

# Structural determinants of imidacloprid-based nicotinic acetylcholine receptor inhibitors identified using 3D-QSAR, docking and molecular dynamics

Qinfan Li · Xiangya Kong · Zhengtao Xiao ·  
Lihui Zhang · Fangfang Wang · Hong Zhang · Yan Li ·  
Yonghua Wang

Received: 24 July 2011 / Accepted: 24 October 2011 / Published online: 26 November 2011  
© Springer-Verlag 2011

**Abstract** Nicotinic acetylcholine receptor (nAChR) is a target for insect-selective neonicotinoid insecticides (NNs), exemplified by imidacloprid (IMI). In the present study, 78 IMI derivatives reported as inhibitors of *Drosophila melanogaster* nAChR (*Dm*-nAChR) and *Musca domestica* nAChR (*Md*-nAChR) were used for three-dimensional quantitative structure–activity relationship (3D-QSAR) studies. Two optimal models with good predictive power were obtained:  $Q^2=0.64$ ,  $R^2_{\text{pred}}=0.72$  for *Dm*-nAChR, and  $Q^2=0.63$ ,  $R^2_{\text{pred}}=0.62$  for *Md*-nAChR. In addition, homology modeling, molecular dynamic (MD) simulation, and molecular docking also showed that amino acids located within loops A, C, D and E play key roles in the interaction of *Dm*-/*Md*-nAChR with NNs. This is highly consistent with the results of graphical analysis of 3D-QSAR contour

plots. Mutation analysis also implicates the Y/S mutation within loop B as being associated closely with NN resistance in *Drosophila* and *Musca*. The results obtained lead to a better understanding not only of interactions between these antagonists and *Dm*-/*Md*-nAChR, but also of the essential features that should be considered when designing novel inhibitors with desired activities.

**Keywords** Nicotinic acetylcholine receptor · Neonicotinoid · 3D-QSAR · Mutation · MD simulation · Molecular docking

## Introduction

The nicotinic acetylcholine receptor (nAChR) is an agonist-controlled cation channel that acts in fast neurotransmission at cholinergic synapses in vertebrates and invertebrates [1]. It belongs to the Cys-loop ligand-gated ion channel superfamily and forms pentamers of either a single type of subunit or different types of homologous subunits, each of which consists of a conserved extracellular N-terminal ligand-binding domain (LBD), four transmembrane helices and a C-terminus facing the extracellular space [2, 3]. The LBD, which is around 210 residues long and contains six loops (A–F), makes up the ligand-binding sites for agonists and competitive antagonists [4, 5]. Its significant role in mediating rapid chemical transmission of signals has led to the development of insecticides targeting this receptor, such as nicotine and neonicotinoid insecticides (NNs) [6–8].

Nicotine is a nAChR agonist but with limited insecticidal efficacy and spectrum, in addition to posing rather a high risk to people, and is now obsolescent [9]. Therefore, several NNs have been developed as insect-selective

Qinfan Li and Xiangya Kong contribute equally

**Electronic supplementary material** The online version of this article (doi:10.1007/s00894-011-1293-z) contains supplementary material, which is available to authorized users.

Q. Li · X. Kong · L. Zhang  
College of Veterinary Medicine, Northwest A & F University,  
Yangling 712100 Shaanxi, China

Z. Xiao · F. Wang · H. Zhang · Y. Wang  
College of Life Science, Northwest A & F University,  
Yangling 712100 Shaanxi, China

Y. Li  
Department of Materials Science and Chemical Engineering,  
Dalian University of Technology,  
Dalian 116023 Liaoning, China

Y. Wang (✉)  
Center of Bioinformatics, Northwest A & F University,  
Yangling 712100 Shaanxi, China  
e-mail: yh\_wang@nwsuaf.edu.cn

nAChR agonists with greatly improved effectiveness for pest management and that act selectively on insect nAChRs [10]. Further analysis postulates that either a nitro or a cyano group of NNs contributes directly to their selectivity [10–12]. Imidacloprid (IMI)—the first member of the NN class of insecticides—was patented in 1985 by Bayer and first marketed in 1991 [13]. Subsequently, some other NNs have also been developed and brought to the market, including nitenpyram (1995—Takeda), acetamiprid (1996—Nippon Soda), thiamethoxam (1998—Syngenta), thiacloprid (2000—Bayer), clothianidin (2002—Takeda and Bayer) and dinotefuran (2002—Mitsui), which are now estimated to have annual worldwide sales of  $\approx$  US \$1 billion [14, 15].

However, NNs, like many systemic insecticides, display prolonged persistence in the environment, which is likely to generate high selection pressure for resistance [16, 17]. Evidence has shown that the resistance is attributed to enhanced oxidative detoxification of NNs by overexpressed cytochrome P450-dependent monooxygenases [18, 19]. Recently, a study of *Nilaparvata lugens* (brown planthopper)—a major rice pest in Asia—has identified a target-site mutation (Y151S) within nAChR  $\alpha$  subunits associated with NNs resistance by radioligand binding assays [20]. Therefore, to avoid the problem of insect resistance to NNs, the continual introduction of new insect control chemicals is urgently needed.

Numerous biochemical studies have shown that the binding site of nAChR is formed by  $\alpha$ -subunit residues contributing to the so-called ‘loops’ A, B and C, whereas the neighboring subunit residues contribute to ‘loops’ D, E and F [21, 22]. These works provide a preliminary description of the agonist binding site, but little is known about the precise binding mode of the ligand within this receptor. The crystal structure of *Aplysia californica* ACh binding protein (*Ac*-AChBP) with homology to the LBD of nAChR  $\alpha$ -subunit was published in 2008 [23, 24]. *Ac*-AChBP is of neither mammalian nor insect origin, but opens the way to the building of a realistic structural model of insect nAChR through in silico modeling, thus providing the means with which to make major breakthrough in the study of the binding interaction between NNs and insect nAChR. Clearly, in vitro assessment remains a labor-intensive and time-consuming operation. Thus, more efficient and economical alternative methods should be employed, such as the in silico molecular modeling approach that is used to predict and prioritize chemicals for subsequent in vitro and in vivo screening. However, only a few computational works to date have used three-dimensional quantitative structure-activity relationship (3D-QSAR) methods and comparative molecular field analysis (CoMFA) on a small data set of 17–25 NNs. These latter studies showed that the electrostatic potential, steric

potential and permeability coefficient are important parameters for the design of new pesticides [25–27].

This work uses the in silico approach for the first time to address the following two questions: (1) what is the binding mechanism of NNs with *Dm/Md*-nAChR? and (2) what are the optimal structural determinants for these novel potent inhibitors. For this purpose, a total of 78 IMI analogs [28–31] were collected and analyzed using theoretical computations, including 3D-QSAR models by CoMFA and comparative molecular similarity indices analysis (CoMSIA) [32]. In addition, homology modeling, mutation analysis, molecular dynamic (MD) simulation, and molecular docking were also performed to further probe the structural properties and probable binding modes of these inhibitors at the allosteric sites of *Dm/Md*-nAChR. The application of all these methods to these inhibitors not only helps to better understand ligand–receptor interactions but also provides useful and rational suggestions for the further design of novel insect nAChR inhibitors.

## Materials and methods

### Data sets

The total of 78 IMI analogues included two datasets, i.e., set-I [28, 29] (43 molecules) with  $K_i$  data from *Drosophila melanogaster* nAChR (*Dm*-nAChR) assays, and set-II [30, 31] (41 molecules) expressed by  $IC_{50}$  to *Musca domestica* nAChR (*Md*-nAChR). All bioactivity data were converted into  $pK_i$  or  $pIC_{50}$  values, respectively. The molecules in the test sets were selected in such a way that their  $pK_i$  or  $pIC_{50}$  values were distributed uniformly in the range of values for the whole sets, so that the predictive power of the model could be evaluated effectively. The structures of the two groups and their activity values are given in the supporting Tables S1–S2 with the test sets marked by <sup>a</sup>.

### Molecular modeling and alignment

Molecular alignment is a crucial step in CoMFA/CoMSIA studies [32]. In the modeling process, the 3D structures of all molecules were first constructed using the Sketch Molecule function in SYBYL 6.9 (Tripos, St. Louis, MO). Geometry optimization of the 3D structures was carried out using the Tripos force field with the Gasteiger-Marsili charges, and repeated minimization was performed using the Powell conjugate gradient method with a convergence criterion of  $0.05 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ .

In this work, two alignment rules were applied. The first, denoted by ligand-based alignment, selected the most potent inhibitor for each class (compounds 31, 72, respectively) as the template. All inhibitors for each class

in the data set were then aligned to a common substructure (shown in Figs. 1a, 2a) of the template using the “align database” command in SYBYL. The second was structure-based alignment, where all molecules were docked into the receptors, and the top scored conformations were then used for further analysis. Figures 1 and 2 show the aligned conformations derived from the two methods.

### 3D-QSAR studies

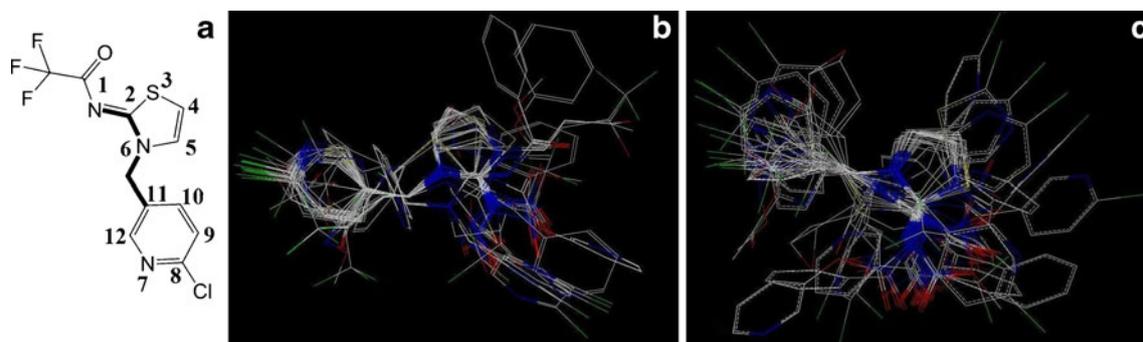
In order to search for ligands with higher affinity, it is helpful to establish a reasonable 3D-QSAR model to analyze the effect of the molecular field on the activities of nAChR inhibitors [32]. A 3D cubic lattice with grid spacing of 2 Å in x, y, and z directions was generated automatically to encompass the aligned molecules and to derive the CoMFA/ CoMSIA descriptor field. CoMFA descriptors steric (S) and electrostatic (E) were calculated using a  $sp^3$  carbon probe atom with a van der Waals radius of 1.52 Å and a charge of +1.0, with a distance dependent dielectric at each lattice point. The steric and electrostatic energy values in CoMFA were truncated at 50 kcal mol<sup>-1</sup>. The CoMSIA method defined explicit hydrophobic (H) and hydrogen-bond (H-bond) donor (D), and acceptor (A) descriptors in addition to the S and E field descriptors used in CoMFA. A  $sp^3$  carbon probe atom with a charge of +1.0, a radius of 1.0 Å, hydrophobicity +1.0, and H-bond donor and acceptor property +1 was used to calculate the CoMSIA respective fields. When building the CoMSIA models, as the five descriptors were not completely independent of each other and may reduce the significance of the model, all 31 possible descriptor combinations were calculated. The attenuation factor was set to 0.3 as default.

To select the best model, a cross-validated coefficient  $Q^2$  and the optimum number of components  $N_c$  was determined by the leave-one-out (LOO) cross-validation method of partial least square algorithm.  $N_c$  was then used to derive the non-cross validated model. The Pearson coefficient  $R^2_{nev}$  and standard error of estimate SEE was subsequently

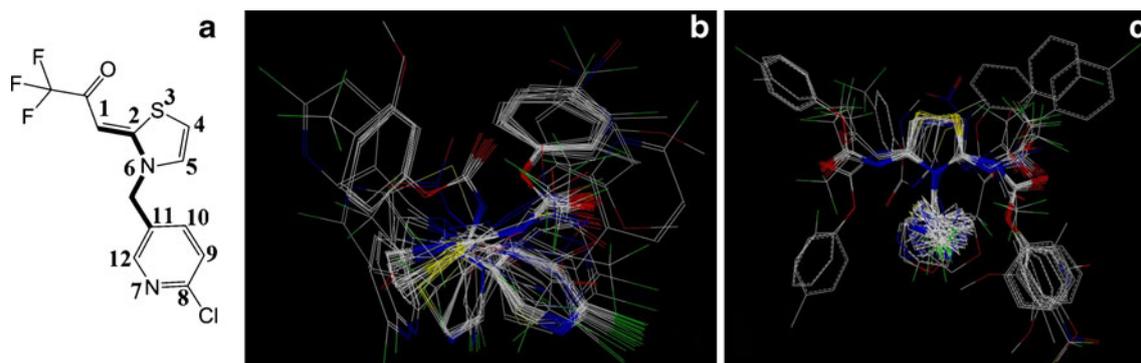
obtained. The predictive ability of the model was expressed by the predictive correlation coefficient  $R_{pred}^2$  of an external test set [32]. Finally, the optimum CoMFA/ CoMSIA results were represented graphically by field contour maps, where the coefficients were generated using the field type “StDev\*Coeff”.

### Homology modeling

Structural information about target proteins and ligands that bind to them specifically are of utmost importance for the rational design of new drugs. Due to the unavailability of the crystal structure of insect nAChR, homology modeling of the protein structure from its primary sequence was performed. The template protein *Ac*-AChBP (PDB ID 3 C79 chains A and B, resolution 2.48 Å) from the Brookhaven Protein Database (<http://www.pdb.org/pdb/home/home.do>) identified by Blast Search (<http://www.ncbi.nlm.nih.gov/BLAS>) was employed here for the construction of the target proteins [33]. The initial alignment of the target and template sequences was carried out using the ClustalW program [34]. The sequence identity between *Dm*-nAChR (GenBank ID NP\_995708 residues from 306 to 520) and *Ac*-AChBP was 32%, and that between *Md*-nAChR (GenBank ID ABJ09672.1 residues from 43 to 232) and *Ac*-AChBP was 30%, but the active site (loop A–F) identity was up to 51% and 49%, respectively (Fig. 3). Homology modeling was performed using SWISS-MODEL (Automated Comparative Protein Modeling Server, Version 3.5, Glaxo Wellcome Experiment Research, Geneva, Switzerland) [35]. In addition, the quality of the resulting homology structure was assessed using the protein structure verification WHAT-CHECK [36] module of the WHATIF on-line server (<http://swift.cmbi.kun.nl/WIWWWI/modcheck.html>). For molecule docking purposes, all hydrogen atoms were added subsequently to the unoccupied valence of heavy atoms at the neutral state (pH 7.0) using the biopolymer module of SYBYL package.



**Fig. 1** a–c Superimposition of the set-I compounds in the training and test sets. **a** Common substructure. **b**, **c** Ligand- and structure-based alignments, respectively



**Fig. 2** a–c Superimposition of the set-II compounds in the training and test sets. **a** Common substructure. **b**, **c** Ligand- and structure-based alignments, respectively

## MD simulations

To gain a better relaxation and a correct arrangement of the atoms, as well as to refine the loop regions, MD simulations of the homology modeled *Dm*-nAChR and *Md*-nAChR were conducted using the Amber 10 package [37]. The amber99SB force field was employed to describe the protein parameters. The initial conformers were neutralized by adding sufficient  $\text{Na}^+$  counterions, and solvated in a rectangular box ( $73.21 \times 98.78 \times 97.42 \text{ \AA}^3$  and  $74.14 \times 99.79 \times 95.44 \text{ \AA}^3$ , respectively) of TIP3P water with a minimum solute-wall distance of  $10 \text{ \AA}$ . During the MD simulation runs, the cut-off distance for computing the nonbonded interactions was truncated at  $10 \text{ \AA}$ ; the SHAKE algorithm was applied to constrain all covalent bonds involving H-atoms; the Particle Mesh Ewald method was employed to calculate the long-range electrostatic interactions with default values.

Prior to MD simulations, every system was minimized energetically with the complex atoms constrained to eliminate possible bad contacts through 2,500 steepest descent steps and another 2,500 conjugate-gradient steps. Following that, MD simulations commence by heating up the systems to 300 K at

<i>Ac</i> -nAChR	DLVYYE[DR]RWKLSL[M]NDPNEYGNITDFRTSAAD[ITW]PDI[AV]SSTRP.V	51-99	
<i>Dm</i> -nAChR	VTNVMLKLEWNDMNLRWNTSDYGGVKDLRIPPHR[IKW]PDVLMYNSADEGF	360-409	
<i>Md</i> -nAChR	TTNAWLNLEWNDYNL[RW]NDSEYGGVKDLRITPNKLRKPDVLMYNSADEGF	73-122	
	Loop D	Loop A	
<i>Ac</i> -nAChR	QVLSPT[AVV]THDGSVMFI[PA]QRLSFMCDP.TGVDSEEGATCAVKFGSWV	100-148	
<i>Dm</i> -nAChR	DGTYQIN[VV]RNGSCLVYPPGIFKSTCKIDI[TFW]PFDDQRC[EM]KFGSWT	410-459	
<i>Md</i> -nAChR	DGTYHIN[VV]KHNGNCLVYPPGIFKSTCKIDI[TFW]PFDDQHC[EM]KFGSWT	123-172	
	Loop E	Loop B	
<i>Ac</i> -nAChR	*YSGFE[IDL]KTDDQ.VDLSYYASSKYEILSATQTRQVHYSCCPPEYI	149-196	
<i>Dm</i> -nAChR	YDGFQLDLQLQDETGGD[IS]SYVLNGEWELLGVPGRNEIYYNCPEPYI	460-508	
<i>Md</i> -nAChR	YDGNQLDLVLNSEDGGDLSDFITNGEWYL[I]AMPGKNTIYYNCPEPYV	173-221	
	Loop B	Loop F	Loop C

**Fig. 3** Multiple sequence alignments of *Aplysia californica* ACh binding protein (*Ac*-AChBP), *Drosophila melanogaster* nicotinic acetylcholine receptor (*Dm*-nAChR) and *Musca domestica* nAChR (*Md*-nAChR). Cyan color denotes amino acid residues conserved in individual columns. The six loops comprising the ligand binding domain are underlined. The location of the mutation residue in this study (equivalent to Y151 of the brown planthopper) is indicated by an asterisk. Key binding site residues are highlighted in black rectangles

a constant force  $2.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ . Then, a 50 ps of density equilibrated was applied at 300 K with the complex atoms constrained. After that, the system was equilibrated with a collision frequency of  $1 \text{ ps}^{-1}$  at a constant temperature by a Berendsen thermostat and pressure. Finally, the simulation was run and continued using a 2-fs time step, where the coordinates were saved every 10 ps for analysis. The total number of atoms in each simulation system was 63,826 and 57,952, respectively, including complex and water atoms.

## Molecular docking

In order to locate the appropriate binding orientations and conformations of IMI analogs in the *Dm*-/*Md*-nAChR binding pocket, and to develop structure-based 3D-QSAR models, molecular docking for the representative structures obtained from the MD simulation trajectories was carried out using the flexible docking module Surflex [38]. Our docking analysis was performed as follows: (1) the target protein structure was aligned with the template protein, then the cocrystallized ligand IMI and water molecules of *Ac*-AChBP were merged into the corresponding sites of the target protein structure, since the crystallized water was important in mediating the interactions between the ligand and the *Ac*-AChBP [28]. (2) Two parameters, i.e., specified 1\_0.2 of bloat and threshold, which determine how far a potential ligand should extend outside of the concavity and how deep into the protein, were applied to define the protomol. When the docking run was finished, it afforded the top ten docking poses of each ligand ranked by total scores using the Hammerhead scoring function [38].

## Results and discussion

### 3D-QSAR statistical analysis

Table 1 summarizes the optimum models derived from CoMFA/CoMSIA studies (see other combinations in

supporting Tables S3–S6). For space saving, only the optimal 3D-QSAR models derived from structure-based CoMSIA of set-I and ligand-based CoMFA of set-II are utilized for further discussion.

### Set-I

The statistical results of the optimal structure-based CoMSIA model for set-I ( $Q^2=0.64$ ,  $R_{ncv}^2=0.93$ ,  $F$  value=75.98, a small SEE=0.42 and high  $R_{pred}^2$  value=0.72) indicate that this model is robust. The steric, hydrophobic and H-bond donor fields contribute 18.8%, 64.0%, and 17.2% to total fields, respectively, demonstrating the major role that the hydrophobic property plays in ligand binding affinity. During the cross-validation procedure, compound 43 is detected as an outlier due to the large residual between the experimental and predicted values ( $> 1.0$  log unit). Further analysis suggests that this compound has no electron-withdrawing atom at position 3, which may be a prerequisite for forming an H-bond with its receptor. The plots of actual vs predicted  $pK_i$  values represent a uniform distribution around the regression line, indicating the satisfactory predictive capability and reliability of the model obtained (Fig. 4a).

### Set-II

The optimal ligand-based CoMFA model for set-II obtained had good internal predictive ability as illustrated by  $Q^2$  of 0.63,  $R_{ncv}^2$  of 0.96,  $F$  value of 118.82, a low SEE value of 0.51, and  $R_{pred}^2$  of 0.62 for the external test set. The steric and electrostatic contributions were 48.7% and 51.3%, respectively. Compound 65, which possesses a low inhibitory activity and is the only molecule to have a methyl group substituent at position -3, is regarded as an outlier. The correlation between the predicted and actual  $pIC_{50}$  values is shown in Fig. 4b. As can be seen from this figure, the data points are distributed uniformly along the regression line, which proves that the model is reasonable.

### 3D-QSAR contour maps

One of the greatest advantages of 3D-QSAR modeling is that the results can be visualized as 3D coefficient contour maps. These are helpful in identifying the important regions where changes in the S, E, H, D and A fields might affect biological activity and the possible interaction sites for the ligand with its target. To aid visualization, the potentially

**Table 1** Summary of statistical results of the optimal three-dimensional quantitative structure–activity relationship (3D-QSAR) models for each class

Parameters	Set-I				Set-II			
	Ligand-based		Receptor-based		Ligand-based		Receptor-based	
	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA	CoMFA	CoMSIA
$Q^2$ <sup>a</sup>	0.52	0.54	0.31	0.64	0.63	0.46	0.08	0.43
$R_{ncv}^2$ <sup>b</sup>	0.94	0.82	0.87	0.93	0.96	0.90	0.62	0.97
SEE <sup>c</sup>	0.40	0.66	0.56	0.42	0.51	0.45	0.84	0.27
$F$ <sup>d</sup>	72.27	46.32	100.56	75.98	118.82	83.42	26.00	145.79
$R_{pred}^2$ <sup>e</sup>	0.66	0.74	0.78	0.72	0.62	0.86	0.61	0.64
SEP <sup>f</sup>	1.14	1.06	1.28	1.07	0.91	1.06	1.36	1.23
$N_c$ <sup>g</sup>	6	3	2	5	5	3	2	5
Field contribution								
S	0.479	-	0.416	0.188	0.487	0.166	0.379	-
E	0.521	0.362	0.584	-	0.513	0.426	0.621	-
D	-	0.157	-	0.172	-	-	-	0.251
H	-	0.481	-	0.640	-	-	-	0.390
A	-	-	-	-	-	0.408	-	0.359

<sup>a</sup> Cross-validated correlation coefficient using the leave-one-out (LOO) method

<sup>b</sup> Non-cross-validated correlation coefficient

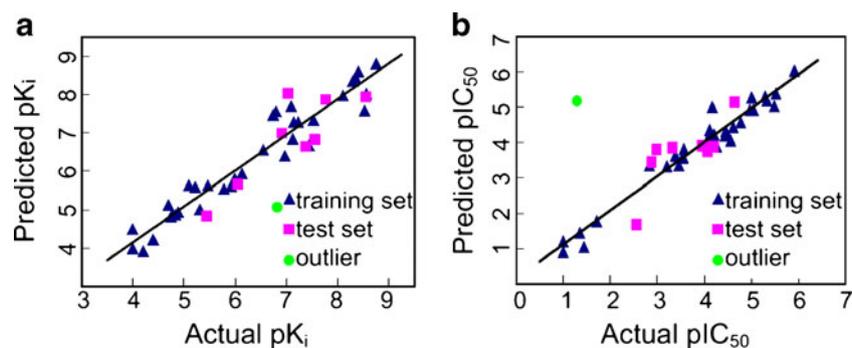
<sup>c</sup> Standard error of estimate

<sup>d</sup> Ratio of  $R_{ncv}^2$  explained to unexplained= $R_{ncv}^2/(1-R_{ncv}^2)$

<sup>e</sup> Predicted correlation coefficient for the test set compounds

<sup>f</sup> Standard error of prediction

<sup>g</sup> Optimal number of principal components



**Fig. 4** Graphs of predicted vs actual  $pK_i$  or  $pIC_{50}$  values of **a** structure-based comparative molecular similarity indices analysis (CoMSIA) model of set-I, and **b** ligand-based comparative molecular field analysis (CoMFA) model of set-II, respectively

active IMI (compound 1) was used as a reference as shown in Figs. 5 and 6.

#### Set-I

As seen in Fig. 5a, a big yellow polyhedron at positions -3, -4, and -5 of the tetrahydroimidazole ring suggests that bulky substituents in these areas will decrease biological activity significantly. For example, compounds 17–20, with a bulky substituent at position -3 have lower activity than compound 1. There is a big green contour region around positions -7, -8 and -9 of chloropyridinyl, which, together with a small green contour near position -1, indicates that a bulky substituent is preferred in these positions to produce higher inhibitory activity. This is also confirmed by the fact that compound 1, with a chloropyridinyl substituent is more active than compounds 2–6 with small alkyl substituents; compounds 32, 33, and 34 with bulky groups [–NC(O)OPh] near position -1 show higher activity than compounds 23, 29, and 37 with small substituents (–NCHO).

The hydrophobic field contour map is depicted in Fig. 5b. A large white contour around positions -1, -6 and -11 shows the region where the polar group is favored, which could explain why compounds 30 and 31 with polar groups –NCOCF<sub>3</sub> at the position -1 have higher activity. Two red-orange contour regions near positions -3, -4 and -

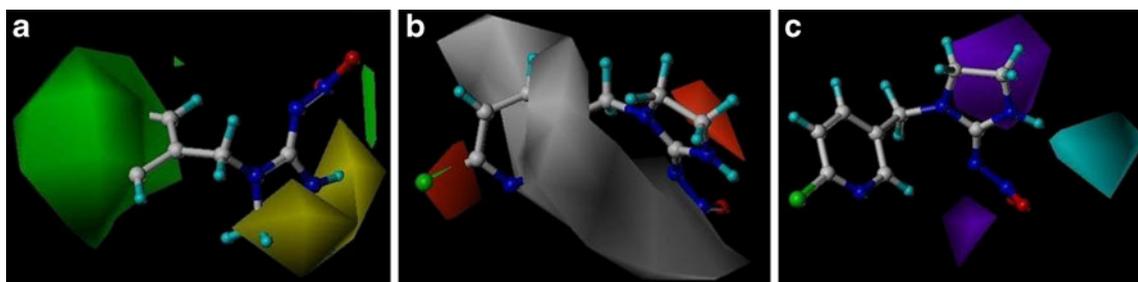
8 suggest that the presence of the hydrophobic group is favorable.

In Fig. 5c, two purple polyhedrons near positions -1, -4 and -5 indicate that these areas are disfavored for H-bond donor interactions, while another cyan polyhedron above position -3 shows a favored region for H-bond donor interactions. Thus, the low potency of compound 24 may be attributed to the lack of H-bond donor atoms at position -3.

#### Set-II

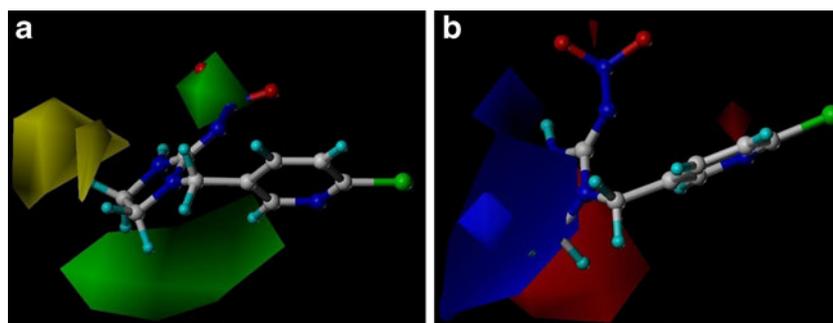
In Fig. 6a, two big green polyhedrons around position -1 of the nitro group and position -12 of chloropyridinyl ring suggest a requirement for bulky substituents to enhance biological activity. This explains why all compounds of this series have large substituents such as a chloropyridinyl ring. Additionally, this is also validated by compound 29 without a benzene ring at position -1, which exhibits lower activity than compounds 44–57. Two yellow polyhedrons at positions -3 and -4 of the tetrahydroimidazole ring confirm that bulky groups are disfavored in this region, which is consistent with the fact that compound 65, with an alkyl substituent at position -3 is less active than compound 64.

In Fig. 6b, a big blue polyhedron beside the tetrahydroimidazole indicates that a positively charged group is



**Fig. 5** a–c CoMSIA contour plots for set-I combined with imidacloprid (IMI). **a** Green Sterically favorable contours, yellow sterically unfavorable contours. **b** Red-orange Hydrophobic favorable contours, white hydrophobic unfavorable contours. **c** Cyan H-bond donor

favorable, purple H-bond donor unfavorable contours. The maps represent 80% (favored) and 20% (disfavored) level contributions, respectively



**Fig. 6 a,b** CoMFA contour plots for set-II combined with IMI. **a** *Green* Sterically favorable contours, *yellow* sterically unfavorable contours. **b** *Blue* Positive charge favorable contours, *red* positive

charge unfavorable contours. The maps represent 80% (favored) and 20% (disfavored) level contributions, respectively

necessary in this area to increase activity. This is revealed by the order of activity for those compounds: compounds 68 (S)>64 (N)>67 (O); compounds 31, 69 (N)>30, 66 (S). Two small red polyhedrons at position -1 of the nitro group and position -7 of chloropyridinyl imply that H-bond interactions could exist in this region. Thus, compound 68, which has a relatively electronegative group (COCF<sub>3</sub>) at position -1, is more active than compound 29 (having COCH<sub>3</sub>). Another single big red polyhedron at position -6 also indicates that this location favors negatively charged substituents.

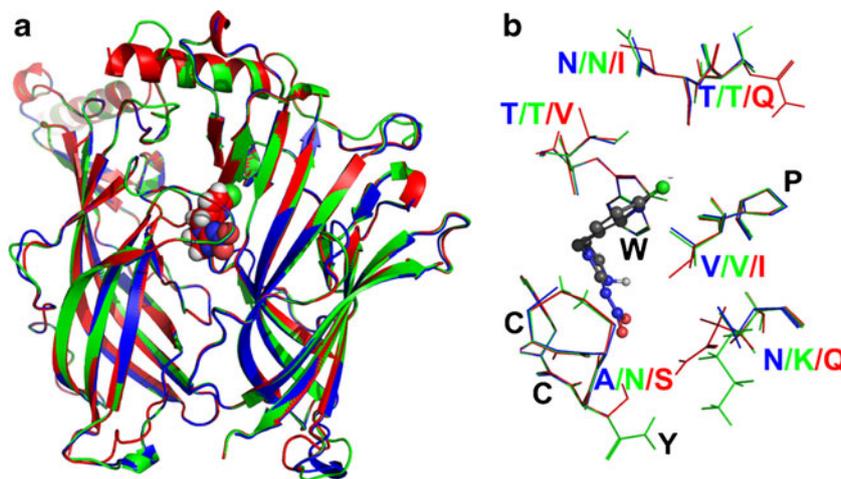
### Homology modeling

Homology modeling, i.e., comparative modeling of protein, is a method used to construct protein structure based on the general observation that proteins with sequence identity greater than 30% share common structural elements [39]. In the present work, the functional sequence (loop A–F) identities between the target (*Dm-/Md-nAChR*) and the

template (*Ac-AChBP*) protein are up to 51% and 49%, respectively. The superimposition of homology models on the template (Fig. 7) indicates that the overall conformations of the targets are very similar to the template, with a root mean square deviation (RMSD) <1 Å. Almost all critical residues of the binding site overlap well in 3D space for these three structures (*Ac-AChBP*, red; *Dm-nAChR*, green; *Md-nAChR*, blue). The two homology modeling structures were then used for further MD analyses.

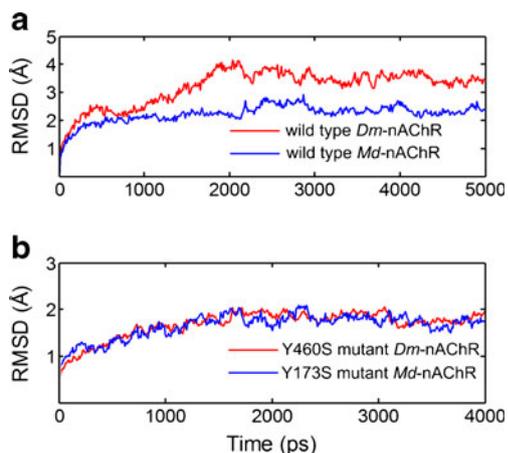
### MD simulation and docking studies

The homology modeled *Dm-nAChR* and *Md-nAChR* were subjected to MD simulation to obtain a more precise and energetically favorable stable receptor conformation. To check the stability of the system throughout the simulation, the RMSD of the wild type protein was plotted vs the simulation time (Fig. 8a). During the simulation, the RMSD values of wild type *Dm-nAChR* and *Md-nAChR* system range from 0.9 Å to 4.1 Å and 1.8 Å to 3.0 Å, respectively.



**Fig. 7** Superimposition of **a** *Ac-AChBP* (red ribbon), *Dm-nAChR* (green ribbon) and *Md-nAChR* (blue ribbon). **b** Enlargement of the active site structure (colored as in **a**) with IMI displayed in ball and

stick format (*gray*). Conserved residues are colored *black*, while variant ones are labeled differently



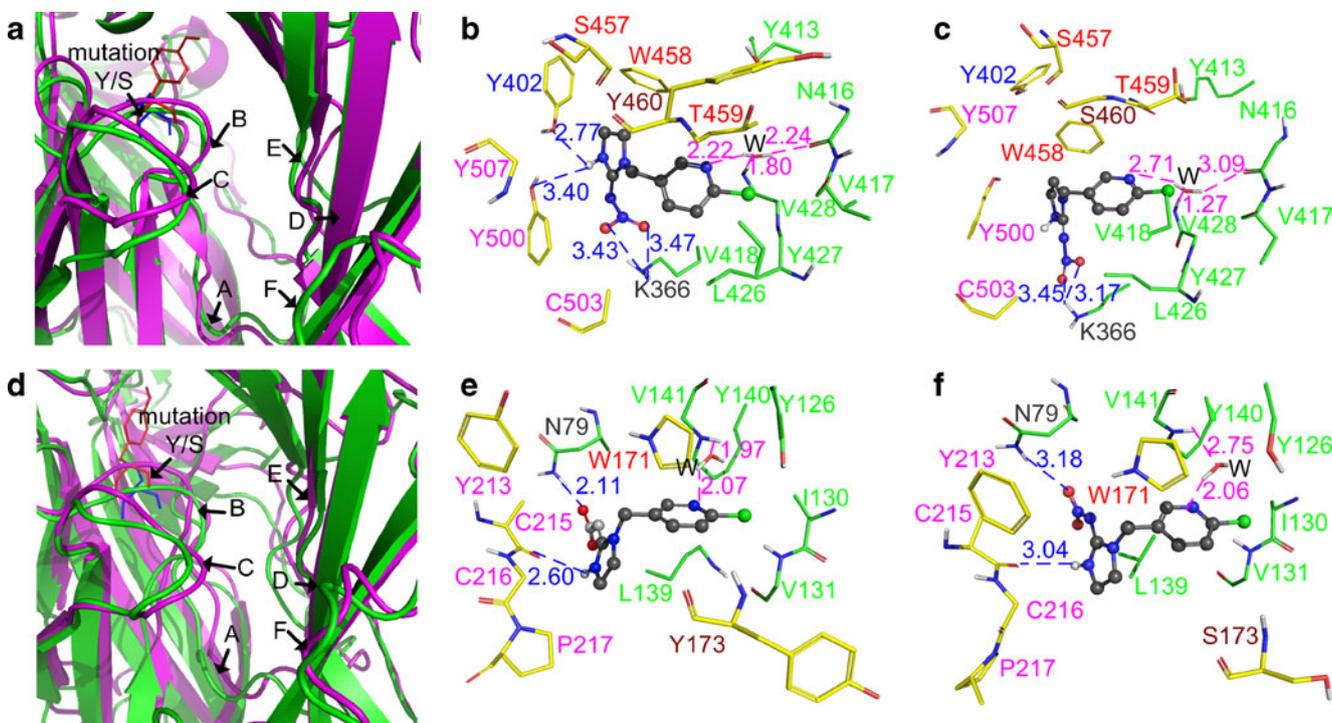
**Fig. 8** RMSD values of (a) wild type and (b) Y/S mutant *Dm/Md*-nAChR vs simulation time

During the last 1 ns, the RMSD values of each system are converged and the C-loop regions are very stable with a RMSD  $<1\text{Å}$ , indicating that the system is stable and well equilibrate. Two representative structures are obtained from the saved frames of the last simulation trajectories, which are then used for further analyses.

In order to understand the nature of the receptor–ligand binding mode, and to develop structure-based CoMFA/CoMSIA models, all conformations of set-I and II were docked into the representative structures of wild type *Dm*-nAChR and *Md*-nAChR, respectively. Figure 9b and 9e show the conformation derived from the allosteric binding site of wild type *Dm*-/*Md*-nAChR, where IMI is suitably localized at the binding site.

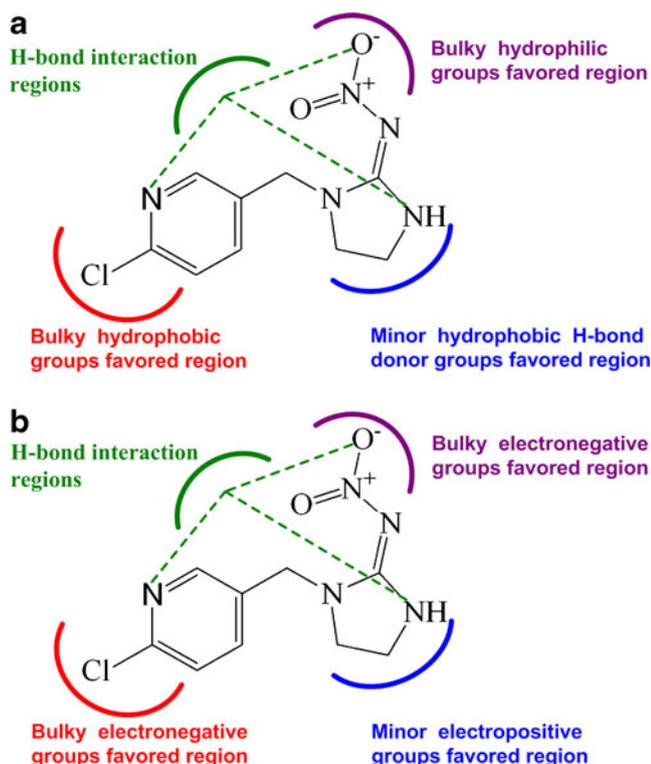
#### *Dm*-nAChR binding analysis

In Fig. 9b, the active site of *Dm*-nAChR contains two major hydrophobic pockets: pocket I is a deep and narrow aromatic box consisting of residues Tyr B402, Trp B458, Tyr B500 and Tyr B507; pocket II is a shallow and wide hydrophobic cavity formed by residues Tyr A413, Val A417, Val A418, Leu A426, Tyr A427 and Val A428. The ligand core is anchored in the binding site via several H-bonds and water-mediated contacts with the protein. The nitrogen atom of chloropyridinyl acts as an acceptor to form an H-bond with Val A428 ( $-\text{O}\cdots\text{HN}$ ,  $1.80\text{Å}$ ,  $176.6^\circ$ ) and Asn A416 ( $-\text{OH}\cdots\text{O}$ ,  $2.24\text{Å}$ ,  $168.1^\circ$ ) through a structural water molecule ( $-\text{N}\cdots\text{HO}$ ,  $2.22\text{Å}$ ,  $141.8^\circ$ ) at pocket II. For



**Fig. 9** Superimposition of the wild type (magenta) and mutant (green) structures of **a** *Dm*-nAChR and **d** *Md*-nAChR. Mutant residue Y (red)/S (blue) is represented in stick form. Molecular docked conformations are derived from IMI with the binding site of **b** wild type *Dm*-nAChR, **c** Y460S mutant *Dm*-nAChR, **e** wild type *Md*-nAChR, and **f** Y173S mutant *Md*-nAChR. Relevant amino acids are represented in stick form (subunit A in green; B in yellow)

by loop A (blue), loop B (red; mutant residue dark red), loop C (pink), loop D (gray), and loop E (green). IMI is shown in ball and stick form (gray). The H-bonds formed directly between residues and molecule, and those mediated indirectly by water are shown as dotted lines with blue and magenta color, respectively. W Water molecules. The nonpolar hydrogen atoms are removed for clarity



**Fig. 10** Comparison of **a** *Dm*-nAChR and **b** *Md*-nAChR binding information with IMI

pocket I, the nitrogen of tetrahydroimidazole is seen involved in H-bond interaction with the Tyr B500 ( $-\text{NH}\cdots\text{O}$ , 3.40 Å, 135.3°) and Tyr B402 ( $-\text{NH}\cdots\text{O}$ , 2.77 Å, 107.9°). In addition, the oxygen of the nitro group forms an H-bond with Lys A366 ( $-\text{O}\cdots\text{HN}$ , 3.47 Å, 103.3°;  $-\text{O}\cdots\text{HN}$ , 3.43 Å, 106.0°), which further enhances the interaction between the ligand and the receptor.

Interestingly, the docking result is consistent with CoMSIA contour map analysis, which further validates the 3D-QSAR model overall. The position -1 of the nitro group and positions -7, -8 and -9 of the chloropyridinyl ring fit nicely into the relatively large pockets I and II. Therefore, a relatively bulky substituent is needed at this position, as also evidenced by the presence of the sterically favorable green contours around these areas seen by the CoMSIA model. However, replacement with substituents that are too large at positions -3, -4, and -5 leads to steric clash with residues Tyr B402, Trp B458, and Tyr B507, as is evident from the presence of a yellow contour nearby. This feature of the pocket is also in agreement with the CoMSIA hydrophobic field contour map. Two major hydrophobic pockets near the tetrahydroimidazole (I) and chloropyridinyl ring (II) may explain the red-orange contour regions near the positions -3, -4 and -8. A large white contour around the positions -1, -6 and -11 shows the region favorable to polar groups as corroborated by the

presence of Ser B457, Thr B459 and Cys B503. The small cyan contour seen near position -3 of tetrahydroimidazole shows an H-bond donor favorable region, as evidenced by the Tyr B500 and Tyr B402 located nearby.

#### *Md*-nAChR binding model

Residues Tyr A126, Tyr A140, Val A141, Leu A139, Ile A130 and Val A131 form a wide hydrophobic pocket (Fig. 9e). The nitrogen atom of chloropyridinyl forms an H-bond with a water molecule ( $-\text{N}\cdots\text{HO}$ , 2.07 Å, 130.8°), which itself forms an H-bond with residue Val A141 ( $-\text{O}\cdots\text{HN}$ , 1.97 Å, 100.2°) of the pocket. The nitro oxygen atom and tetrahydroimidazole nitrogen atom are involved in H-bond interactions with the Asn A79 ( $-\text{OH}\cdots\text{N}$ , 2.11 Å, 148.3°) and Cys B215 ( $-\text{OH}\cdots\text{N}$ , 2.60 Å, 100.1°), respectively.

Similarly, the docking result is in agreement with the overall CoMFA maps. Residues Tyr B213, Cys B215, Cys B216 and Pro B217 occupy the area above the tetrahydroimidazole; thus, large substituents in this area would conflict with these residues and are not favored for increasing the molecular activity. On the contrary, the lack of residues near chloropyridinyl and nitro groups indicate that bulk substituents are favored in this region. The electronegative favorable red contour observed near the nitro oxygen group and chloropyridinyl nitrogen suggests an H-bond favorable region, which is also corroborated by the docking results.

#### Y/S mutation analysis

Certain amino acids in insect nAChR might be critical for NNs selectivity, including important residues in loop A–F [20]. Mutation of these amino acids to other residues that are identical or similar to the corresponding residues in vertebrate subunits may also contribute to the development of resistance to these insecticides [20]. A first point mutation in loop B, Y151S mutant of *Nilaparvata lugens* nAChR has been identified as being associated with target-site resistance to IMI and cross-resistance to other NNs [20]. As yet, no work has established the prevalence of the Y151S mutation in field populations of *Drosophila* and *Musca*. However, Y/S mutant studies of *Dm*-/*Md*-nAChR are needed to investigate this point in conjunction with ongoing surveys of NNs resistance. Thus, MD simulations and molecular docking studies of Y460S/Y173S mutants *Dm*-/*Md*-nAChR were performed to obtain reliable mutation modes.

The RMSD of the mutated protein vs simulation time is shown in Fig. 8b. During the simulation, the RMSD values of the trajectory with respect to their initial wild type structures range from 0.5 to 2.0 Å (*Dm*-nAChR) and 0.7 to

2.1 Å (*Md*-nAChR), respectively. The small RMSD variations reached about 1.8 Å after 2 ns, indicating that each system is retained and equilibrates well. Then, two representative mutant structures obtained from the final simulation trajectories are used for docking analysis.

The comparisons of binding modes of wild type and mutant *Dm*-/*Md*-nAChR from docking studies are shown in Fig. 9. Note that the H-bonds with Tyr500 (loop C) and Tyr402 (loop A) of *Dm*-nAChR are lacking in Y460S mutants. For the Y173S mutation *Md*-nAChR, the H-bond number does not appear to differ significantly from the wild type, while the transformation of weakened H-bond nonetheless contributes an important factor to the reduced binding affinity of IMI. Thus, the Y/S mutants of loop B in *Dm*-nAChR and *Md*-nAChR both cause a shift to a lower binding affinity for IMI, which is consistent with the enhanced IMI resistance seen in pests [20]. Despite the fact that Tyr may not be involved directly in the binding of NNs, it seems plausible that the mutation induces a conformational change within the *Dm*-/*Md*-nAChR binding site region that involves other amino acids, which is essential for the binding of the NNs. This could lead to a strategy for the rational design of novel NNs effective against target-based resistant pests.

#### Comparison of binding modes for each species

To explore the similarities and differences between *Dm*-nAChR and *Md*-nAChR, we compared the binding mode of each species (Fig. 10). Overall, the nAChR agonist-binding site is present at the interface of adjacent subunits and is formed by loops A–C in  $\alpha$  subunits, together with loops D–F in homomer-forming  $\alpha$  subunits [40]. In detail, for loop E, the interaction between the nitrogen atom of chloropyridinyl and Val428, Asn416 of *Dm*-nAChR or Val141 of *Md*-nAChR by water mediation are conserved. Corringer et al. [5] also generated 3D models of wild-type chicken  $\alpha 7$  nAChR and discovered that Leu118 of loop E (a position analogous to Val428/Val141 of *Dm*-/*Md*-nAChR) is situated close to the ligand binding site and influences ligand–protein interactions, which further confirms this interaction. In loop C, H-bonds are formed between the nitrogen of tetrahydroimidazole and Tyr500/Cys215 of *Dm*-/*Md*-nAChR. As described previously, the YXCC motif of loop C is conserved and critical for nAChR [41]. For *Dm*-nAChR loop A, Tyr402 is also involved in the H-bond interaction with this nitrogen atom. This may account for the difference that *Dm*-nAChR (pIC<sub>50</sub> of 5.3) shows a little higher binding affinity of IMI than *Md*-nAChR (pIC<sub>50</sub> of 5.0). In addition, the NO<sub>2</sub> group of IMI interacts through H-bond with Lys366/Asn79 of *Dm*-/*Md*-nAChR loop D. The effects of loop D could also be interpreted by the earlier findings of Shimomura et al. [42], in which the

corresponding residue Thr77 of the chicken nAChR  $\beta 2$  subunit was shown to be responsible for NN selectivity.

By contrast, the NN binding modes of each species does not change significantly. Therefore, we confirm that residues located within insect-specific loops A, C, D and E play key roles in the interactions of homo-pentamers *Dm*-/*Md*-nAChR with NNs, thus further promoting our understanding of NN-*Dm*-/*Md*-nAChR interactions.

#### Conclusions

nAChR is an agonist-gated ion channel complex involved in rapid excitatory neurotransmission. It is distributed widely in the insect central nervous system and constitutes a major target for NN action. Therefore, the design of inhibitors of nAChR has attracted much research interest in the field of developing novel insecticides. In this study, two optimal 3D-QSAR models of *Dm*-/*Md*-nAChR were obtained, and the statistical results indicate that the models are sufficiently reliable to predict the inhibitory efficiency of these compounds. Furthermore, homology modeling, MD simulation and molecular docking analysis show that residues located within loops A, C, D and E play key roles in the interactions of *Dm*-/*Md*-nAChR with NNs. Overall, the 3D contour maps have a good correlation with the molecular docking analysis, which further proves the reliability of the models and promotes understanding of ligand–receptor interactions. Following mutation analysis, we also suggest that the Y/S mutation within loop B is associated closely with NN resistance in *Drosophila* and *Musca*. It can be concluded that some key structural determinants of novel potent inhibitors against the *Dm*-/*Md*-nAChR are as follows (IMI as a reference): (1) the chloropyridinyl ring region requires bulky, electronegative, and hydrophobic groups. (2) Substituents of the tetrahydroimidazole nitrogen area should be small, electropositive, and hydrophobic. (3) Larger, electronegative, and polar groups at nitro region are favorable for inhibitory activity.

**Acknowledgments** This work is supported by the high-performance computing platform of northwest A & F University. The research is supported financially by the Fund of northwest A & F University. The authors are grateful to Prof. L. Yang for access to Sybyl software.

#### References

- Ortells MO, Lunt GG (1995) Trends Neurosci 18:121–127
- Karlin A (2002) Nat Rev Neurosci 3:102–114
- Tomizawa M, Millar NS, Casida JE (2005) Insect Biochem Mol 35:1347–1355
- Revah F, Bertrand D, Galzi JL, Devillers TA, Mulle C, Hussy N, Beryrand S, Ballivet M (1991) Changeux JP 353:846–849
- Corringer PJ, Le NN (2000) Changeux JP 40:431–458

6. Breer H, Sattelle DB (1987) *J Insect Physiol* 33:771–790
7. Bloomquist JR (1996) *Annu Rev Entomol* 41:163–190
8. Casida JE, Quistad GB (1998) *Annu Rev Entomol* 43:1–16
9. Yamamoto I, Tomizawa M, Saito T, Miyamoto T, Walcott EC, Sumikawa K (1998) *Arch Insect Biochem* 37:24–32
10. Tomizawa M, Lee DL, Casida JE (2000) *J Agr Food Chem* 48:6016–6024
11. Tomizawa M, Zhang NJ, Durkin KA, Olmstead MM, Casida JE (2003) *Biochemistry* 42:7819–7827
12. Tomizawa M, Casida JE (2009) *Accounts Chem Res* 42:260–269
13. Bai D, Lummis SCR, Leicht W, Breer H, Sattelle DB (1991) *Pestic Sci* 33:197–204
14. Tomizawa M, Casida JE (2005) *Annu Rev Pharmacol* 45:247–268
15. Jeschke P, Nauen R (2008) *Pest Manag Sci* 64:1084–1098
16. Taylor CE, Georghiou GP (1982) *Environ Entomol* 11:746–750
17. Nauen R, Denholm I (2005) *Arch Insect Biochem* 58:200–215
18. Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Sharakhova MV, Unger MF, Collins FH, Feyereisen R (2002) *Science* 298:179–181
19. Rauch N, Nauen R (2003) *Arch Insect Biochem* 54:165–176
20. Liu Z, Williamson MS, Lansdell SJ, Denholm I, Han Z, Millar NS (2005) *Proc Natl Acad Sci USA* 102:8420–8425
21. Galzi JL, Revah F, Black D, Goeldner M, Hirth C, Changeux JP (1990) *J Biol Chem* 265:30–37
22. Sine SM (2002) *J Neurobiol* 3:431–446
23. Smit AB, Syed NI, Schaap D, Van MJ, Klumperman J, Kits KS, Lodder H, van der Schors RC, Van Elk R, Sorgedragger B, Brejc K, Sixma TK, Geraerts WPM (2001) *Nature* 411:261–268
24. Talley TT, Harel M, Hibbs RE, Radic Z, Tomizawa M, Casida JE, Taylor P (2008) *Proc Natl Acad Sci USA* 105:7606–7611
25. Okazawa A, Akamatsu M, Ohoka A, Nishiwaki H, Cho WJ, Nakagawa Y, Nishimura K, Uenol T (1998) *Pestic Sci* 54:134–144
26. Okazawa A, Akamatsu M, Nishiwaki H, Nakagawa Y, Miyagawa H, Nishimura K, Uenol T (2000) *Pest Manag Sci* 56:509–515
27. Akamatsu M (2011) *J Agric Food Chem* 59:2909–2917
28. Ohno I, Tomizawa M, Durkin KA, Casida JE, Kagabu S (2009) *J Agric Food Chem* 57:2436–2440
29. Ohno I, Tomizawa M, Durkin KA, Naruse Y, Casida JE, Kagabu S (2009) *Chem Res Toxicol* 22:476–482
30. Ohno I, Tomizawa M, Aoshima A, Kumazawa S, Kagabu S (2010) *J Agric Food Chem* 58:4999–5003
31. Ohno I, Tomizawa M, Miyazu N, Kushibikib G, Noda K, Hasebe Y, Durkin KA, Miyake T, Kagabu S (2010) *Bioorg Med Chem Lett* 20:5933–5935
32. Sivan SK, Manga V (2010) *J Mol Model* 16:1169–1178
33. Tomizawa M, Maltby D, Talley TT, Durkin KA, Medzihradzky KF, Burlingame AL, Taylor P, Casida JE (2008) *Proc Natl Acad Sci USA* 105:1728–1732
34. Thompson JD, Higgins DG, Gibson TJ (1994) *Nucleic Acids Res* 22:4673–4680
35. Arnold K, Bordoli L, Kopp J, Schwede T (2006) *Bioinformatics* 22:195–201
36. Hooft RW, Vriend G, Sander C, Abola EE (1996) *Nature* 381:272–272
37. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) *J Comput Phys* 23:327–341
38. Bisson WH, Westera G, Schubiger PA, Scapozza L (2008) *J Mol Model* 14:891–899
39. Ginalski K (2006) *Curr Opin Struct Biol* 16:172–177
40. Brejc K, Dijk WJ, Klaassen RV, Schuurmans M, Oost J, Smit AB, Sixma TK (2001) *Nature* 411:269–276
41. Shimomura M, Yokota M, Matsuda K, Sattelle DB, Komai K (2004) *Neurosci Lett* 363:195–198
42. Shimomura M, Yokota M, Ihara M, Akamatsu M, Sattelle DB, Matsuda K (2006) *Mol Pharmacol* 70:1255–1263