Competitive docking model for prediction of the human nicotinic acetylcholine receptor α7 binding of tobacco constituents

Hui Wen Ng[1],[*], Carmine Leggett[2],[*], Sugunadevi Sakkiah[1],[*], Bohu Pan[1], Hao Ye[1], Leihong Wu[1], Chandrabose Selvaraj[1], Weida Tong[1] and Huixiao Hong[1]

[1]Division of Bioinformatics and Biostatistics, Office of Research, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, AR 72079, USA

[2]Division of Non-clinical Science, Office of Science, Center for Tobacco Products, U.S. Food and Drug Administration, Silver Spring, MD 20993, USA

[*]These authors made equal contributions to the work

Correspondence to: Huixiao Hong, *email:* huixiao.hong@fda.hhs.gov

ABSTRACT

The detrimental health effects associated with tobacco use constitute a major public health concern. The addiction associated with nicotine found in tobacco products has led to difficulty in quitting among users. Nicotinic acetylcholine receptors (nAChRs) are the targets of nicotine and are responsible for addiction to tobacco products. However, it is unknown if the other >8000 tobacco constituents are addictive. Since it is time-consuming and costly to experimentally assess addictive potential of such larger number of chemicals, computationally predicting human nAChRs binding is important for in silico evaluation of addiction potential of tobacco constituents and needs structures of human nAChRs. Therefore, we constructed three-dimensional structures of the ligand binding domain of human nAChR α7 subtype and then developed a predictive model based on the constructed structures to predict human nAChR α7 binding activity of tobacco constituents. The predictive model correctly predicted 11 out of 12 test compounds to be binders of nAChR α7. The model is a useful tool for high-throughput screening of potential addictive tobacco constituents. These results could inform regulatory science research by providing a new validated predictive tool using cutting-edge computational methodology to high-throughput screen tobacco additives and constituents for their binding interaction with the human α7 nicotinic receptor. The tool represents a prediction model capable of screening thousands of chemicals found in tobacco products for addiction potential, which improves the understanding of the potential effects of additives.

INTRODUCTION

Continuing to smoke increases the risk of diseases that are caused by the toxicants in tobacco smoke to smokers themselves [1, 2]. Furthermore, exposure to tobacco smoke in the environment leads to adverse health effects such as respiratory symptoms, impaired lung function, coronary heart disease, nasal irritation, stroke and lung cancer in nonsmokers as is seen in smokers

[3–6]. The Third National Health Survey showed that a large fraction of the US population has detectable serum cotinine levels due to environmental tobacco smoke exposure [7]. The Surgeon General's Report and recent studies indicate that children exposed to tobacco smoke in the environment have a high likelihood of ear disease, respiratory symptoms, lower respiratory illness, and sudden infant death syndrome [8]. A longitudinal study of a large cohort of middle-aged and older adults found

that many individuals, especially women, were passively exposed to tobacco smoke at home, public transport stations, workplaces, and recreational places [9]. Thus, tobacco smoke has detrimental effects for nonsmokers as well as smokers and is a public health issue.

Nicotinic acetylcholine receptors (nAChRs) belong to the superfamily of ligand-gated ion channels. Pentameric in nature, nAChRs consist of an extracellular domain approximately 200 amino acids in length (where the ligand binding domain [LBD] resides), a transmembrane region (consisting of four transmembrane helices), and an intracellular domain that connects the third and fourth transmembrane helices [10]. There are two major types of nAChRs: the neuronal and the muscle-type nAChRs; the former have been a major target in studies of addiction arising from tobacco use. The neuronal nAChRs formed by 12 known subunits, α2-α10 (n.b. α8 found in avian family and not mammals) and β2-β4 [10, 11]. The α7-α10 subunits can form functional homopentamers, while the α2-α6 require other α or β subunits to form a functional heteropentamer [11–14]. The α7 nAChR has been found to play a role in reinforcing and inducing nicotine dependence [15–19].

Various studies have identified thousands of chemicals in tobacco smoke, with the highest estimate at over 9,500 chemicals [20]. While the harmful effects of some tobacco constituents, such as the list of 93 harmful and potentially harmful constituents (HPHCs) established by the FDA, have been well studied, the biological effects of the majority of tobacco constituents remain unknown. Accordingly, understanding the effects of these chemicals on the body may be useful for strategies to reduce tobacco product harm.

The biological effects of tobacco constituents are typically tested with a battery of laboratory experiments. However, acquiring experimental data on the biological actions of these chemicals is both time-consuming and costly. *In silico* techniques offer a rapid approach to study and prioritize laboratory experiments needed to study tobacco constituents. Among available computational methods such as pharmacophore modeling [21–24], comparative molecular field analysis [25], decision tree [26], decision forest [27–33], support vector machine [34, 35], and other machine learning methods [36–38], molecular docking is one of the most established and widely-used approaches to assess the binding activity of chemicals. Molecular docking involves the prediction of how chemicals interact with proteins [39–42]. Knowing the binding potential of a chemical is important as receptor binding often initiates a cascade of chemical-induced biological actions. Apart from predicting how a chemical binds to the active site of a receptor, the inherent flexibility of proteins in ligand-receptor recognition is another important factor that needs to be considered as the conformation of a protein is closely linked to its function. Many times, upon binding to a ligand, a protein

changes its conformation to perform different functions in complex biological processes [43]. However, majority of docking studies are still performed under the "fully flexible chemical vs. rigid protein" condition due to the high computational cost required to allow modeling flexibility. Indeed, even limited flexibility introduced to the protein (i.e., only on a few key residues in the active site) considerably increases the calculation time in the docking of a given chemical. Therefore, assessing a large library of compounds with molecular docking using fully flexible proteins is impractical.

Our previous study investigated the interactions of chemicals with the ligand binding domain of the α4β2 nAChR [44]. A similar study was conducted for the LBD of the α7 nAChR (α7 nAChR-LBD), not only to investigate how chemicals interact with the receptor but also to develop a model to predict the potential binding of chemicals to α7 nAChR-LBD. To date, the three dimensional (3D) structure of human α7 nAChR-LBD has not been elucidated experimentally. The closest available 3D structure to human α7 nAChR-LBD is α7 nAChR chimera (Protein Data Bank (PDB) ID: 3SQ6). Therefore, it would be useful to construct a 3D structure of human α7 nAChR-LBD, but also for a prediction model that incorporates the protein flexibility involved in the ligand-receptor recognition process. Here, we describe the development of a competitive docking model (CDM) based on an approach similar with our previously published model [40] for predicting estrogen receptor binding activity [32, 45] to help mitigate the shortcoming of rigid-protein docking by favoring the more energetically favorable receptor-ligand complex. The ability of CDM to predict the binding of chemicals to human α7 nAChR-LBD was assessed with a set of compounds whose human α7 nAChR-LBD binding had been experimentally tested. Finally, the important interactions that occurred when the receptor was bound by different chemicals were also investigated with molecular dynamics (MD) simulations [32, 46–48]. We elucidated the potential key residues in human α7 nAChR-LBD that take part in chemical binding, and revealed good performance of the model in chemical binding prediction. This model focuses on the interaction of these molecules with the receptor's ligand binding domain, but does not account for interactions at other receptor sites that may modulate receptor activity. This α7 nAChR binding activity prediction model may be useful for screening tobacco constituents that may have addiction potential as well as for regulatory priority setting for laboratory testing these compounds.

Figure 1 gives the overview of this study. The CDM, which accounts for protein flexibility, was developed to predict the binding potential of chemicals to the human α7 nAChR-LBD. Developed using a training set of ligands obtained from the PDB, the CDM was used to predict, with good accuracy, the binding potential for a set of test compounds whose binding activity was experimentally

validated. The potential key ligand binding interactions of human α7 nAChR were elucidated using MD simulations. These interactions include non-polar interactions (e.g., hydrophobic contacts, pi-pi interactions), as well as polar interactions (e.g., hydrogen bonding, pi-cation interactions). Key residues involved in these binding interactions have been identified. The elucidated binding interactions clarified the understanding of chemical binding potential to human α7 nAChR. The developed CDM shows applicability in computational high-throughput screening of tobacco constituents with addiction potential and may play a role in efforts to protect and improve public health.

RESULTS

Analysis of ligand binding pockets and selection of templates for homology modeling

The PDB complexes (PDB IDs and the associated ligands) that were chosen as potential templates for constructing human α7 nAChR-LBD are listed in Table 1. The receptors in these PDB structures were in complex with ligands (Figure 2) that were found to have human α7 nAChR binding data (with the exception of AN4 and AN5 bound to 2WNL, which had rat α7 nAChR binding data). Among these, 16 complexes were acetylcholine binding proteins (AChBP), while the remaining complex was a chimeric nAChR-LBD. Apart from the latter, which shared the highest percent sequence identity (63%) with human nAChR-LBD, all other templates were found to share approximately 24-26% sequence identity with the target. While three of these complexes were resolved at a resolution below 2 Å (2WNJ, 2WN9, 2XYS), the majority of the structures had a resolution of 2 to 3 Å.

Figure 3 depicts the results of the clustering analysis performed on the ligand binding pockets of the 17 PDB complexes. These complexes were clustered based on the RMSD of the residues lining their ligand binding pockets. Two distinct groups of protein conformations were observed; one representative
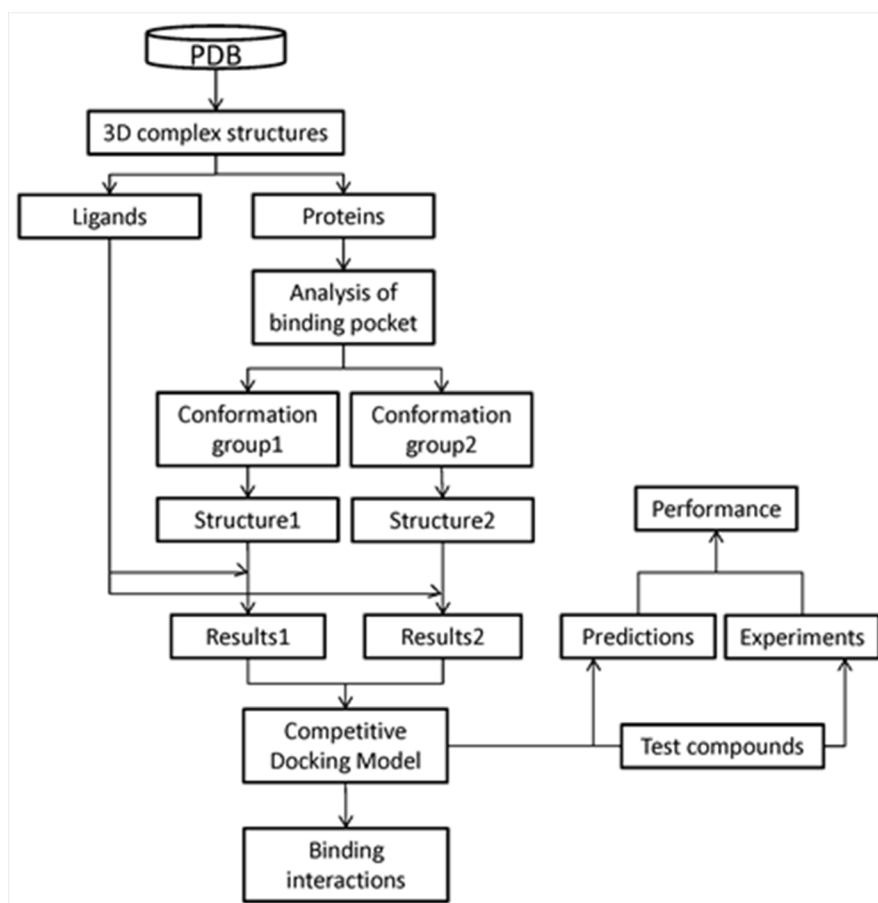

Figure 1: Overall modeling scheme. The 3D structures of complexes of nAChBPs with different ligands were retrieved from the PDB. Ligands and proteins were separated. Binding pocket similarity was analyzed, resulting two groups. One template structure was selected from each group based on sequence similarity with human nAChR. Two 3D structures of the human α7 nAChR-LBD were constructed using homology modeling based on the selected templates. Using these two structures, the CDM was developed with the training set ligands. The CDM was used to predict the binding activities of the test set compounds. The prediction results were validated through experimental determination of the binding activities of the test set ligands.

complex was selected from each of the two groups (3SQ6 and 2XYT) as templates for construction of human α7 nAChR-LBD structures. The complex structure 3SQ6 had the highest sequence identity with human α7 nAChR-LBD and thus was selected as a template for the homology modeling; the complex structure 2XYT was selected as another template for the homology modeling mainly due to its good resolution (2 Å) combined with the relatively rigid and large ligand (TC9). The main difference between the two complexes was that the receptor conformation of the complex 3SQ6 had a more enclosed/smaller ligand binding pocket compared to the receptor conformation of the complex 2XYT, which had a more open/larger ligand binding pocket.
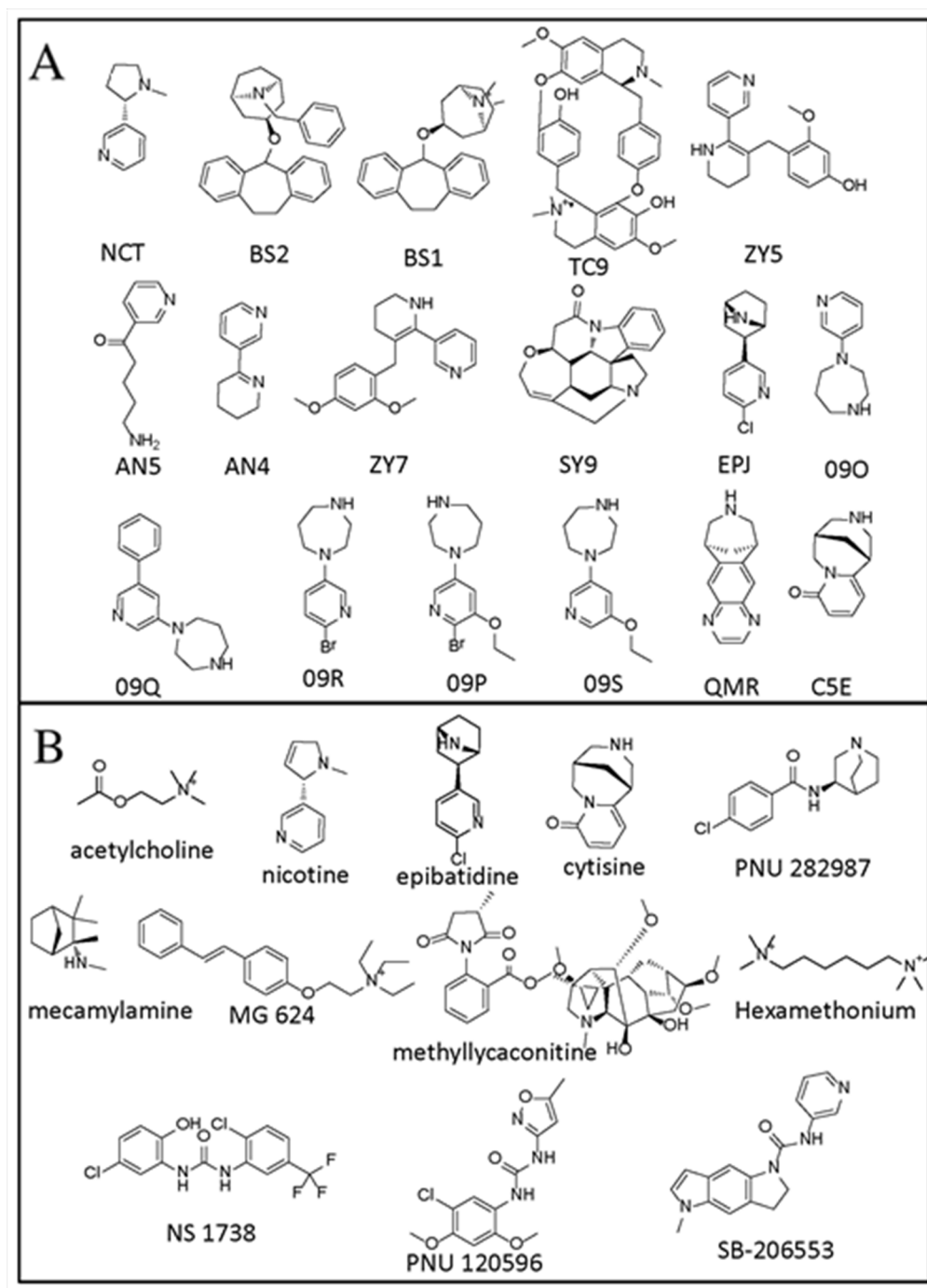

Figure 2: Structures of the compounds. The training **(A)** and test **(B)** compounds identifications that were used in this paper are noted below the structures.

Table 1: The PDB complexes shortlisted as potential templates to construct the human α7 nAChR-LBD, with associated details and available experimental α7 nAChR binding data

PDB ID	Ligand ID	Protein	Species	α7 binding data[*]		Sequence identity (%)	Resolution (Å)
				Ki(nM)	Ref.		
3SQ6	EPJ	α7-nrc[*]	HS, LS	18	47	63.24	2.80
1UW6	NCT	AChBP	LS	170±65	47	24.75	2.20
3U8J	09O	AChBP	LS	136	47	23.88	2.35
3U8L	09Q	AChBP	LS	>10000	47	23.88	2.32
3U8M	09R	AChBP	LS	190	47	23.88	2.70
3U8K	09P	AChBP	LS	9970	47	23.88	2.47
3U8N	09S	AChBP	LS	847	47	23.88	2.35
2W8G	BS2	AChBP	AC	79.4	48	26.24	2.60
2W8F	BS1	AChBP	AC	79.4	48	26.24	2.70
2XYT	TC9	AChBP	AC	2975	49	26.24	2.05
2WN9	ZY5	AChBP	AC	235[&]	50	25.24	1.75
2XYS	SY9	AChBP	AC	4854	49	26.24	1.91
2WNL	AN5	AChBP	AC	200[&]	50	25.24	2.70
2WNL	AN4	AChBP	AC	200[&]	50	25.24	2.70
2WNJ	ZY7	AChBP	AC	130	50	25.24	1.80
4AFT	QMR	AChBP	AC	322	51	26.24	3.2
4BQT	C5E	AChBP	AC	4200	51	26.24	2.88

Legends: [*] α7 nAChR binding data from human except AN4 and AN5 from rat; & Kd (nM).
Abbreviations: α7-nrc: α7-nicotinic receptor chimera; AChBP: Acetylcholine Binding Protein; HS: Homo sapiens; LS: Lymnaea stagnalis; AC: Aplysia Californica.

3D human α7 nAChR-LBD structures

Figure 4 shows the structures 3SQ6 and 2XYT and alignments of their protein sequences with the primary sequence of human α7 nAChR. The colored sections indicate the regions containing identical residues between the templates and human α7 nAChR. The protein in 3SQ6 shared a high sequence identity (63%) with human α7 nAChR-LBD (Figure 4A and 4C), while the protein in 2XYT shared 26% sequence identity with human α7 nAChR-LBD (Figure 4B and 4D). It was noted that the loop sections and the secondary structure regions (α-helices and β-sheets) of the protein in 3SQ6 aligned better with human α7 nAChR-LBD than those of the protein in 2XYT.

Figure 5 displays the initial 3D human α7 nAChR-LBD structures constructed by homology modeling using 3SQ6 (Figure 5A) and 2XYT (Figure 5B) as the templates. The Ramachandran plot of the initial 3D structure constructed from 3SQ6 (Figure 5C) shows that 92.2% of residues were placed in the favored zone (the red regions) and 99.1% of residues were placed in the allowed zone (the red and yellow regions), indicating that the 3D structure is of good quality and at a stable state. In the same way, the Ramachandran plot depicted in Figure 5D shows that 92.4% of residues were placed in the favored zone (the red regions) and 99.2% of residues were placed in the allowed zone (the red and yellow regions), indicating the initial homology structure based on 2XYT is of good quality and at a stable state.

The initial 3D structures were optimized through MD simulations. The optimized 3D structures of human α7 nAChR-LBD constructed from 3SQ6 and 2XYT were superimposed with the initial 3D structures from the homology modeling as shown in Figure 6A and 6B, respectively. Comparison of the optimized structures with their initial structures indicates that the MD simulations led to relatively larger changes in the conformation of the loops, while the helixes had very small changes in the optimizations. The Ramachandran plot of the optimized 3D structure from 3SQ6 (Figure 6C) shows that 93.1% and 99.4% of residues were placed in the favored zone

(the red regions) and the allowed zone (the red and yellow regions), respectively, higher than the percentages of residues (92.2% in the favored zone and 99.1% in the allowed zone) for the initial structure, indicating the optimized 3D structure is of a higher quality and at a more stable state. The Ramachandran plot of the optimized 3D structure from 2XYT (Figure 6D) shows that 93.7% and 99.6% of residues were placed in the favored zone (the red regions) and the allowed zone (the red and yellow regions), respectively, higher than the percentages of residues (92.4% in the favored zone and 99.4% in the allowed zone) for the initial structure, revealing the optimized 3D structure is of a higher quality and at a more stable state.

Competitive docking model

The results of docking the 17 ligands in the training set to the two human α7 nAChR-LBD structures are presented in Table 2. Three ligands (ID: TC9, BS1, BS2) were only docked in the human α7 nAChR-LBD structure constructed from 2XYT, and their docking scores were used in the development of the CDM. The other 14 ligands were successfully docked to both human α7 nAChR-LBD structures. According to the CDM rules, the lower docking scores were used for training the CDM. It was noted that the lower of the two docking scores for most of the ligands came from docking to the 3SQ6-based human α7 nAChR-LBD, except for SY9, which was the only ligand that obtained a lower docking score in the 2XYT-based human α7 nAChR-LBD structure than in the 3SQ6-based human α7 nAChR-LBD structure.

With the 17 winning docking scores listed in the last column of Table 2 as the training data, the predictive CDM was constructed using and is expressed by equation (1).

$$CDM = 1 - \frac{score + 8.725}{4.507} \qquad (1)$$

Key ligand binding interactions

The 17 human α7 nAChR-LBD-ligand complexes from the CDM were subjected to 20 ns MD simulations to investigate potential key ligand binding interactions. The RMSD plots of the MD simulations are shown in Supplementary Figure 1. These plots indicate that the 17 human α7 nAChR-LBD-ligand complexes reached a more or less stable state throughout the 20 ns MD simulations, following the sharp increase in the initial stage.

The key binding interactions and the residue fluctuations of human α7 nAChR-LBD were investigated using the frames from the last ns of the MD simulations. Supplementary Figure 2 shows the RMSF values that depict the fluctuations of residues of human α7 nAChR in the MD simulations. In general, the fluctuations for the residues remained between 0.5-2Å. Larger fluctuations were observed in the loop regions compared to the α-helices and β-sheets, as expected; human α7 nAChR complexes bound with ligands AN4, AN5, BS2 and 09R showed slightly more restricted movement compared to the complexes bound with other ligands, which may be due to the small size and rigid structures.

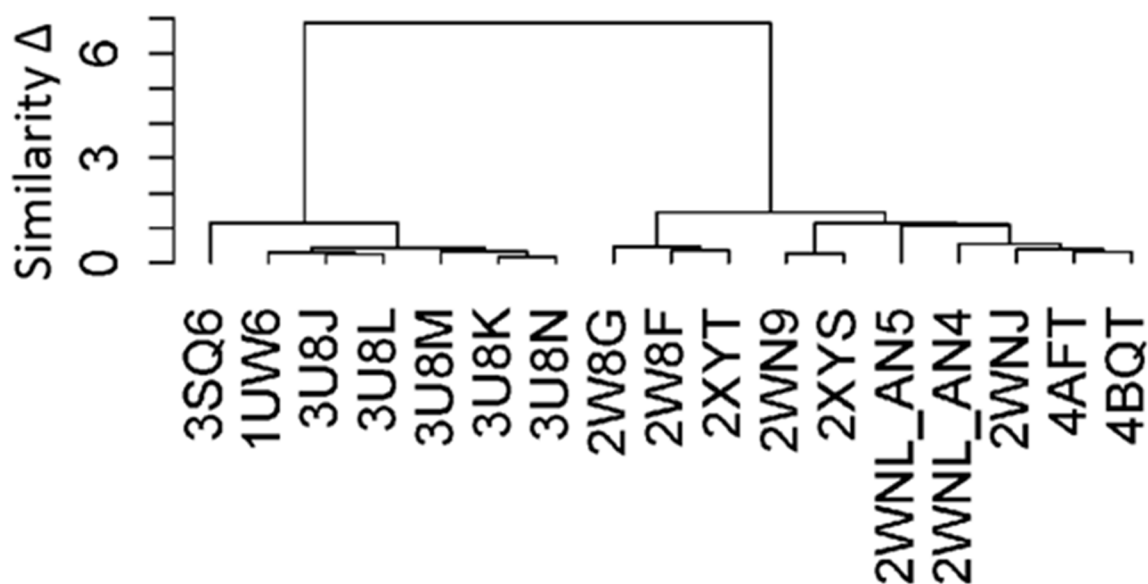The key binding interaction fractions calculated over the last 1 ns are shown in Supplementary Figures 3

Figure 3: The hierarchical clustering of the binding pockets of 17 complex structures. The similarity (y-axis) was measured using the RMSDs between the pocket conformations. The PDB ID (x-axis) was used to indicate the complex.

for hydrophobic contact, S4 for pi-pi interactions, and 5 for hydrogen bonds and pi-cation (See Supplementary Figures 6–22 for the diagrams of interactions for the individual ligands). Table 3 summarizes the important residues involved in the ligand binding of human α7 nAChR. The residues involved in hydrophobic contacts were Tyr93, Phe104, Val108, Leu109, Tyr118, Leu119, Pro121, Trp149, Tyr188 and Tyr 195 in chain A, and Trp55, Tyr93, Leu109, Leu119, Trp149, Ile169, Pro170, Tyr188, Cys190 and Tyr195 in chain B. Among these residues, the most important residue was Trp149 of chain B, which was involved in hydrophobic contact for most of the ligands. The other hydrophobic interaction residues were Tyr195 of chain B and Leu119 of chains A and B. The residues involved in pi-pi interactions

were Phe104, Trp149, Tyr195 and Tyr188 in chain A and Trp55, His115, Trp149, Tyr195 and Tyr188 in chain B. The residues involved in hydrogen bonding interactions were Arg79, Asn107, and Leu119 in chain A and Arg79, Gln117, Trp149, Asp164, Ser166, Tyr188, and Tyr195 in chain B. Hydrogen bonding with the backbone of Leu119 of chain A was particularly important – eight of the 17 ligands formed hydrogen bonds with this residue. Interestingly, hydrogen bonding interactions were not found for BS1, BS2, EPJ, and QMR. In addition to hydrogen bonding, another polar interaction type (i.e., pi-cation) was observed between some ligands (BS1, ZY5, ZY7 and AN4) and human α7 nAChR-LBD. While the pi-cation interaction between human α7 nAChR-LBD and BS1 arose from the interactions between the positive
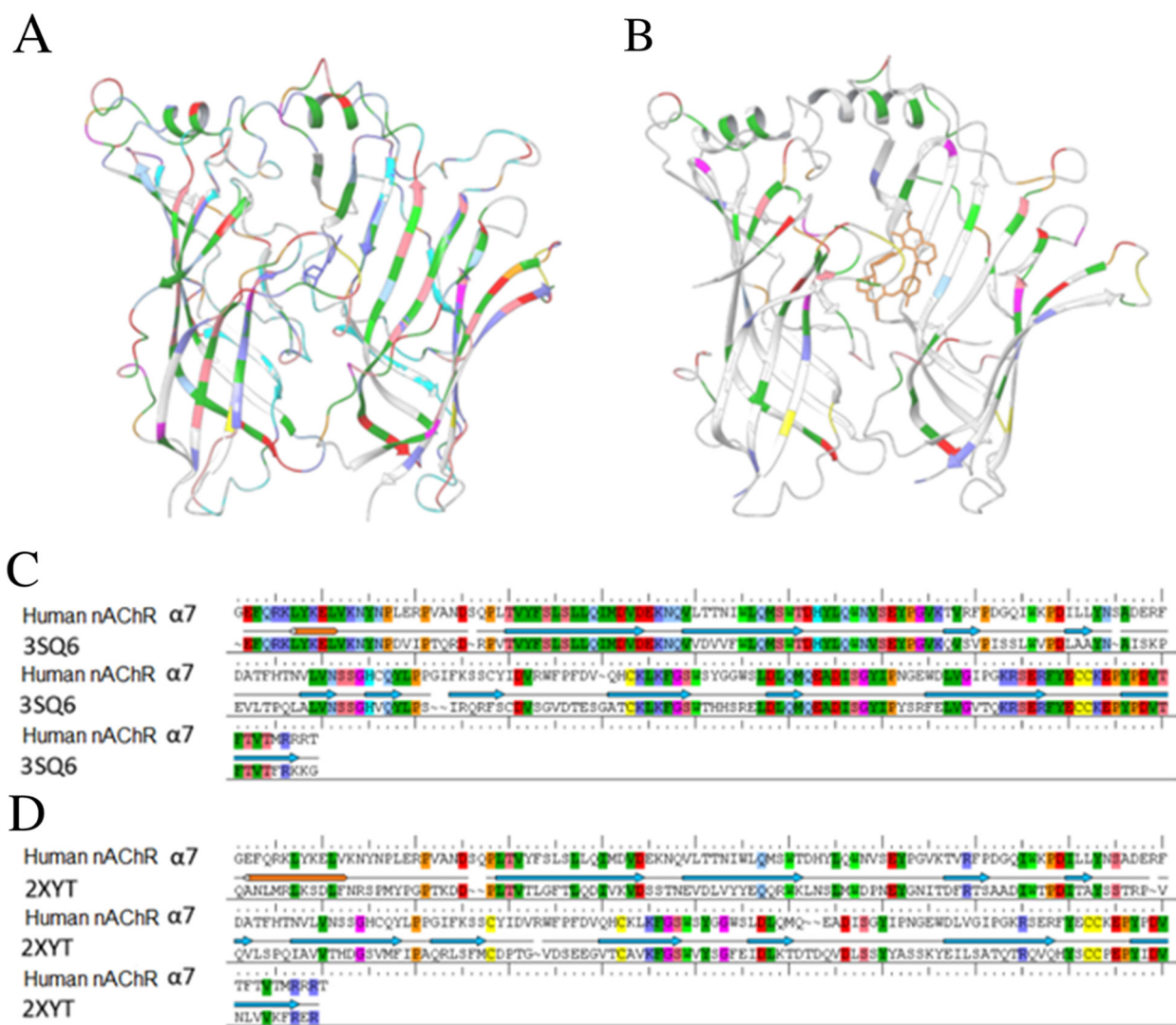

Figure 4: The template structures used in homology modeling of human α7 nAChR-LBD structures. 3D structures are shown for 3SQ6 **(A)** and 2XYT **(B)**. The alignment results of the sequence of the human α7 nAChR with the template proteins in 3SQ6 **(C)** and 2XYT **(D)** are shown. The colored regions in a-d indicate the regions that were well-aligned between the human α7 nAChR and 3SQ6 or 2XYT.

charge found in the quaternary nitrogen of the ligand and the two tyrosine residues (Tyr188 and Tyr195) in chain A of the receptor, the pi-cation interactions between human α7 nAChR-LBD and ZY5, ZY7 and AN4 arose from the interactions between the aromatic rings in these residues with the positive charge of Arg79.

CDM predictions of test compounds

Human α7 nAChR binding activity was predicted for the 12 test compounds with the CDM expressed by equation (1). MG624 and hexamethonium were docked only to the 2XYT-based human α7 nAChR-LBD and the docking scores were used in the CDM predictions. The remaining 10 compounds were successful in molecular docking to both structures and, thus, the lower docking scores were used in the CDM predictions as listed in Table 4. The CDM prediction value for hexamethonium was -0.318 and, thus, hexamethonium was predicted as a non-binder. Hexamethonium exhibited binding affinity for nicotine binding sites, but did not show inhibition in methyllycaconitine binding, demonstrating that hexamethonium interacts with α4β2 but not α7 receptor subtypes, respectively [49]. The CDM prediction values for the remaining 11 compounds were positive (see the last column of Table 4) and, thus, these compounds were predicted as binders of human α7 nAChR. Out of these 11 predicted α7 binders, acetylcholine [50], epibatidine [51], nicotine [52], PNU282987 [53], methyllycaconitine [54], NS1738 [55], PNU120596 [56], and SB206553 [57] were confirmed by the experimental data reported in the literature.
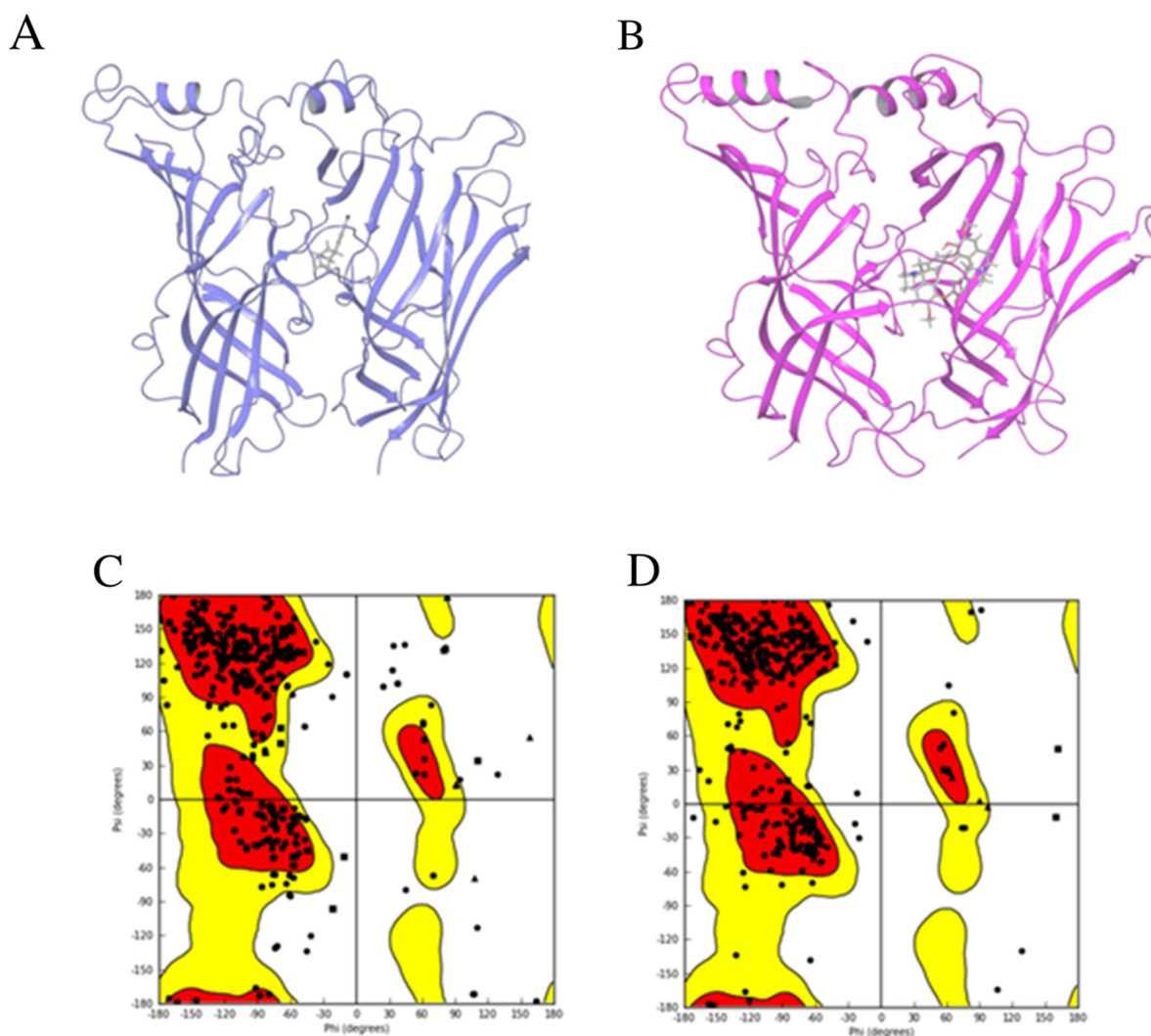

Figure 5: Homology modeling results. The initial structures of the human α7 nAChR-LBD constructed using homology modeling based on the template structures 3SQ6 **(A)** and 2XYT **(B)**. The proteins were drawn in ribbon models and the ligands were plotted in stick models. The Ramachandran plots are given for the corresponding initial structures based on 3SQ6 **(C)** and 2XYT **(D)**. The x-axis and y-axis indicate the two dihedral angles of amino acid residues in the homology model structures. Each point represents an amino acid residue. The favored regions are color-coded in red, the allowed regions in yellow, and the not allowed regions in white.

DISCUSSION

Computational works such as MD simulations on nAChR α7 receptor of nonhuman species were reported [58, 59]. To date, the structure of a human α7 nAChR-LBD has yet to be elucidated. In order to inform the assessment of the addiction potential of tobacco constituents, 3D structures of human α7 nAChR-LBD are of great value, especially in the identification of tobacco constituents that have potential to bind human α7 nAChR. Knowledge of human α7 nAChR binding potential of tobacco constituents would be useful, as binding human α7 nAChR is a biological process that involves a cascade of biological responses in addiction and can be used for

screening tobacco constituents with addiction potential. Therefore, the results could potentially be used to model a reduction in addiction potential of tobacco products by decreasing the amount of tobacco constituents that bind human α7 nAChR.

Homology modeling was used to generate the 3D structure of human α7. The template was selected based on the size of the ligand binding pocket of AChBP from different species. The 17 crystal structures were classified into two group. One template was selected from each group (3SQ6 and 2XYT) to model the human α7 nAChR. The size of the binding pocket of 2XYT was larger than 3SQ6 due to the binding of large ligand (TC9). The sequence alignments between the human
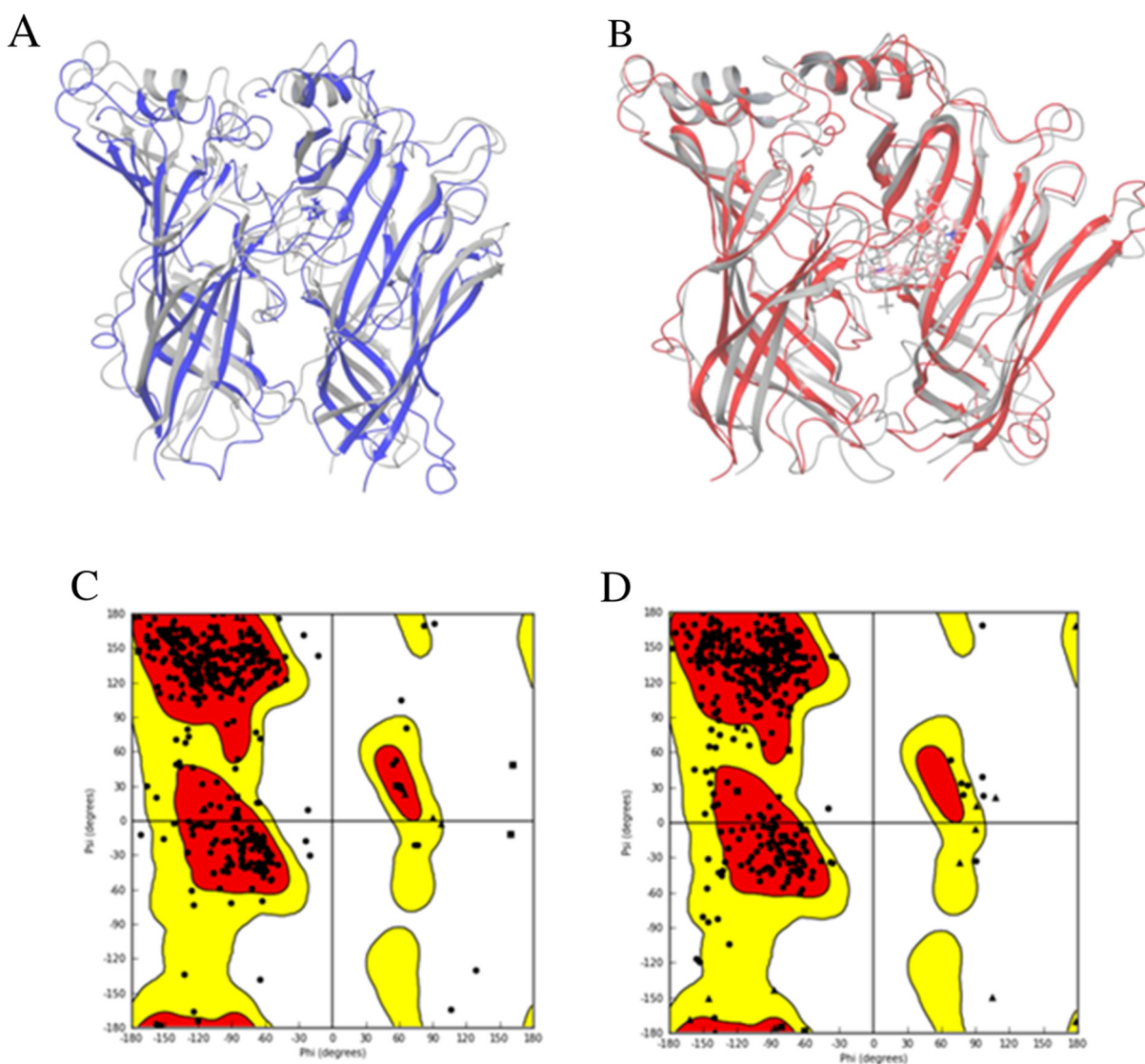

Figure 6: Optimized structures. The optimized structures of the human α7 nAChR-LBD based on the template structures 3SQ6 **(A, purple)** and 2XYT **(B, red)**. The initial homology structures are shown in grey. The Ramachandran plots are given for the corresponding optimized structures based on 3SQ6 **(C)** and 2XYT **(D)**. The favored regions are color-coded in red, the allowed regions in yellow, and the not allowed regions in white.

Table 2: Docking scores for the training set ligands obtained from docking to both 3SQ6-based and 2XYT-based α7 nAChR-LBD structures as well as the winning score

ID	XPGscore[*]		Winning score
	3SQ6-based α7 nAChR-LBD	2XYT-based α7 nAChR-LBD	
EPJ	-8.725	-5.2	-8.725
ZY5	-8.082	-5.601	-8.082
ZY7	-8.047	-5.537	-8.047
09S	-8.016	-4.093	-8.016
09Q	-7.704	-5.072	-7.704
AN5	-7.673	-4.681	-7.673
09R	-7.666	-5.109	-7.666
NCT	-7.512	-3.1	-7.512
09O	-7.439	-3.656	-7.439
QMR	-7.353	-4.135	-7.353
09P	-7.257	-3.616	-7.257
AN4	-6.982	-4.674	-6.982
C5E	-6.362	-3.65	-6.362
TC9	ND	-5.583	-5.583
BS1	ND	-5.015	-5.015
BS2	ND	-4.36	-4.36
SY9	-3.596	-4.218	-4.218

Abbreviations: [*] XPGscore: Extra-precision Glide docking score.

α7 nAChR and the templates were performed using ClustalW from Maestro. The 3SN6 template had shown a higher sequence identity than 2XYT. Prime module from Maestro was used to build the 3D structure of human a7 nAChR. The generated models were assessed using Ramachandran plot. The 3SN6 and 2XYT based model structures had shown 99.1% and 99.2% residues in the allowed and favored regions in the Ramachandran plots. The initial models were subject to 100 ns molecular dynamics simulations to refine the side chain orientations and to relax the loop regions. The final frame from the trajectory file was selected as a representative structure. The refined models were better than the initial models in the Ramachandran plots and were used for subsequent molecular docking studies. Docking is a well-established method for assessing the binding potential of chemicals to the target receptor in the body. This technique has been used for predicting potential binders of nAChRs [60–64]. However, protein conformation flexibility remains an issue due to the rigid-body treatment of the protein structure during the rigid docking process, the most common docking protocol. A flexible docking protocol that allows changes in protein conformation is very expensive to implement. Application of flexible docking in a big number of molecules such as high-throughput screening is difficult, if not impossible, and currently is not a popular docking protocol due to the challenges of conformational sampling and energy weighting. Different methods have been developed to incorporate protein flexibility into molecular docking through balancing the computational cost and the conformational sampling space; for example, the flexible docking protocol uses experimentally-determined conformations to reduce conformation sampling space and to guide weighting energy of protein conformations. The CDM developed in this study was constructed in a similar way, taking into consideration the conformational flexibility inherent in proteins by incorporating the different protein (or more importantly, pocket) conformations most likely to be adopted. It is worth noting that issues related to potential false positive predictions are often found when using models based on the docking scores of CDM. These models utilize homology model structures based on docking scores used to measure the fitness of ligands in a binding pocket. Homology model structure are also determined by the orientation, shape, and energetic

Table 3: Potential key binding interactions obtained from MD simulations performed on the docked complexes of the training set ligands

Interactions	Protein	Residues
Hydrophobic contact	A Chain	Leu119
	B Chain	Leu119, Trp149, Tyr195
Pi-pi	B Chain	Trp149, Tyr188
H-bonding	A Chain	Leu119
Pi-cation	A Chain	Arg79, Tyr188, Tyr195

Table 4: Binding activity prediction results of the test set ligands. A positive CDM prediction score indicates that the ligand is a human α7 nAChR-LBD binder while a negative score indicates otherwise

Ligands	XPGScore		Winning score	CDM
	2XYT-based α7 nAChR-LBD	3SQ6-based α7 nAChR-LBD		
Acetylcholine	-4.68	-4.812	-4.812	0.132
epibatidine	-5.2	-8.725	-8.725	1.000
nicotine	-3.1	-7.512	-7.512	0.731
cytisine	-3.65	-6.362	-6.362	0.476
PNU282987	-4.518	-5.878	-5.878	0.368
mecamylamine	-3.564	-4.873	-4.873	0.145
MG624	-4.954	failed to dock	-4.954	0.163
methyllycaconitine	-5.829	failed to dock	-5.829	0.357
Hexamethonium	-2.786	-1.001	-2.786	-0.318
NS1738	-7.108	-8.513	-8.513	0.953
PNU120596	-4.992	-6.188	-6.188	0.437
SB206553	-4.373	-6.671	-6.671	0.544

interactions of small molecules with the receptor with an approximation method for binding energy.

Clustering analysis of the templates shortlisted from the PDB was a key step in constructing 3D structures of human α7 nAChR in accounting for the protein flexibility in CDM. While structure 3SQ6 in PDB appeared to be the obvious template of choice for constructing human α7 nAChR-LBD, a second human α7 nAChR-LBD constructed based on structure 2XYT (despite the lower percentage sequence identity) demonstrated its usefulness during the docking of TC9, BS1 and BS2 (i.e., ligands that failed to dock to the 3SQ6-based human α7 nAChR-LBD but were known to bind to human α7 nAChR). This clearly showed the merit of using the CDM in the assessment of chemical binding potential to human α7 nAChR-LBD. More favorable scores were often observed in the 3SQ6-based rather than the 2XYT-based human α7 nAChR-LBD structures for majority of the ligands. This was not surprising, as the narrower binding pocket of the 3SQ6-

based human α7 nAChR-LBD allowed the ligands to fit more snugly to the pocket (thus forming more interactions) compared to the 2XYT-based structure, which had a wider binding pocket opening.

Comparison of the residues in the active site of the human α7 nAChR structures that involved interactions with ligands revealed that 14 and 23 residues from Chain A and B interacted with the ligands. The interacting residues in Chain A and B are 148S, 149W, 150S, 151Y, 152G, 186R, 187F, 188Y, 190C, 191C, 192K, 193E, 194P, 195Y and 32Y, 33F, 34S, 55W, 56L, 57Q, 58M, 59S, 60W, 77T, 79R, 107W, 108V, 109L, 110V, 111N, 115H, 116C, 117Q, 118Y, 119L, 120P, 121P, respectively.

Binding interaction between nAChRs and some ligands were investigated experimentally [65, 72]. Apart from developing the CDM, this study also investigated potential key binding interactions between human α7 nAChR-LBD and the ligands in the training set. It was

found that most ligands were anchored in the binding pocket of human α7 nAChR-LBD through rather non-specific hydrophobic contacts with a range of hydrophobic residues in the binding pocket. Residues Trp149 and Tyr195 in the B chain were found to interact strongly (i.e., observed through most of the frames extracted from the MD trajectories) with many of these ligands. Trp149 has been found to interact with ligands through hydrogen bonding and cation–π interactions in literature [66]. Leu119 in chain A was also found to interact with many of these ligands but with apparent weaker contacts. Considering that aromatic rings were contained in many residues of the binding pocket and most of the ligands, the formation of pi-pi interactions between human α7 nAChR-LBD and the ligands was expected. Experimental data confirmed binding interaction of Leu119 with large agonists of α7 [67]. Trp149 and Tyr188 in chain B were found to form rather stable pi-pi interactions with the ligands which were confirmed with the findings in the literature [68, 69, 71]. Similarly, polar interactions such as hydrogen bonding and pi-cation interactions were found to play an important role in receptor-ligand binding. In general, the hydrogen bonds observed were strong and were maintained in most of the trajectories. Similar with the hydrogen bonding interactions, the pi-cation interactions observed between the receptor (Arg79, Tyr188 and Tyr195 of chain A) and the ligands (BS1, ZY5, ZY7, and AN4) were also fair strong. While in the case of BS1, the positive charge originated from the ligand, the opposite was true for AN4, ZY5 and ZY7 (i.e., the aromatic rings from these ligands interacted with the positive charge from Arg79). With ZY5 and ZY7, the pi-cation interaction with Arg79 could be attributable to the electron-donating double hydroxyl/methoxy substituents of one of the aromatic rings. The mutation study demonstrated that Arg79 involved binding interaction with ligands of α7 [70].

MATERIALS AND METHODS

Overall modeling scheme

The overall modeling scheme of this study is depicted in Figure 1. First, the PDB was searched for suitable protein-ligand complexes to be used as potential templates to construct the 3D human α7 nAChR-LBD structures. During the search, only complexes containing ligands with known human α7 nAChR binding data were selected (see Table 1). These complexes were then subjected to a binding pocket analysis to examine the similarity or difference among the ligand binding pockets. Based on this analysis, two homology structures of human α7 nAChR-LBD were constructed. The CDM was developed using these 3D structures and the crystal-bound ligands (training set; see Figure 2A). The key binding interactions between the ligands and the 3D human α7 nAChR-LBD structures were examined with

MD simulations. The model was then used to predict the binding activity of a series of test compounds (see Figure 2B) and the model prediction results were compared with the experimentally-obtained findings to assess model performance.

Binding pocket analysis

Seventeen receptor-ligand complexes were obtained through the PDB search. The residues lining these ligand binding pockets were identified. Taking the PDB crystal structure 1UW6 as an example, the residues of its ligand binding pocket are: Asp86, Ala88, Tyr90, Lys140, Gly142, Ser143, Trp144, Thr145, His146, His147, Val184, Thr185, Tyr186, Ser187, Cys188, Cys189, Glu191, Ala192, Tyr193, Glu194, Lys241, Ile243, Trp260, Gln261, Gln262, Thr263, Gln280, Ser282, Thr306, Pro307, Leu309, Ala310, Arg311, Val312, Val318, Leu319, Tyr320, Met321, Pro322, Ser323, and Tyr371. Based on these residues, 17 3D structures of the α7 nAChR complexes with ligands were aligned and the root-mean-squared deviations (RMSDs) of these residues among the 17 ligand binding pockets were calculated using Pymol (https://www.pymol.org/). The receptor-ligand complexes were clustered according to their binding pocket residue RMSDs using the *hclust* tool (method: ward) in R (https://www.rdocumentation.org/packages/stats/versions/3.4.3/topics/hcl).

Homology modeling of human α7 nAChR-LBD

The clustering results from the binding pocket analysis identified two distinct groups of binding pocket conformations among the 17 receptor-ligand complexes. One structure was selected from each of the two clusters (i.e., PDB IDs: 3SQ6 and 2XYT) as a template to construct the 3D structures of human α7 nAChR-LBD.

A series of homology modeling steps were performed using the Prime structure prediction wizard (http://www.schrodinger.com/Prime/) within Maestro (http://www.schrodinger.com/Maestro/) as follows. First, the primary sequences of the target protein and the template structure were compared. Sequence alignment was performed using ClustalW (http://www.ch.embnet.org/software/ClustalW.html) between the primary sequences of the two subunits (chain A and chain B) of human α7 nAChR (UniProt ID: P36544) and the template proteins 3SQ6 and 2XYT. The model was built based on the alignment between the target and template sequence with the ligands (EPJ from 3SQ6 and TC9 from 2XYT), through the knowledge-based and heteromultimer methods. The heteromultimer helps to build the model with more than one chains at a single run. There are four steps in the homology model construction. Initially the PRIME module copied the backbone atoms of the aligned region and the side chains of the conserved residues. Subsequently, the side chain was optimized and the non-template residues were minimized. Finally, the

missing residues in the alignment file were inserted. The constructed human α7 nAChR-LBD structures based on 3SQ6 and 2XYT were minimized using the OPLS2005 force field. The minimized models were assessed using Ramachandran plots for the structural integrity and steric hindrance of the resulting.

Structure refinement through MD simulations

The two 3D human α7 nAChR-LBD structures from homology modeling were subjected to MD simulations for structural optimization to remove any potential atomic clashes that might arise during construction. MD simulations were performed using Desmond from DE Shaw (https://www.deshawresearch.com/resources_desmond.html). OPLS2005 force field was used for both the proteins and ligands. In DESMOND, OPLS2005 has enough parameters to parametrize the ligands. The complex structures with OPLS force field and the DESMOND configuration file (parameter file) were used to run the molecular dynamics simulations. His residues in the modelled human α7 were single protonated. MD simulation system—consisting of an orthorhombic box with borders of 10Å from the receptor-ligand complex, containing simple point charge water molecules, and having a Na+ and Cl- salt concentration of 0.15M—was constructed using System Builder. Prior to production simulations, both receptor-ligand systems were subjected to a relaxation stage using the default setting as follows: two steps of energy minimization of the protein–ligand complexes with and without restraint on the solute; 12 picosecond (ps) of constant number, volume, and temperature ensemble simulation carried out at 10 kelvin (K) with a Berendsen barostat and a fast temperature relaxation constant, velocity resampling set at every 1 ps, and restraints applied on all heavy atoms of the protein-ligand complex; 12 ps of constant number, pressure, and temperature (NPT) ensemble simulation carried out at 10 K, 1 atmospheric pressure (atm) using a Berendsen thermostat and barostat with a fast temperature relaxation constant to heat the system from 10 K to 300 K, and a slow pressure relaxation constant, velocity resampling set at every 1 ps, and restraints applied on all heavy atoms of the protein–ligand complexes; 12 ps of NPT simulation carried out at 300 K, 1 atm using the same thermostat, barostat, temperature relaxation constants, pressure relaxation constants, velocity resampling steps, and restraints; and finally, 24 ps of NPT simulation at 300 K, 1 atm using a fast temperature, and normal pressure relaxation constants. After this, production MD simulations were performed under NPT condition, 300 K, and pressure of 1 atm for 100 nanoseconds (ns) with energy being recorded at every 5 ps and trajectory at every 20 ps. The structural integrity of the final receptor structures obtained at the end of the MD simulations was checked using the Ramachandran plot.

Preparation of ligands for docking

Two sets of ligands—training and test sets—were used in this study. The training set consisted of ligands extracted from the 17 receptor-ligand complexes obtained from the PDB. The ligands were imported to Maestro and checked for correct bond order and type before hydrogen atoms were added. The ligands in the test set were downloaded from PubChem (https://pubchem.ncbi.nlm.nih.gov/) in 3D SDF format. All ligands were protonated at pH 7.

Preparation of human α7 nAChR-LBD structures for docking

A docking grid was generated around the binding pocket for the two human α7 nAChR-LBD structures constructed from homology modeling and refined by MD simulations. For the 3SQ6-based human α7 nAChR-LBD, the centroid of the ligand (EPJ) was set as the center of the grid and the size of the grid box was set to be suitable for ligands with a length of 10 Å. For the 2XYT-based human α7 nAChR-LBD, the centroid of the ligand (TC9) was set as the center of the grid, while the size of the grid box was set to be suitable for ligands with a length of 17 Å. Compared the residues in the active site between the two human α7 homology modeled protein to identify the conformation of the residues due to the ligand binding.

Competitive docking model (CDM)

The ligands in the training set were docked to the grid boxes of the two 3D human α7 nAChR-LBD structures using Glide [72–74] in Maestro (http://www.schrodinger.com/Glide). The extra precision docking method was used while the rest of the docking parameters were maintained as default: van der Waals radii scaling factor of 0.80, partial charge cutoff of 0.15, flexible ligand sampling with regard to nitrogen inversions and ring conformations, and Epik state penalties applied to the docking scores. The number of ligand poses to include for post-docking minimization was set at five, and only one pose per ligand was written out. Three possible scenarios were anticipated from the docking of a ligand to the two human α7 nAChR-LBD structures: (1) the ligand was successfully docked to only one of the two structures: the 3SQ6-based or the 2XYT-based human α7 nAChR-LBD structure; (2) the ligand was successfully docked to both structures; and (3) the ligand could not be docked to either structure. In scenario (1), the ligand would be considered to bind human α7 nAChR in the conformation similar to the 3SQ6-based or 2XYT-based human α7 nAChR-LBD structure; therefore, the associated docking score would be used for the CDM development. In scenario (2), the ligand would be considered to bind the human α7 nAChR-LBD structure whose docking score was the lower of the

two (more favorable), as obtained from the ligand docking to the 3SQ6-based and 2XYT-based human α7 nAChR-LBD structures, and the lower docking score would be used in the CDM. In scenario (3), the ligand would be considered as a non-binder of human α7 nAChR. Based on this rule, each ligand from the training set would be associated with a favored docked complex and docking score (or none if it was a non-binder). The predictive CDM was developed using these docking scores according to equation (2), below:

$$CDM = 1 - \frac{score - score\ lowest}{score\ highest - score\ lowest} \qquad (2)$$

"Score" indicates the winning docking score (the lower docking score between the two human α7 nAChR for the same chemical), while "score highest" and "score lowest" are the highest and lowest winning scores among the 17 scores. When the CDM value calculated from docking is less than zero, the chemical is predicted to be a non-binder of human α7 nAChR; the chemical is predicted to be a binder if its CDM value is greater than or equal to zero.

Elucidation of key binding interactions through MD simulations

In order to define the potential key binding interactions between human α7 nAChR-LBD and its binding ligands, the 17 complexes obtained from the CMD were subjected to MD simulations. The MD simulations were performed for each complex in a manner similar to the human α7 nAChR-LBD structure-refinement step. First, a relaxation stage identical to what was described in 3D structure optimization was performed; next, a production of 20 ns of MD simulations was performed. All simulation parameters remained unchanged. To assess simulation stability, the RMSD of the 20 ns simulations was plotted against time. The key binding interactions (i.e., hydrophobic contact, pi-pi interactions, pi-cation and hydrogen bonds, root-mean-squared fluctuations [RMSFs] of the residues) were investigated using the frames obtained from the last ns of the MD simulations. These analyses were performed using the simulation interactions diagram tool in Maestro (http://www.schrodinger.com/Maestro/).

Binding activity predictions of the compounds for experimental validation

After development of the CDM, the compounds in the test set were docked to the two 3D human α7 nAChR-LBD structures, and their respective favored docking scores were used to make predictions, based on equation (1), of the binding activity of the respective compounds.

Experimental validation

The CDM performance in the prediction of ligand binding activity was assessed by comparing the prediction results with the findings obtained from experimental testing.

A Center for Tobacco Products (CTP) sent 12 compounds for blind testing of the CDM. The 12 compounds were assayed for their activity to the human α7 nAChRs subtype in a CTP-contracted project in which 786 compounds were profiled in five human nicotinic subtypes including the α7 [75]. The results were sent to the CTP scientist for assessing CDM performance.

Abbreviations

AChBP: Acetylcholine binding protein; Atm: Atmospheric pressure; CDM: Competitive docking; model; HPHCs: Harmful and potentially harmful constituents; LBD: Ligand-binding domain; MD: Molecular dynamics; nAChRs: Nicotinic acetylcholine receptors; ns: Nano second; ps: Pico second; RMSDs: Root mean squared deviation; RMSFs: Root mean squared fluctuations.

Author contributions

HH and CL designed and led the project. HWN and SS collected the data and implemented the methods. HWN, SS, BP, HY, LW, SC, WT, LV and HH discussed the data analysis and the results. HWN, HH and CL wrote the manuscript.

CONFLICTS OF INTEREST

The authors have no conflicts of interest to declare.

(FDA). This publication represents the views of the author(s) and does not represent FDA/CTP position or policy.

REFERENCES

1. Centers for Disease Control and Prevention. (2010) A Report of the Surgeon General: How Tobacco Smoke Causes Disease: What It Meands to You. https://www.cdc.gov/tobacco/data_statistics/sgr/2010/consumer_booklet/pdfs/consumer.pdf (Accessed on July 16, 2017).

2. National Center for Chronic Disease Prevention, Health Promotion Office on Smoking, Health. Reports of the Surgeon General. The Health Consequences of Smoking-50 Years of Progress: A Report of the Surgeon General. (Atlanta, GA). US: Centers for Disease Control and Prevention; 2014.

3. Jain P, Danaei G, Robins JM, Manson JE, Hernán MA. Smoking cessation and long-term weight gain in the Framingham Heart Study: an application of the parametric g-formula for a continuous outcome. Eur J Epidemiol. 2016; 31:1223–29. https://doi.org/10.1007/s10654-016-0200-4.

4. Borland R, Pierce JP, Burns DM, Gilpin E, Johnson M, Bal D. Protection from environmental tobacco smoke in California. The case for a smoke-free workplace. JAMA. 1992; 268:749–52. https://doi.org/10.1001/jama.1992.03490060081028.

5. Fontham ET, Correa P, Reynolds P, Wu-Williams A, Buffler PA, Greenberg RS, Chen VW, Alterman T, Boyd P, Austin DF, Liff J. Environmental tobacco smoke and lung cancer in nonsmoking women. A multicenter study. JAMA. 1994; 271:1752–59. https://doi.org/10.1001/jama.1994.03510460044031.

6. Stockwell HG, Goldman AL, Lyman GH, Noss CI, Armstrong AW, Pinkham PA, Candelora EC, Brusa MR. Environmental tobacco smoke and lung cancer risk in nonsmoking women. J Natl Cancer Inst. 1992; 84:1417–22. https://doi.org/10.1093/jnci/84.18.1417.

7. Pirkle JL, Flegal KM, Bernert JT, Brody DJ, Etzel RA, Maurer KR. Exposure of the us population to environmental tobacco smoke: the third national health and nutrition examination survey, 1988 to 1991. JAMA. 1996; 275:1233–40. https://doi.org/10.1001/jama.1996.03530400021033.

8. Yeramaneni S, Dietrich KN, Yolton K, Parsons PJ, Aldous KM, Haynes EN. Secondhand tobacco smoke exposure and neuromotor function in rural children. J Pediatr. 2015; 167:253–9.e1. https://doi.org/10.1016/j.jpeds.2015.03.014.

9. Tran TT, Yiengprugsawan V, Chinwong D, Seubsman SA, Sleigh A. Environmental tobacco smoke exposure and health disparities: 8-year longitudinal findings from a large cohort of Thai adults. BMC Public Health. 2015; 15:1217. https://doi.org/10.1186/s12889-015-2547-y.

10. Albuquerque EX, Pereira EF, Alkondon M, Rogers SW. Mammalian nicotinic acetylcholine receptors: from structure to function. Physiol Rev. 2009; 89:73–120. https://doi.org/10.1152/physrev.00015.2008.

11. Karlin A. Emerging structure of the nicotinic acetylcholine receptors. Nat Rev Neurosci. 2002; 3:102–14. https://doi.org/10.1038/nrn731.

12. Anand R, Conroy WG, Schoepfer R, Whiting P, Lindstrom J. Neuronal nicotinic acetylcholine receptors expressed in Xenopus oocytes have a pentameric quaternary structure. J Biol Chem. 1991; 266:11192–98.

13. Cooper E, Couturier S, Ballivet M. Pentameric structure and subunit stoichiometry of a neuronal nicotinic acetylcholine receptor. Nature. 1991; 350:235–38. https://doi.org/10.1038/350235a0.

14. Elgoyhen AB, Johnson DS, Boulter J, Vetter DE, Heinemann S. Alpha 9: an acetylcholine receptor with novel pharmacological properties expressed in rat cochlear hair cells. Cell. 1994; 79:705–15. https://doi.org/10.1016/0092-8674(94)90555-X.

15. Nomikos GG, Schilström B, Hildebrand BE, Panagis G, Grenhoff J, Svensson TH. Role of alpha7 nicotinic receptors in nicotine dependence and implications for psychiatric illness. Behav Brain Res. 2000; 113:97–103. https://doi.org/10.1016/S0166-4328(00)00204-7.

16. Mansvelder HD, Keath JR, McGehee DS. Synaptic mechanisms underlie nicotine-induced excitability of brain reward areas. Neuron. 2002; 33:905–19. https://doi.org/10.1016/S0896-6273(02)00625-6.

17. De Luca V, Wong AH, Muller DJ, Wong GW, Tyndale RF, Kennedy JL. Evidence of association between smoking and alpha7 nicotinic receptor subunit gene in schizophrenia patients. Neuropsychopharmacology. 2004; 29:1522–26. https://doi.org/10.1038/sj.npp.1300466.

18. Li ZZ, Guo ZZ, Zhang Z, Cao QA, Zhu YJ, Yao HL, Wu LL, Dai QY. Nicotine-induced upregulation of VCAM-1, MMP-2, and MMP-9 through the α7-nAChR-JNK pathway in RAW264.7 and MOVAS cells. Mol Cell Biochem. 2015; 399:49–58. https://doi.org/10.1007/s11010-014-2231-z.

19. Lam DC, Girard L, Ramirez R, Chau WS, Suen WS, Sheridan S, Tin VP, Chung LP, Wong MP, Shay JW, Gazdar AF, Lam WK, Minna JD. Expression of nicotinic acetylcholine receptor subunit genes in non-small-cell lung cancer reveals differences between smokers and nonsmokers. Cancer Res. 2007; 67:4638–47. https://doi.org/10.1158/0008-5472.CAN-06-4628.

20. Rodgman A, Perfetti TA. The chemical components of tobacco and tobacco smoke. Boca Raton (FL): CRC Press; 2013. https://doi.org/10.1201/b13973.

21. Neamati N, Hong H, Owen JM, Sunder S, Winslow HE, Christensen JL, Zhao H, Burke TR Jr, Milne GW, Pommier Y. Salicylhydrazine-containing inhibitors of HIV-1 integrase: implication for a selective chelation in the integrase active site. J Med Chem. 1998; 41:3202–09. https://doi.org/10.1021/jm9801760.

22. Hong H, Neamati N, Winslow HE, Christensen JL, Orr A, Pommier Y, Milne GW. Identification of HIV-1 integrase inhibitors based on a four-point pharmacophore.

Antivir Chem Chemother. 1998; 9:461–72. https://doi.org/10.1177/095632029800900602.

23. Neamati N, Hong H, Sunder S, Milne GW, Pommier Y. Potent inhibitors of human immunodeficiency virus type 1 integrase: identification of a novel four-point pharmacophore and tetracyclines as novel inhibitors. Mol Pharmacol. 1997; 52:1041–55. https://doi.org/10.1124/mol.52.6.1041.

24. Hong H, Neamati N, Wang S, Nicklaus MC, Mazumder A, Zhao H, Burke TR Jr, Pommier Y, Milne GW. Discovery of HIV-1 integrase inhibitors by pharmacophore searching. J Med Chem. 1997; 40:930–36. https://doi.org/10.1021/jm960754h.

25. Hong H, Fang H, Xie Q, Perkins R, Sheehan DM, Tong W. Comparative molecular field analysis (CoMFA) model using a large diverse set of natural, synthetic and environmental chemicals for binding to the androgen receptor. SAR QSAR Environ Res. 2003; 14:373–88. https://doi.org/10.1080/10629360310001623962.

26. Hong H, Tong W, Fang H, Shi L, Xie Q, Wu J, Perkins R, Walker JD, Branham W, Sheehan DM. Prediction of estrogen receptor binding for 58,000 chemicals using an integrated system of a tree-based model with structural alerts. Environ Health Perspect. 2002; 110:29–36. https://doi.org/10.1289/ehp.0211029.

27. Ng HW, Doughty SW, Luo H, Ye H, Ge W, Tong W, Hong H. Development and validation of decision forest model for estrogen receptor binding prediction of chemicals using large data sets. Chem Res Toxicol. 2015; 28:2343–51. https://doi.org/10.1021/acs.chemrestox.5b00358.

28. Tong W, Hong H, Fang H, Xie Q, Perkins R. Decision forest: combining the predictions of multiple independent decision tree models. J Chem Inf Comput Sci. 2003; 43:525–31. https://doi.org/10.1021/ci020058s.

29. Hong H, Tong W, Xie Q, Fang H, Perkins R. An in silico ensemble method for lead discovery: decision forest. SAR QSAR Environ Res. 2005; 16:339–47. https://doi.org/10.1080/10659360500203022.

30. Xie Q, Ratnasinghe LD, Hong H, Perkins R, Tang ZZ, Hu N, Taylor PR, Tong W. Decision forest analysis of 61 single nucleotide polymorphisms in a case-control study of esophageal cancer; a novel method. BMC Bioinformatics. 2005; 6:S4. https://doi.org/10.1186/1471-2105-6-S2-S4.

31. Hong H, Tong W, Perkins R, Fang H, Xie Q, Shi L. Multiclass Decision Forest—a novel pattern recognition method for multiclass classification in microarray data analysis. DNA Cell Biol. 2004; 23:685–94. https://doi.org/10.1089/dna.2004.23.685.

32. Shen J, Xu L, Fang H, Richard AM, Bray JD, Judson RS, Zhou G, Colatsky TJ, Aungst JL, Teng C, Harris SC, Ge W, Dai SY, et al. EADB: an estrogenic activity database for assessing potential endocrine activity. Toxicol Sci. 2013; 135:277–91. https://doi.org/10.1093/toxsci/kft164.

33. Hong H, Shen J, Ng HW, Sakkiah S, Ye H, Ge W, Gong P, Xiao W, Tong W. A rat alpha-fetoprotein binding activity

prediction model to facilitate assessment of the endocrine disruption potential of environmental chemicals. Int J Environ Res Public Health. 2016; 13:372. https://doi.org/10.3390/ijerph13040372.

34. Hong H, Hong Q, Perkins R, Shi L, Fang H, Su Z, Dragan Y, Fuscoe JC, Tong W. The accurate prediction of protein family from amino acid sequence by measuring features of sequence fragments. J Comput Biol. 2009; 16:1671–88. https://doi.org/10.1089/cmb.2008.0115.

35. Liu J, Mansouri K, Judson RS, Martin MT, Hong H, Chen M, Xu X, Thomas RS, Shah I. Predicting hepatotoxicity using ToxCast in vitro bioactivity and chemical structure. Chem Res Toxicol. 2015; 28:738–51. https://doi.org/10.1021/tx500501h.

36. McPhail B, Tie Y, Hong H, Pearce BA, Schnackenberg LK, Ge W, Valerio LG Jr, Fuscoe JC, Tong W, Buzatu DA, Wilkes JG, Fowler BA, Demchuk E, Beger RD. Modeling chemical interaction profiles: I. Spectral data-activity relationship and structure-activity relationship models for inhibitors and non-inhibitors of cytochrome P450 CYP3A4 and CYP2D6 isozymes. Molecules. 2012; 17:3383–406. https://doi.org/10.3390/molecules17033383.

37. Tie Y, McPhail B, Hong H, Pearce BA, Schnackenberg LK, Ge W, Buzatu DA, Wilkes JG, Fuscoe JC, Tong W, Fowler BA, Beger RD, Demchuk E. Modeling chemical interaction profiles: II. Molecular docking, spectral data-activity relationship, and structure-activity relationship models for potent and weak inhibitors of cytochrome P450 CYP3A4 isozyme. Molecules. 2012; 17:3407–60. https://doi.org/10.3390/molecules17033407.

38. Luo H, Ye H, Ng HW, Shi L, Tong W, Mendrick DL, Hong H. Machine learning methods for predicting HLA-peptide binding activity. Bioinform Biol Insights. 2015; 9:21–29.

39. Ng HW, Shu M, Luo H, Ye H, Ge W, Perkins R, Tong W, Hong H. Estrogenic activity data extraction and in silico prediction show the endocrine disruption potential of bisphenol A replacement compounds. Chem Res Toxicol. 2015; 28:1784–95. https://doi.org/10.1021/acs.chemrestox.5b00243.

40. Ng HW, Zhang W, Shu M, Luo H, Ge W, Perkins R, Tong W, Hong H. Competitive molecular docking approach for predicting estrogen receptor subtype α agonists and antagonists. BMC Bioinformatics. 2014; 15:S4. https://doi.org/10.1186/1471-2105-15-S11-S4.

41. Drake RR, Neamati N, Hong H, Pilon AA, Sunthankar P, Hume SD, Milne GW, Pommier Y. Identification of a nucleotide binding site in HIV-1 integrase. Proc Natl Acad Sci USA. 1998; 95:4170–75. https://doi.org/10.1073/pnas.95.8.4170.

42. Luo H, Du T, Zhou P, Yang L, Mei H, Ng H, Zhang W, Shu M, Tong W, Shi L, Mendrick DL, Hong H. Molecular docking to identify associations between drugs and class I human leukocyte antigens for predicting idiosyncratic drug reactions. Comb Chem High Throughput Screen. 2015; 18:296–304. https://doi.org/10.2174/1386207318666150305144015.

43. Ng HW, Perkins R, Tong W, Hong H. Versatility or promiscuity: the estrogen receptors, control of ligand selectivity and an update on subtype selective ligands. Int J Environ Res Public Health. 2014; 11:8709–42. https://doi.org/10.3390/ijerph110908709.

44. Shu M, Ng HW, Orr M, Luo H, Ye H, Ge W, Tong W, Hong H. Homology model and ligand binding interactions of human α4β2 nicotinic acetylcholine receptor. J Biomed Sci Eng. 2016; 9:41–100. https://doi.org/10.4236/jbise.2016.91005.

45. Ding D, Xu L, Fang H, Hong H, Perkins R, Harris S, Bearden ED, Shi L, Tong W. The EDKB: an established knowledge base for endocrine disrupting chemicals. BMC Bioinformatics. 2010; 11:S5. https://doi.org/10.1186/1471-2105-11-S6-S5.

46. Ng HW, Laughton CA, Doughty SW. Molecular dynamics simulations of the adenosine A2a receptor: structural stability, sampling, and convergence. J Chem Inf Model. 2013; 53:1168–78. https://doi.org/10.1021/ci300610w.

47. Ng HW, Laughton CA, Doughty SW. Molecular dynamics simulations of the adenosine A2a receptor in POPC and POPE lipid bilayers: effects of membrane on protein behavior. J Chem Inf Model. 2014; 54:573–81. https://doi.org/10.1021/ci400463z.

48. Compadre CM, Singh A, Thakkar S, Zheng G, Breen PJ, Ghosh S, Kiaei M, Boerma M, Varughese KI, Hauer-Jensen M. Molecular dynamics guided design of tocoflexol: a new radioprotectant tocotrienol with enhanced bioavailability. Drug Dev Res. 2014; 75:10–22. https://doi.org/10.1002/ddr.21162.

49. Ayers JT, Dwoskin LP, Deaciuc AG, Grinevich VP, Zhu J, Crooks PA. bis-Azaaromatic quaternary ammonium analogues: ligands for alpha4beta2* and alpha7* subtypes of neuronal nicotinic receptors. Bioorg Med Chem Lett. 2002; 12:3067–71. https://doi.org/10.1016/S0960-894X(02)00687-X.

50. Bolchi C, Valoti E, Binda M, Fasoli F, Ferrara R, Fumagalli L, Gotti C, Matucci R, Vistoli G, Pallavicini M. Design, synthesis and binding affinity of acetylcholine carbamoyl analogues. Bioorg Med Chem Lett. 2013; 23:6481–85. https://doi.org/10.1016/j.bmcl.2013.09.023.

51. Ivy Carroll F, Yokota Y, Ma W, Lee JR, Brieaddy LE, Burgess JP, Navarro HA, Damaj MI, Martin BR. Synthesis, nicotinic acetylcholine receptor binding, and pharmacological properties of 3′-(substituted phenyl) deschloroepibatidine analogs. Bioorg Med Chem. 2008; 16:746–54. https://doi.org/10.1016/j.bmc.2007.10.027.

52. Jensen AA, Zlotos DP, Liljefors T. Pharmacological characteristics and binding modes of caracurine V analogues and related compounds at the neuronal alpha7 nicotinic acetylcholine receptor. J Med Chem. 2007; 50:4616–29. https://doi.org/10.1021/jm070574f.

53. Bodnar AL, Cortes-Burgos LA, Cook KK, Dinh DM, Groppi VE, Hajos M, Higdon NR, Hoffmann WE, Hurst RS, Myers JK, Rogers BN, Wall TM, Wolfe ML, Wong E. Discovery and structure-activity relationship of quinuclidine benzamides as agonists of alpha7 nicotinic acetylcholine receptors. J Med Chem. 2005; 48:905–08. https://doi.org/10.1021/jm049363q.

54. Baker SR, Boot J, Brunavs M, Dobson D, Green R, Hayhurst L, Keenan M, Wallace L. High affinity ligands for the alpha7 nicotinic receptor that show no cross-reactivity with the 5-HT3 receptor. Bioorg Med Chem Lett. 2005; 15:4727–30. https://doi.org/10.1016/j.bmcl.2005.07.070.

55. Faghih R, Gopalakrishnan M, Briggs CA. Allosteric modulators of the alpha7 nicotinic acetylcholine receptor. J Med Chem. 2008; 51:701–12. https://doi.org/10.1021/jm070256g.

56. Capelli AM, Castelletti L, Salvagno C, Oliosi B, Di Lenarda E, Virginio C, Lightfoot A, Kew JN, Teague S. Identification of novel alpha7 nAChR positive allosteric modulators with the use of pharmacophore in silico screening methods. Bioorg Med Chem Lett. 2010; 20:4561–65. https://doi.org/10.1016/j.bmcl.2010.06.014.

57. Mazurov AA, Speake JD, Yohannes D. Discovery and development of α7 nicotinic acetylcholine receptor modulators. J Med Chem. 2011; 54:7943–61. https://doi.org/10.1021/jm2007672.

58. Cheng X, Ivanov I, Wang H, Sine SM, McCammon JA. Molecular-dynamics simulations of ELIC-a prokaryotic homologue of the nicotinic acetylcholine receptor. Biophys J. 2009; 96:4502–13. https://doi.org/10.1016/j.bpj.2009.03.018.

59. Cheng X, Wang H, Grant B, Sine SM, McCammon JA. Targeted molecular dynamics study of C-loop closure and channel gating in nicotinic receptors. PLOS Comput Biol. 2006; 2:e134. https://doi.org/10.1371/journal.pcbi.0020134.

60. Kombo DC, Mazurov AA, Chewning J, Hammond PS, Tallapragada K, Hauser TA, Speake J, Yohannes D, Caldwell WS. Discovery of novel α7 nicotinic acetylcholine receptor ligands via pharmacophoric and docking studies of benzylidene anabaseine analogs. Bioorg Med Chem Lett. 2012; 22:1179–86. https://doi.org/10.1016/j.bmcl.2011.11.090.

61. Kombo DC, Mazurov A, Tallapragada K, Hammond PS, Chewning J, Hauser TA, Vasquez-Valdivieso M, Yohannes D, Talley TT, Taylor P, Caldwell WS. Docking studies of benzylidene anabaseine interactions with α7 nicotinic acetylcholine receptor (nAChR) and acetylcholine binding proteins (AChBPs): application to the design of related α7 selective ligands. Eur J Med Chem. 2011; 46:5625–35. https://doi.org/10.1016/j.ejmech.2011.09.033.

62. Kombo DC, Bencherif M. Comparative study on the use of docking and Bayesian categorization to predict ligand binding to nicotinic acetylcholine receptors (nAChRs) subtypes. J Chem Inf Model. 2013; 53:3212–22. https://doi.org/10.1021/ci400493a.

63. Spurny R, Debaveye S, Farinha A, Veys K, Vos AM, Gossas T, Atack J, Bertrand S, Bertrand D, Danielson UH, Tresadern G, Ulens C. Molecular blueprint of allosteric binding sites in a homologue of the agonist-binding domain

of the α7 nicotinic acetylcholine receptor. Proc Natl Acad Sci USA. 2015; 112:E2543–52. https://doi.org/10.1073/pnas.1418289112.

64. Blum LC, van Deursen R, Bertrand S, Mayer M, Bürgi JJ, Bertrand D, Reymond JL. Discovery of α7-nicotinic receptor ligands by virtual screening of the chemical universe database GDB-13. J Chem Inf Model. 2011; 51:3105–12. https://doi.org/10.1021/ci200410u.

65. Marotta CB, Lester HA, Dougherty DA. An unaltered orthosteric site and a network of long-range allosteric interactions for pnu-120596 in α7 nicotinic acetylcholine receptors. Chem Biol. 2015; 22:1063–73. https://doi.org/10.1016/j.chembiol.2015.06.018.

66. Blum AP, Van Arnam EB, German LA, Lester HA, Dougherty DA. Binding interactions with the complementary subunit of nicotinic receptors. J Biol Chem. 2013; 288:6991–97. https://doi.org/10.1074/jbc.M112.439968.

67. Van Arnam EB, Blythe EE, Lester HA, Dougherty DA. An unusual pattern of ligand-receptor interactions for the α7 nicotinic acetylcholine receptor, with implications for the binding of varenicline. Mol Pharmacol. 2013; 84:201–07. https://doi.org/10.1124/mol.113.085795.

68. Tong M, Arora K, White MM, Nichols RA. Role of key aromatic residues in the ligand-binding domain of alpha7 nicotinic receptors in the agonist action of beta-amyloid. J Biol Chem. 2011; 286:34373–81. https://doi.org/10.1074/jbc.M111.241299.

69. Huang X, Zheng F, Stokes C, Papke RL, Zhan CG. Modeling binding modes of alpha7 nicotinic acetylcholine receptor with ligands: the roles of Gln117 and other residues of the receptor in agonist binding. J Med Chem. 2008; 51:6293–302. https://doi.org/10.1021/jm800607u.

70. Shimomura M, Okuda H, Matsuda K, Komai K, Akamatsu M, Sattelle DB. Effects of mutations of a glutamine residue in loop D of the alpha7 nicotinic acetylcholine receptor on agonist profiles for neonicotinoid insecticides and related ligands. Br J Pharmacol. 2002; 137:162–69. https://doi.org/10.1038/sj.bjp.0704848.

71. Puskar NL, Xiu X, Lester HA, Dougherty DA. Two neuronal nicotinic acetylcholine receptors, alpha4beta4 and alpha7, show differential agonist binding modes. J Biol Chem. 2011; 286:14618–27. https://doi.org/10.1074/jbc.M110.206565.

72. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem. 2004; 47:1739–49. https://doi.org/10.1021/jm0306430.

73. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. J Med Chem. 2006; 49:6177–96. https://doi.org/10.1021/jm051256o.

74. Halgren TA, Murphy RB, Friesner RA, Beard HS, Frye LL, Pollard WT, Banks JL. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. J Med Chem. 2004; 47:1750–59. https://doi.org/10.1021/jm030644s.

75. Fischer M, Coleman RG, Fraser JS, Shoichet BK. Incorporation of protein flexibility and conformational energy penalties in docking screens to improve ligand discovery. Nat Chem. 2014; 6:575–83. https://doi.org/10.1038/nchem.1954.