Electronic Supplementary Material (ESI) for RSC Advances.

Appendix A: Supplementary Information

Predicting the mixture effects of three pesticides by integrating the molecular simulation with the concentration addition

Fu Chen [a], Shu-Shen Liu [a,b,*], Xin-Tian Duan [a], and Qian-Fen Xiao [a,b]

[a] Key Laboratory of Yangtze River Water Environment, Ministry of Education, College of Environmental Science and Engineering, Tongji University, Shanghai 200092, PR China

[b] State Key Laboratory of Pollution Control and Resource Reuse, College of Environmental Science and Engineering, Tongji University, Shanghai 200092, PR China

[*] Corresponding author: Prof. Shu-Shen Liu

College of Environmental Science and Engineering, Tongji University,

1239 Siping Road, Shanghai 200092, P. R. China

Tel: 86-021-65982767

E-mail address: ssliuhl@263.net (S.-S. Liu)

Sequential docking protocol

Sequential docking protocol was carried out to predict the locations of two ligands within a given binding site. In the first step of the sequential docking, pesticide A was individually docked into binding site with the same parameters described in Molecular docking and molecular dynamics simulation section. In the second step, pesticide B was docked into the complex of LUC with bound pesticide A, whose docking box expanded into 16×20×20 Å. The last step was selecting the pose of pesticide B with the most negative score. The molecular modelling of A-B-LUC was built. The results were shown in Fig. S6. As can be seen from Fig. S6, the former BAY occupies the LH_2 pocket, while the latter pesticide (BAY, MET, or VEL) just occupies the pocket of AMP that is adjacent to LH_2 pocket but separate. This phenomenon is found in other ternary complex. Furthermore, comparing the two complexes having the same second pesticide, it is clear that the RMSD of second pesticide is small. It is deduced that the former bound ligand does not influence the binding site of the second ligand.

Molecular dynamics procedure

The structural parameters of pesticide ligand and LUC protein were prepared by using antechamber module and tleap module in AMBER 12 software package. The atom types and atomic charges were assigned by the AM1-BCC[1] model in the antechamber module and the missing bonded parameters were gained from the generalized AMBER force field (GAFF)[2, 3] for the pesticide ligand. The ff99SB force field[4] was selected for the LUC protein. The tleap module was used to correct all missing hydrogen atoms of both the protein and pesticide. The pesticide-LUC complex was solvated by the TIP3P waters, with a minimum distance of 8.5 Å from the complex surface. Thus, the total number of atoms were 42672, 42665, and 42680 for the BAY-LUC, MET-LUC, and VEL-LUC systems, respectively. All three systems having a total charge of -7 were neutralized by adding seven Na^+ ions.

MD was conducted by the standard procedure, which comprises energy-minimization, gradual heating of the systems, and isothermal isobaric ensemble (NPT) molecular dynamics. To reduce bad steric interaction, the solvated system was minimized by three steps[3]: The steepest descent (2000 steps) and conjugate gradient (1000 steps) minimizations were firstly carried out on the water molecules alone, then on the backbone of complex and water system, and finally on the whole system. The equilibration phase and production phase were obtained by using the PMEMD[5] module in AMBER. The periodic boundary with NPT (constant particle, pressure, and temperature) ensemble at 1 atm was applied. The SHAKE method was used to provide an integration time step of 2 fs while keeping all bonds to the hydrogen atoms rigid. The cutoff distance for the long-range van der Waals was set as 10 Å. The long-range Coulombic interactions were handled using the particle mesh Ewald (PME) method. The whole system was gradually heated from 0 to 300 K in 50 ps and subsequently simulated at 300 K for equilibration and production phases. The whole system was equilibrated for 4 ns, and the MDs were continued for another 4 ns. One hundred snapshots of the simulated structures within the last 1 ns with a step of 10 ps were sampled.

Simulation system was monitored through the convergences of energy and temperature[3]. To verify the system stability and measure protein dynamics, the ptraj module was used to analyze the root mean-square displacements (RMSD) between the trajectory structures and the first snapshot structure in 1th ns trajectory.

Binding free energy calculation

The binding free energy (ΔG_i) of single pesticide bound to LUC was computed from the free energy of the complex ($G_{complex}$), protein ($G_{protein}$), and ligand (G_{ligand}),

$$\Delta G_i = G_{complex} - \left(G_{protein} + G_{ligand}\right) \quad (1)$$

On the other hand, ΔG_i can be expressed as follows,

$$\Delta G_i = \Delta H - T\Delta S \approx \Delta E_{MM} + \Delta G_{sol} - T\Delta S \quad (2)$$

where ΔH is the enthalpy, ΔE_{MM} is the molecular mechanics free energy, ΔG_{sol} is the solvation free energy, and $T\Delta S$ represents the entropy term.

The molecular mechanics free energy was calculated as follows:

$$\Delta E_{MM} = \Delta E_{ele} + \Delta E_{vdw} \quad (3)$$

where ΔE_{ele} and ΔE_{vdw} represent the Coulomb and van der Waals interactions, respectively.

The solvation free energy was composed of polar and nonpolar components:

$$\Delta G_{sol} = \Delta G_{ele,sol} + \Delta G_{nonpolar,sol} \quad (4)$$

where $\Delta G_{ele,sol}$ is the polar contribution to solvation and $\Delta G_{nonpolar,sol}$ is the nonpolar solvation term. The $\Delta G_{ele,sol}$ was calculated by using the generalized Born (GB) model. The $\Delta G_{nonpolar,sol}$ was computed based on solvent accessible surface area (SASA). The bigger absolute value of ΔG negative value means the more binding ability.

Entropy contributions ($T\Delta S$) arising from changes in the translational, rotational and vibrational degrees of freedom was

calculated using normal-mode analysis[6] by the nmode program in AMBER 12. The normal-mode analysis is high computationally demanding, so that $T\Delta S$ was averaged over only 20 snapshots of the last 1 ns MD trajectory.[7, 8] Before the calculation, each snapshot was optimized for 50000 steps using a distance-dependent dielectric of $4r_{ij}$ (r_{ij} is the distance between atoms i and j) until the root-mean-square deviation of the gradient vector was less than 0.0001 kcal mol^{-1} $Å^{-2}$.[8]

Ligand-residue interaction decomposition

The interaction between each residue of luciferase and pesticide were computed using the MM/GBSA approach applied in AMBER 12. This decomposition was performed only for molecular mechanics and solvation energies but not for entropies.[9] The binding free energy of each complex was decomposed per residue including four terms: van der Waals contribution (ΔE_{vdW}), electrostatic contribution (ΔE_{ele}), polar solvation contribution ($\Delta G_{ele,sol}$), and nonpolar solvation contribution ($\Delta G_{nonpol,sol}$):

$$\Delta G_r = \Delta E_{vdW} + \Delta E_{ele} + \Delta G_{ele,sol} + \Delta G_{nonpol,sol} \quad (5)$$

where ΔE_{ele} and ΔE_{vdw} represent the Coulomb and van der Waals interactions, respectively. $\Delta G_{ele,sol}$ is the polar contribution to solvation and $\Delta G_{nonpolar,sol}$ is the nonpolar solvation term. ΔG_r is pesticide-residue pair energy between each pesticide and each individual residue

The microplate toxicity analysis procedure.

The toxicity of a pesticide or mixture was expressed as a percentage inhibition (E or x), which was calculated as follows:

$$E = x = \left(1 - \frac{L}{L_0}\right) \times 100\% \quad (6)$$

where L_0 is an average RLU of the controls (12 parallels), L is an average RLU of the treatments (3 parallels).

To quantitatively describe the toxicities of various concentrations, the observed concentration-response datas were fitted to nonlinear function, Weibull, called concentration response curve fitting (CRC)[10],

$$E = 1 - \exp\left(-\exp\left(\alpha + \beta \log_{10}(c)\right)\right) \quad (7)$$

where α and β are the parameters to be estimated, c is the concentration of test pesticide or mixture. The regression analysis was performed using nonlinear least-squares fit. The coefficient of determination (R^2) and the root-mean-square error (RMSE) were essential to evaluate the fitting. As a quantitative measure of the uncertainty, the 95% observation-based confidence interval was also determined.[11]

References

1. M. Han and J. Z. H. Zhang, *J. Chem. Inf. Model.*, 2010, **50**, 136-145.
2. J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157-1174.
3. W. Khuntawee, T. Rungrotmongkol and S. Hannongbua, *J. Chem. Inf. Model.*, 2012, **52**, 76-83.
4. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins: Struct., Funct., Bioinf.*, 2006, **65**, 712-725.
5. D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668-1688.
6. J. Kottalam and D. A. Case, *Biopolymers*, 1990, **29**, 1409-1421.
7. R. Cao, Y. Jin and D. Xu, *J. Phys. Chem. B*, 2012, **116**, 6087-6096.
8. H. Zhang, Y. Yao, H. Yang, X. Wang, Z. Kang, Y. Li, G. Li and Y. Wang, *Insect Biochem. Mol. Biol.*, 2012, **42**, 583-595.
9. Y. Yang, J. Qin, H. Liu and X. Yao, *J. Chem. Inf. Model.*, 2011, **51**, 680-692.
10. M. Scholze, W. Boedeker, M. Faust, T. Backhaus, R. Altenburger and L. H. Grimme, *Environ. Toxicol. Chem.*, 2001, **20**, 448-457.
11. X. W. Zhu, S. S. Liu, H. L. Ge and Y. Liu, *Water Res.*, 2009, **43**, 1731-1739.

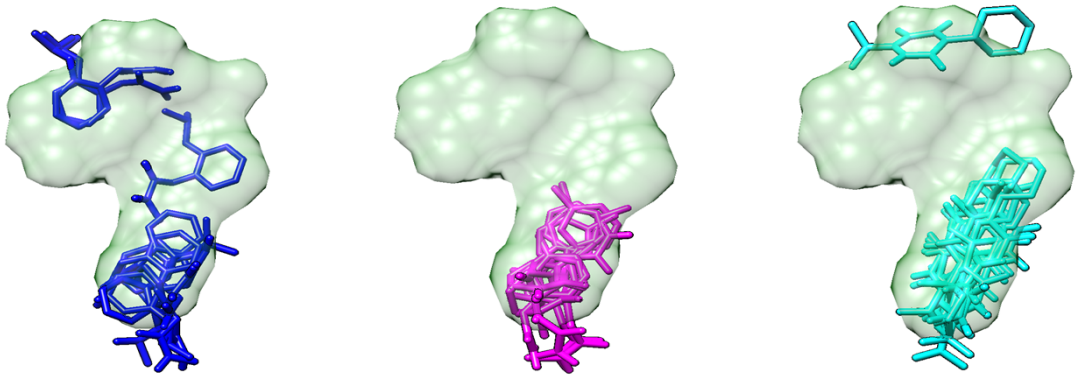
Table S1. The fitted concentration-response curve model parameters (α and β), statistics (RMSE and R^2), EC_{50} (10^{-3} mol/L), and binding free energy (ΔG_i, kcal/mol) of three pesticides

Chemicals	α	β	RMSE	R^2	EC_{50}	ΔG_i	EC_{50}[a]
BAY	5.36	2.02	0.019	0.993	1.463	-13.0	1.442
MET	4.75	1.99	0.011	0.998	2.692	-5.7	2.682
VEL	8.57	3.02	0.017	0.997	1.100	-18.4	1.110

[a]: the average value of the EC_{50}s estimated from two linear models (Eqs. 3 - 5)

Table S2. Binding free energy (kcal/mol) and its components for the three pesticide-LUC systems

Component	BAY-LUC	MET-LUC	VEL-LUC
ΔE_{vdw}	-37.6	-26.7	-39.2
ΔE_{ele}	-11.3	-13.6	-29.0
ΔE_{MM}	-48.9	-40.3	-68.2
$\Delta G_{ele,sol}$	20.3	20.5	33.7
$\Delta G_{nonpolar,sol}$	-4.4	-3.9	-5.0
ΔG_{sol}	15.8	16.6	28.7
$\Delta E_{vdw}+\Delta G_{nonpolar,sol}$	-42.0	-30.6	-44.2
$\Delta E_{ele}+\Delta G_{ele,sol}$	9.0	6.9	4.7
ΔH_i	-33.1	-23.7	-39.5
$T\Delta S_{Translational}$	-12.4	-12.2	-12.6
$T\Delta S_{Rotational}$	-9.6	-9.2	-9.8
$T\Delta S_{Vibrational}$	1.9	3.4	1.4
$T\Delta S$	-20.1	-18.0	-21.1
ΔG_i	-13.0	-5.7	-18.4
EC_{50}(mol/L)	1.463E-03	2.692E-03	1.100E-03



BAY-LUC MET-LUC VEL-LUC

Fig. S1. The 10 best scoring poses of pesticides bound to LUC.

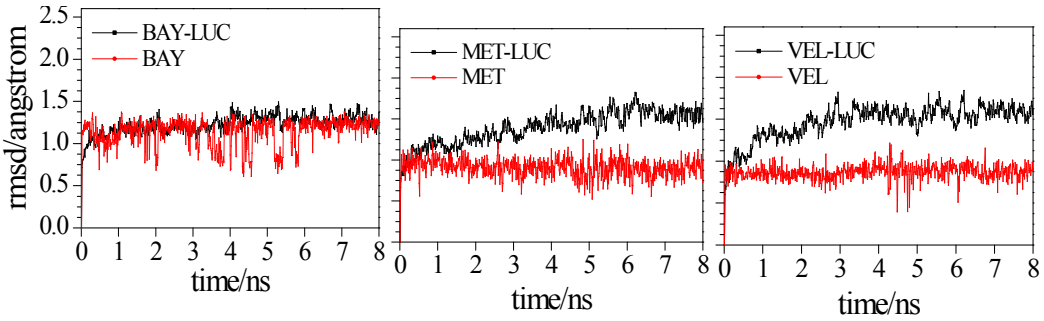


Fig. S2. Plots of root mean-square deviation (RMSD) vs. time for the simulated systems of BAY-LUC, MET- LUC, and VEL- LUC

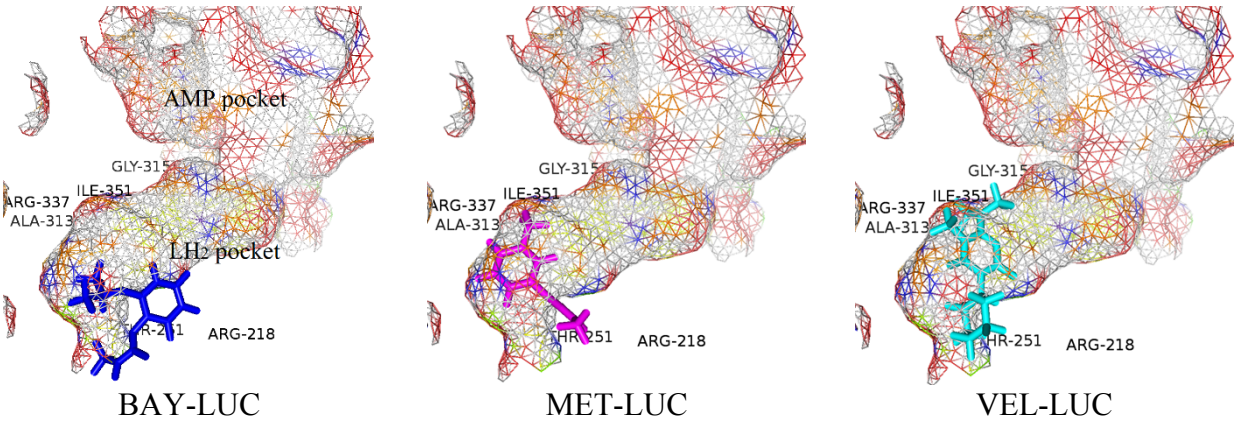


Fig. S3. The binding sites of three pesticides in the LUC pocket where the stick shown in blue, magenta, and cyan represent BAY, MET, and VEL, respectively.

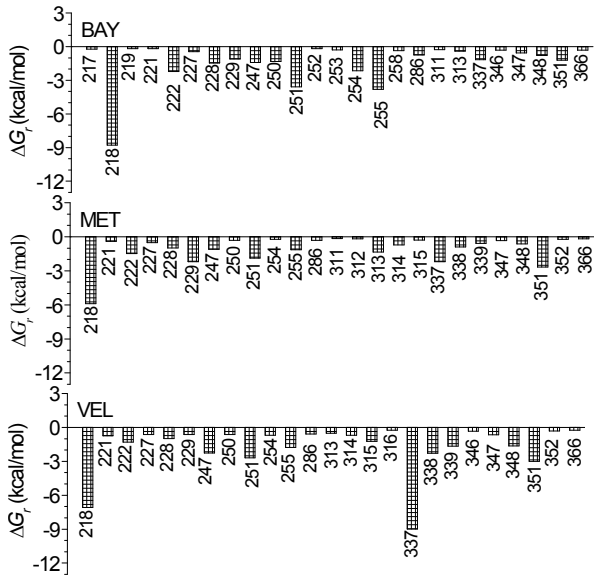


Fig. S4. Decomposition of the binding free energy on a residue-based (ΔG_r) for key residues in the pesticide-LUC.

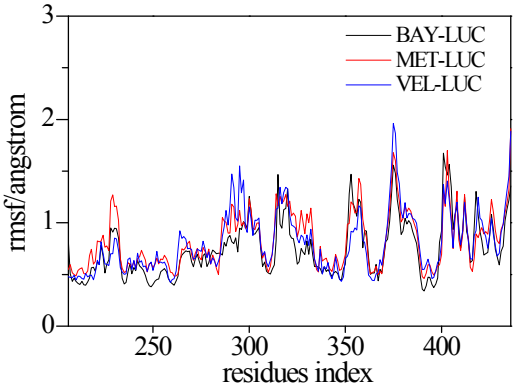


Fig. S5. Plots of root mean-square fluctuation (RMSF) vs. residue index for the simulated systems.

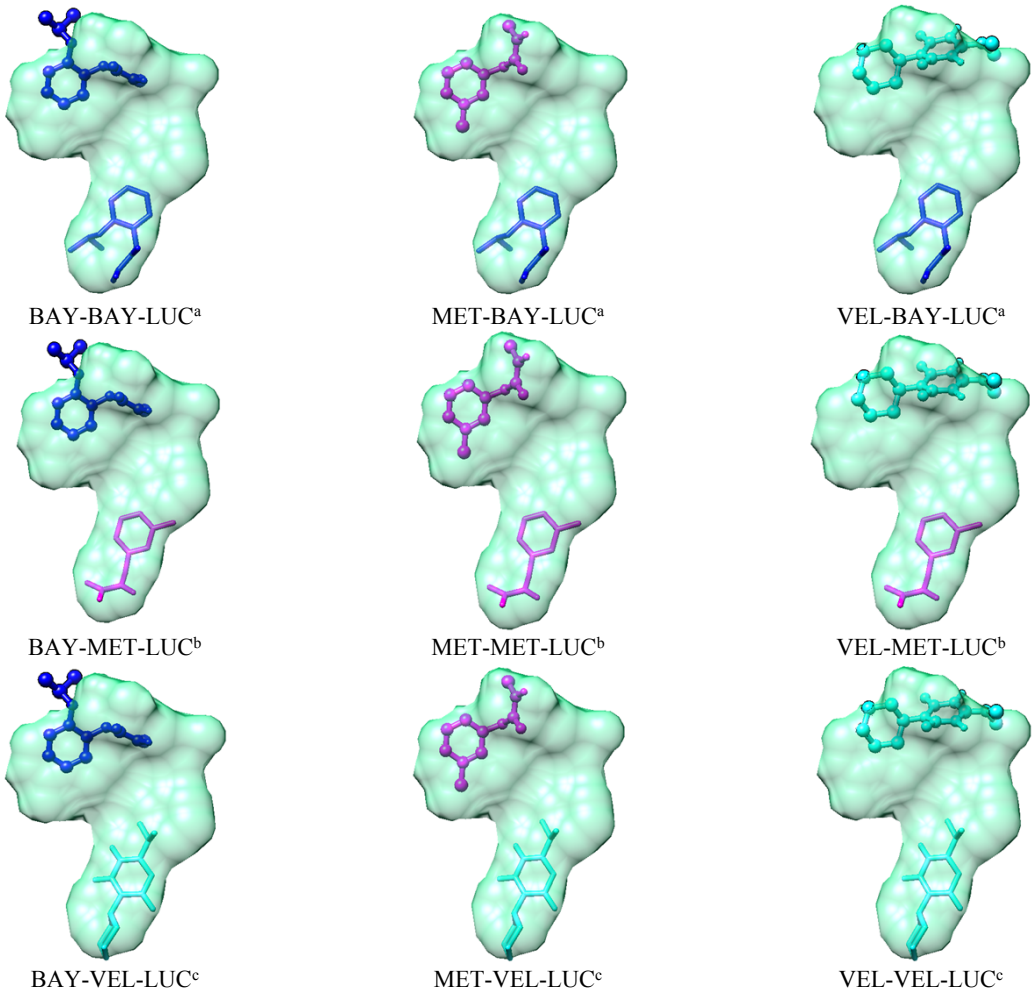


Fig. S6. sequential simulations. The former pesticide is shown as stick models, the latter pesticide is shown as ball and stick models. [a] sequential docking based on BAY-LUC, [b] sequential docking based on MET-LUC, [c] sequential docking based on VEL-LUC.