental assays to confirm the synergy between traditional computational techniques and innovative AI-enhanced methodologies in the search for effective insecticidal agents.

5 Discussion

Our search for better inhibitors to the American cockroach AChE reveals the potential use of AI-powered VS processes in the development of more efficient insecticides. The performed analysis began with the challenge of determining the tertiary structure of the AChE protein, as it remains unresolved in the literature; a key step for screening of the DB library of compounds. The application of the AlphaFold algorithm in this specific framework demonstrated its predictive capacity, generating a tertiary structure very much in line with the most classical homology processes. Accordingly, structural bioinformatics based on AI introduces a novel framework that improves the current state of the art, yielding considerable time savings for molecular modelers and facilitating broader access to these tools for the scientific community.

The integration of metaheuristic algorithms within the METADOCK 2 framework through GPU acceleration might help in the development of enhanced VS methodologies. The comparison of METADOCK 2 *vs.* the standard Schrörindger Glide workflow demonstrated the comprehensive nature of metaheuristic searches while confirming the ability of HPC to increase the efficiency and precision of these algorithms. METADOCK 2 exploits the parallel processing capabilities inherent to GPUs, facilitating a broad and intensive exploration of the docking landscape. This exhaustive approach permits the investigation of an expansive array of ligand–receptor interactions across the entire protein surface, surpassing the traditional targeted docking methodologies that focus on predefined active sites.

Furthermore, our work demonstrates that AI algorithms in general and metaheuristics in particular must be correctly parameterized to optimize the balance between computational cost and predictive capacity. As it has been shown, the most powerful metaheuristics (in general the models) do not necessarily correlate the best results. On the contrary, the computational strategy might require the specific adaptation of the model to the problem domain to optimize the quality/performance ratio. The convergence of metaheuristics and HPC exemplifies a powerful toolkit for modern bioinformatics, setting a new benchmark for drug discovery.

We complete our study by assessing the binding energies of novel ligands to the AChE model generated for the American cockroach. The candidate poses selected by the two employed screening methods (*i.e.*, Glide and METADOCK 2) were refined using MMGSBA, culminating in a set of prospective inhibitors

of this protein. Furthermore, another action mode for insecticides targeting the American cockroach, specifically the sodium channel, which has been documented in the literature,[32] was employed as a reference. The interaction energies of compounds selected through our investigative process with this receptor were presented, thereby providing an enriched perspective of this screening to the reader. Notably, the compound Nadide was highlighted as active against both action modes, rendering it an exceptionally promising candidate for subsequent *in vivo* assays. Additionally, several other compounds exhibiting significant activity emerged as auspicious contenders for further investigation of their efficacy within this biological model that might also include higher levels of theory, including quantum mechanical simulations and/or more exhaustive MMGBSA sampling trajectories in during molecular dynamic simulations.[58–62]

6 Conclusions and outlook

This article aims to contribute to the convergence of HPC and AI by paving the way towards the discovery of novel agents for pest control. The ability of the AlphaFold algorithm to predict the structure of major targets for insecticides, *e.g.*, AChE, has been first assessed. This strategy overcomes the limitations imposed by the lack of empirical crystallographic data. The use of new metaheuristic algorithms enabled by the power of GPUs opened the door to a more exploratory and comprehensive approach to virtual detection compared to the standard VS workflows. The application of these advanced computational techniques leads to the identification of promising AChE inhibitors, which may offer new avenues for controlling pest populations, specifically the American cockroach. The compound Nadide, among others identified in this study, emerges as a particularly notable candidate due to its activity in multiple modes of action, suggesting potential for broad-spectrum efficacy.

The effort to refine the interface of AI and HPC will continue to be the cornerstone of our research. Next steps will be focused on the development of multi-target AI-based models for VS processes. Such models would leverage the insights gained in the current study and integrate them into an even more sophisticated framework capable of identifying compounds with multifaceted modes of action. It is expected that this will improve the efficiency of the screening process and also provides new potential bioactive compounds. Furthermore, we aim to advance from computational modeling to empirical validation through *in vivo* testing of the identified compounds. These animal assays will help us to corroborate VS predictions as well as to evaluate the real-world efficacy and safety profiles of proposed compounds, bridging the gap between computational predictions and tangible impact.

Data availability

The code for METADOCK can be found at **https://Baldoimbernon@bitbucket.org/Baldoimbernon/metadock_2.git**. The version of the code employed for this study is version [METADOCK 2.0].

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work has been funded by the Ramon y Cajal Grant num. RYC2018-025580-I, funded by MCIN/AEI/10.13039/501100011033 and by FSE invest in your future. Bio Logic Crop Science, S.L. and UCAM industrial doctorate program have also participated in the financing of this work. Funding for open access charge: CRUE-Universitat Politècnica de València.

References

1 H. Ranson, C. Claudianos, F. Ortelli, C. Abgrall, J. Hemingway, M. V. Sharakhova, M. F. Unger, F. H. Collins and R. Feyereisen, *Science*, 2002, **298**, 179–181.
2 D. G. Heckel, *Science*, 2012, **337**, 1612–1614.
3 J. Hemingway and H. Ranson, *Annu. Rev. Entomol.*, 2000, **45**, 371–391.
4 J. P. Cerón-Carrasco, D. Jacquemin, J. Graton, S. Thany and J.-Y. Le Questel, *J. Phys. Chem. B*, 2013, **117**, 3944–3953.
5 J. S. Raj and J. V. Ananthi, *J. Inf. Technol.*, 2019, **1**, 38–47.
6 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, *et al.*, *Nature*, 2021, **596**, 583–589.
7 J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, *et al.*, *Proteins: Struct., Funct., Bioinf.*, 2021, **89**, 1711–1721.
8 W. Ma, S. Zhang, Z. Li, M. Jiang, S. Wang, W. Lu, X. Bi, H. Jiang, H. Zhang and Z. Wei, *J. Chem. Inf. Model.*, 2022, **62**, 4008–4017.
9 B. Imbernón, J. Prades, D. Giménez, J. M. Cecilia and F. Silla, *Future Gener. Comput. Syst.*, 2018, **79**, 26–37.
10 J. Drews, *Science*, 2000, **287**, 1960–1964.
11 U. Rester, *Curr. Opin. Drug Discovery Dev.*, 2008, **11**, 559–568.
12 J. M. Rollinger, H. Stuppner and T. Langer, *Natural Compounds as Drugs*, Springer, 2008, vol. I, pp. 211–249.
13 G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson, *J. Comput. Chem.*, 1998, **19**, 1639–1662.
14 R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *J. Med. Chem.*, 2004, **47**, 1739–1749.
15 S. McIntosh-Smith, J. Price, R. B. Sessions and A. A. Ibarra, *Int. J. High Perform. Comput. Appl.*, 2015, **29**, 119–134.
16 T. J. Ewing, S. Makino, A. G. Skillman and I. D. Kuntz, *J. Comput.-Aided Mol. Des.*, 2001, **15**, 411–428.
17 I. Sánchez-Linares, H. Pérez-Sánchez, J. M. Cecilia and J. M. García, *BMC Bioinf.*, 2012, **13**, 1–14.
18 B. Imbernón, J. M. Cecilia, H. E. Pérez and D. Giménez, *Int. J. High Perform. Comput. Appl.*, 2017, 1–15.

19 B. Imbernón, A. Serrano, A. Bueno-Crespo, J. L. Abellán, H. Pérez-Sánchez and J. M. Cecilia, *Bioinformatics*, 2021, **37**, 1515–1520.
20 G. Schneider, *Drug Discovery Today*, 2002, **7**, 64–70.
21 A. N. Jain, *Curr. Protein Pept. Sci.*, 2006, **7**, 407–420.
22 Z. M. Wolfe and M. E. Scharf, *Sci. Rep.*, 2021, **11**, 24196.
23 P. Rezende-Teixeira, R. G. Dusi, P. C. Jimenez, L. S. Espindola and L. V. Costa-Lotufo, *Environ. Pollut.*, 2022, **300**, 118983.
24 V. A. Ingham, L. Grigoraki and H. Ranson, *Entomol. Gen.*, 2023, **43**, 515–526.
25 D. S. Wishart, C. Knox, A. C. Guo, D. Cheng, S. Shrivastava, D. Tzur, B. Gautam and M. Hassanali, *Nucleic Acids Res.*, 2008, **36**, D901–D906.
26 R. da Silva Ramos, J. da Silva Costa, R. Campos Silva, G. Vilhena da Costa, A. Bruno Lobato Rodrigues, É. de Menezes Rabelo, R. Nonato Picanço Souto, C. Anthony Taft, C. H. Tomich de Paula da Silva, J. M. Campos Rosa, *et al.*, *Pharmaceuticals*, 2019, **12**, 20.
27 F. Gang, X. Li, C. Yang, L. Han, H. Qian, S. Wei, W. Wu and J. Zhang, *J. Agric. Food Chem.*, 2020, **68**, 11665–11671.
28 Y. Ding, S. Chen, H. Liu, T. Liu and Q. Yang, *J. Agric. Food Chem.*, 2023, **71**, 8769–8777.
29 B. S. Zhorov and K. Dong, *Insects*, 2022, **13**, 745.
30 A. Tahir, A. R. Siddiqi, A. Maryam, S. Chaitanya Vedithi and T. L. Blundell, *J. Biomol. Struct. Dyn.*, 2023, 1–14.
31 Y. Kong, C. Zhou, D. Tan, X. Xu, Z. Li and J. Cheng, *J. Agric. Food Chem.*, 2024, 5145–5152.
32 B. Chafer-Dolz, J. M. Cecilia, B. Imbernón, E. Núñez-Delicado, V. Casaa-Giner and J. P. Cerón-Carrasco, *New J. Chem.*, 2023, **47**, 17234–17243.
33 I. D. Kuntz, *Science*, 1992, **257**, 1078–1082.
34 J. Kopp and T. Schwede, *Nucleic Acids Res.*, 2004, **32**, D230–D234.
35 A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, *et al.*, *Nucleic Acids Res.*, 2018, **46**, W296–W303.
36 A. Fiser, R. K. G. Do and A. Šali, *Protein Sci.*, 2000, **9**, 1753–1773.
37 Q. Wang, A. A. Canutescu and R. L. Dunbrack Jr, *Nat. Protoc.*, 2008, **3**, 1832–1847.
38 E. W. Sayers, M. Cavanaugh, K. Clark, J. Ostell, K. D. Pruitt and I. Karsch-Mizrachi, *Nucleic Acids Res.*, 2019, **48**, D84–D86.
39 Q. Han, D. M. Wong, H. Robinson, H. Ding, P. C. H. Lam, M. M. Totrov, P. R. Carlier and J. Li, *Insect Sci.*, 2018, **25**, 721–724.
40 K. D. Singh, R. K. Labala, T. B. Devi, N. I. Singh, H. D. Chanu, S. Sougrakpam, B. S. Nameirakpam, D. Sahoo and Y. Rajashekar, *Sci. Rep.*, 2017, **7**, 12483.
41 A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. de Beer, C. Rempfer, L. Bordoli, R. Lepore and T. Schwede, *Nucleic Acids Res.*, 2018, **46**, W296–W303.
42 R. Evans, M. O'Neill, A. Pritzel, N. Antropova, A. Senior, T. Green, A. Žídek, R. Bates, S. Blackwell, J. Yim, *et al.*, *bioRxiv*, 2021, preprint, DOI: **10.1101/2021.10.04.463034**.
43 M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov and M. Steinegger, *Nat. Methods*, 2022, **19**, 679–682.
44 E. Callaway, *Nature*, 2020, **588**, 203–205.
45 T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard and J. L. Banks, *J. Med. Chem.*, 2004, **47**, 1750–1759.
46 R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, *J. Med. Chem.*, 2004, **47**, 1739–1749.
47 R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin and D. T. Mainz, *J. Med. Chem.*, 2006, **49**, 6177–6196.
48 T. A. Halgren, *J. Chem. Inf. Model.*, 2009, **49**, 377–389.
49 G. Rozenberg, T. Bck and J. N. Kok, *Handbook of Natural Computing*, Springer, 2011.
50 K. Sörensen, *Int. Trans. Oper. Res.*, 2015, **22**, 3–18.
51 A. Llanes, A. Muñoz, A. Bueno-Crespo, T. García-Valverde, A. Sánchez, F. Arcas-Túnez, H. Pérez-Sánchez and J. M. Cecilia, *Curr. Drug Targets*, 2016, **17**, 1626–1648.
52 M. P. Jacobson, R. A. Friesner, Z. Xiang and B. Honig, *J. Mol. Biol.*, 2002, **320**, 597–608.
53 M. P. Jacobson, D. L. Pincus, C. S. Rapp, T. J. Day, B. Honig, D. E. Shaw and R. A. Friesner, *Proteins: Struct., Funct., Bioinf.*, 2004, **55**, 351–367.
54 P.-C. Su, C.-C. Tsai, S. Mehboob, K. E. Hevener and M. E. Johnson, *J. Comput. Chem.*, 2015, **36**, 1859–1873.
55 S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, *et al.*, *Nucleic Acids Res.*, 2021, **49**, D1388–D1395.
56 J. Wang, *J. Chem. Inf. Model.*, 2020, **60**, 1549–9596.
57 *Schrödinger Release 2021-3: LigPrep*, Schrödinger, LLC, New York, NY, 2017.
58 Z. Alamiddine, S. Thany, J. Graton and J.-Y. Le Questel, *ChemPhysChem*, 2018, **19**, 3069–3083.
59 Z. Alamiddine, B. Selvam, J. Graton, A. D. Laurent, E. Landagaray, J. Lebreton, M. Mathé-Allainmat, S. H. Thany and J.-Y. Le Questel, *J. Chem. Inf. Model.*, 2019, **59**, 3755–3769.
60 A. Cartereau, E. Taillebois, B. Selvam, C. Martin, J. Graton, J.-Y. Le Questel and S. H. Thany, *Front. Physiol.*, 2020, **11**, 481.
61 B. Selvam, E. Landagaray, A. Cartereau, A. D. Laurent, J. Graton, J. Lebreton, S. H. Thany, M. Mathé-Allainmat and J.-Y. Le Questel, *Bioorg. Med. Chem. Lett.*, 2023, **80**, 129124.
62 Z. Bouchouireb, S. H. Thany and J.-Y. Le Questel, *J. Comput. Chem.*, 2024, **45**, 377–391.