Interactive computational and experimental approaches improve the sensitivity of periplasmic binding protein-based nicotine biosensors for measurements in biofluids

Nandan Haloi*[1], Shan Huang*[2], Aaron N. Nichols[2], Eve J. Fine[2], Christopher B. Marotta[3], Dennis A. Dougherty[3], Erik Lindahl[1,4], Rebecca J. Howard[1], Stephen L. Mayo[2], Henry A. Lester#[2]

*Co-first authors

Corresponding author

[1]Department of Applied Physics, Science for Life Laboratory, KTH Royal Institute of Technology, 10044 Stockholm, Sweden

[2]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena CA 91125 USA

[3]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena CA 91125 USA

[4]Department of Biochemistry and Biophysics, Science for Life Laboratory, Stockholm University, 10691 Stockholm, Sweden

Abstract (237 / 250 words)

To develop more sensitive fluorescent protein sensors for nicotine, we combined computational protein design, site-saturated, site-directed, and combinatorial mutagenesis with fluorescence assays, molecular dynamics simulations, and absorbance measurements. The data showed that the resulting molecules, iNicSnFR11 and iNicSnFR12, have higher sensitivity to nicotine than previously reported constructs. In the linear portion of the dose-response relation at sub-μM [nicotine] for iNicSnFR12, $\Delta F/F_0$ increased with a proportionality constant (S-slope) of 2.6 μM^{-1}, representing a 6.5-fold higher sensitivity than iNicSnFR3a. Molecular dynamics calculations enabled identification of a binding pose for nicotine previously indeterminate from experimental data. Further comparative simulations based on this model revealed a tilt in helix 4 in the optimized sensor, likely altering allosteric networks involving the ligand binding site. The absorbance data showed that the fluorescence activation results from increased absorption rather than increased quantum yield for fluorescence. iNicSnFR12 resolved nicotine in diluted mouse and human serum at the peak concentration (100-200 nM) that occurs during smoking or vaping, but also at the decaying concentrations (< 100 nM) during the intervals between smoking or vaping sessions. NicSnFR12 was roughly as sensitive to varenicline or acetylcholine as to nicotine; the sensitivity to choline was at least one order of magnitude less. None of these drugs would markedly distort measurements in human biofluids such as sweat and interstitial fluid. Therefore, iNicSnFR12 is a promising candidate as the molecular sensor that could underlie a continuous nicotine monitor for human biofluids.

Introduction

The iNicSnFR series of intensity-based fluorescent nicotine-sensing fluorescent reporters was initially developed to test the hypothesis that nicotine enters the cytoplasm and organelles at the concentrations relevant to the brain of smokers and vapers [1]. This entry allows nicotine to act as a pharmacological chaperone for nascent nicotinic acetylcholine receptors in the endoplasmic reticulum. This process leads to the upregulation thought to play a key role in nicotine dependence [2]. Previous pharmacokinetic measurements on the blood of smokers and vapers indicate that the peak of [nicotine] in the blood is 100-200 nM and in the interval between puffing sessions, [nicotine] declines to concentrations as low as 10 nM [3]. The initial sensors, iNicSnFR3a and iNicSnFR3b, lack the sensitivity for measurements in this range; but the linearity of the sensors allowed us to extrapolate to lower concentrations in experiments on various cultured cell types. We therefore considered the hypothesis verified [1].

Previous pharmacokinetic measurements on the blood of smokers and vapers utilize intravenous catheters followed by HPLC- and mass spectrometry-based assays. These measurements are invasive and tedious. Our goals now include incorporating fluorescent reporters into minimally invasive and non-invasive devices that provide continuous measurement of nicotine pharmacokinetics in human interstitial fluid or sweat during smoking or vaping sessions. This application requires iNicSnFR constructs that do detect nicotine at 10 nM, with a time resolution of ~ 5 min.

Therefore, we sought to improve the sensitivity of the iNicSnFR series. Each member of the iNicSnFR series is a merger of two proteins: a suitably mutated, interrupted, OpuBC periplasmic binding protein, and a circularly permuted GFP (cpGFP). The present fluorescent sensors are related both to the acetylcholine (ACh) sensor iAChSnFR [4] and to several other sensors which we term iDrugSnFRs because the OpuBC binding site been mutated to selectively bind individual central nervous system-acting drugs [1, 5-7]. In all cases but ACh, the drug is a weak base ($6.5 < pKa < 10$), and the charged form of the drug binds to the PBP. Because atomic-scale structural data are available for the entire iNicSnFR3a protein, in both apo and varenicline-

bound state [6], it is now possible to test structure-based hypotheses to improve and visualize the sensitivity.

Here we combined computational protein design with mutagenesis, fluorescence, and absorbance assays. The resulting sensor, iNicSnFR12, is ~ nine-fold more sensitive than previous constructs, allowing measurements in biofluids at concentrations below the 100 nM level. Although diffuse electron density in X-ray data [6](PDB 7S7T) previously vitiated an atomic-level description of nicotine-protein interactions, here we further used molecular simulations to model an experimentally consistent ligand pose, and to propose an allosteric basis for altered sensitivity in iNicSnFR12. Therefore, we report good progress toward the goal of a protein biosensor that can minimally invasive devices to continuously monitor nicotine in the blood of smokers or vapers.

Methods

Mutagenesis and library screening

SSM libraries were generated employing the 22c method [8], with the modification that mixtures of primers, with each primer incorporated with one desired mutation, were used for PCR. The PCR products were gel purified, digested with DpnI, ligated using the Gibson Assembly Master Mix (New England BioLabs), and used to directly transform *E. coli* BL21-Gold (DE3) chemically competent cells (Agilent Technologies, Santa Clara, CA). The expression and screening of iNicSnFR variants were performed in 96-well plate format, and 93 variants per SSM library and 186 variants per combinatorial mutagenesis (CM, see below) library were screened. Individual colonies from iNicSnFR libraries were cultivated in 1 ml of ZYM5052 autoinduction media [9] supplemented with 100 μg/ml ampicillin for 30 h at 30 °C, 250 rpm. Then the cells were harvested (3000 × g, 10 min, 4 ºC), washed with PBS, pH 7.40, and resuspended in 3×PBS, pH 7.40. Resuspended cells were lysed by freezing and thawing using liquid nitrogen and a room-temperature water bath. Intact cells and cell debris were removed by centrifugation at 3500 × g for 30 min at 4°C. The supernatants of lysates were tested with excitation at 485 nm and emission at 535 nm in the presence or absence of a nicotine concentration that is near the EC_{50} of the parent construct for the library being screened. A Spark M10 96-well fluorescence plate reader (Tecan, Männedorf, Switzerland) was used to measure baseline fluorescence (F_0) and nicotine-induced fluorescence (ΔF). Promising clones were amplified and sequenced, and the beneficial mutations were confirmed by the measurement of dose-response relation against nicotine (described below). The optimally responding construct in each round of mutagenesis was used as a template for the next round of SSM, site-directed mutagenesis (SDM), or CM.

Purification of iNicSnFRs

Proteins were expressed in *E. coli* BL21-Gold (DE3) cells using 50 ml ZYM5052 autoinduction media [9]. Cells were collected by centrifugation and stored at –80°C until use. For purification, frozen cell pellets were resuspended in PBS, pH 7.40, and lysed by sonication. Intact cells and cell debris were removed by centrifugation at 15,000 × g for 30 min at 4°C. The

supernatant was collected and loaded onto a prewashed Ni NTA column with wash buffer at 4°C. Ni NTA wash buffer contained 20 mM imidazole in PBS, pH 7.4. Elution was achieved using an imidazole gradient (20–200 mM). Proteins were concentrated by centrifugation through Amicon Ultra 15 filter units (Millipore, Burlington, MA) with a 30-kD cutoff and then dialyzed against 3×PBS, pH 7.40. The dialyzed protein was then subject to dose-response studies to characterize its responses to various drugs.

Fluorescence dose-response relation measurement

Purified biosensors were mixed with drug solutions to yield the mixture of 100 nM biosensor and the drug with desired concentrations. Samples were tested in at least three wells (see individual figures below). All solutions were in 3×PBS, pH 7.4. A Tecan Spark 10M plate reader was used to read the plate with 485 nm excitation and 535 nm emission wavelengths to measure GFP fluorescence (F_0 and ΔF). The resulting data were fitted with the Origin 9.2 software (OriginLabs) to the Hill equation [1, 6]. Human plasma was purchased from Sigma-Aldrich (St. Louis MO)

Absorbance measurements

iNicSnFR12 was purified using the His6 sequence with an AKTÄ Start FPLC (GE Healthcare, Chicago, IL). A 1 mM nicotine solution was made using nicotine tartrate (Thermo Fisher, Waltham, MA) dissolved in 3×PBS, pH 7.0, with subsequent adjustment to pH 7.0. For measurements, 100 µL of 50 µM iNicSnFR12 (in 3×PBS, pH 7.0) was added to a single well of a CELLSTAR 340 µL clear, flat-bottom 96-well plate (Greiner Bio-One, Monroe, NC). 100 µL of 3×PBS, pH 7.0 was added to an empty well to serve as control. The absorption spectra of both the iNicSnFR12 and 3×PBS, pH 7.0 wells were measured between 350 nm and 900 nM using the absorbance scan feature on the Tecan Spark 20M (Männedorf, Switzerland). After the initial and eight subsequent absorbance scans, 1 mM nicotine in 3×PBS was added to both iNicSnFR12 and 3×PBS, pH 7.0 containing wells at volumes of 2.1, 2.1, 4.6, 4.8, 5.4, 6.0, 6.6, and 7.3 µL. After each addition, the absorbance spectra were gathered again. The process of adding nicotine and measuring absorbance spectra was repeated across a total of nine [nicotine]. Concentration adjustments were made for iNicSnFR12 and nicotine based on the dilution by the added solutions.

No adjustment was required for path length. No corrections were made for depletion of nicotine by binding to iNicSnFR12.

Computational search for docking poses of nicotine

A workflow resembling a previous study [10] was developed to systematically search for the optimal binding pose of nicotine in the diffuse electron density near the binding site in the X-ray map of iNicSnFR3a (PDB: 7S7U) [6]. To reduce the search space, we used knowledge of the well-resolved varenicline-iNicSnFR3a interaction (PDB ID: 7S7X). Both ligands share a protonated amine group that is known to form multiple cation-π interactions within the binding site of 7S7X [1, 6]. Assuming that similar interactions are preserved for nicotine as well, we first aligned the protonated amine of nicotine to the secondary amine of varenicline. We defined a vector, from the pyrrolidine amine in nicotine the to a carbon atom of the pyridine ring. We mapped this vector onto a Fibonacci spherical lattice (FSL) [11] of 25 points, generating multiple orientations. During rotation, the amino group defined the center of the FSL, while the carbon atom was arranged on the FSL. Each of the 25 generated orientations at each grid point was then self-rotated along, with an interval of $\Psi = 45°$, creating a total of 200 poses. To account for the internal degrees of freedom, multiple ligand conformations were generated using the Open Babel toolkit [PMID: 21982300]. The procedure was repeated for each generated conformer, resulting in a total of 1000 poses.

During our search process, bad contacts between protein/drugs, such as clashes or piercings, are inevitable. Poses associated with clashes and ring piercings were detected [10] and removed, resulting in a total of ~ 600 distinct poses.

Each pose was then subjected to 1,000 steps of energy minimization using the generalized Born implicit solvent (GBIS) module implemented in NAMD2 [12, 13]. GBIS calculates molecular electrostatics in solvent by representing water as a dielectric continuum as described by the Poisson-Boltzmann equation. Therefore, GBIS accounts for the effect of solvation on nicotine-protein interactions without including explicit solvent, which greatly accelerates modeling efficiency. During minimization, the protein side chains and the drug were allowed to move, while the protein backbone was kept fixed in order to allow the drug to relax in its environment without causing a significant conformational change in the protein. We chose to maintain the global

conformation of the protein, because the X-ray structure of iNicSnFR3a (PDB 7S7U), is already in the nicotine-bound form.

The energy-minimized poses were then used to start molecular dynamics (MD) simulations of the nicotine-iNicSnFR3a complex with explicit water and ion neutralized system. To start simulations from structurally different states, we first clustered the poses based on root-mean square deviation (RMSD) of the drug with a cutoff of 1.5 Å and chose 10 clusters. From each cluster, a representative pose was selected (close to the centroid) and solvated with explicit water and neutralized with 150 mM KCl, resulting in a total of ~ 120,000 atoms with the dimension of 10.9 X 10.9 X 10.9 nm. Each system was energy-minimized using the steepest-descent algorithm for 5,000 steps and then relaxed with the protein restrained, independently. The position restraints were gradually released during the first 4 ns to allow adjustment of the nicotine within the binding pocket (Table 1). Then, unrestrained simulations were performed for all 10 simulations, each for 200 ns, gathering a total of 2 microseconds of simulations.

The effect of the nicotine binding on the protein stability was monitored in each simulation starting from the 10 poses. In seven simulations, the PBP undergoes a conformational change from the holo- to apo- like state (i.e., opening of the binding cleft), suggesting that the particular nicotine poses in these simulations are not favorable. The three simulations that showed most stability of the PBP (Figure S1) were analyzed further to model the final nicotine-bound structure. We re-clustered these simulation data and found only three major clusters with populations greater than 10%. To determine the pose that can fit best to the X-ray map, we first converted the X-ray structure factor file to a real-space map file using the *phenix.mtz2map* command in *phenix* [14] and then calculated the cross-correlation of a representative pose from each cluster (structure closed to the centroid) with that of the generated map. The pose from the highest populated cluster and with the greatest cross-correlation is chosen as the best representation of the nicotine-bound structure.

Step	Backbone restraint force-constant ($kJ\ mol^{-1}\ nm^{-2}$)	Side-chain restraint force-constant ($kJ\ mol^{-1}\ nm^{-}$)	Nicotine heavy atom restraints ($kJ\ mol^{-1}\ nm^{-}$)	Time step (fs)	Simulation Time (ns)
1	400	40	400	1	0.5
2	200	20	200	1	0.5
3	20	20	20	1	1
4	20	20	20	2	2

Table 1

Protocol for gradually releasing the restraints during MD simulations of the 10 nicotine-iDrugSnFR3a complexes, during the search for nicotine docking poses.

MD simulations

MD simulations were performed using GROMACS2022 [15] utilizing AMBER ff19SB [16] and GAFF2 [17] force field parameters for proteins and ligand, respectively. The force field parameters for the chromophore were generated using the antechamber tool of AmberTools22. Bonded and short-range nonbonded interactions were calculated every 2 fs, and periodic boundary conditions were employed in all three dimensions. The particle mesh Ewald (PME) method [18] was used to calculate long-range electrostatic interactions with a grid density of 0.1 nm^{-3}. A force-based smoothing function was employed for pairwise nonbonded interactions at a distance of 1 nm with a cutoff of 1.2 nm. Pairs of atoms whose interactions were evaluated were searched and updated every 20 steps. A cutoff (1.2 nm) slightly longer than the nonbonded cutoff was applied to search for the interacting atom pairs. Constant pressure was maintained at a target of 1 atm using the Parrinello-Rahman algorithm [19]. Temperature coupling was maintained at 300 K with the V-rescale algorithm [20]. The iNicSnFR12 system was generated by computationally mutating E476K, N11E, R467S, Q431D in iNicSnFR3a.

Dynamic network analysis

Dynamic network analysis for iNicSnFR3a and iNicSnFR12 in the nicotine-bound form was performed using the NetworkView plugin in VMD [21-23]. For each system, a network map or graph was created where nodes are the Cα atoms of each protein residue. The edges are added between a pair of nodes if the heavy atoms of the protein residues assigned to these nodes are within 4.5 Å distance for > 75% of the simulation time. The edge weights W_{ij} are derived from pairwise correlations (C_{ij}) which define the probability of information transfer across a given edge:

$$W_{ij} = -log(|C_{ij}|) ,$$

$$C_{ij} = \frac{\Delta r_i(t) \cdot \Delta r_j(t)}{(\langle \Delta r_i(t)^2 \rangle \langle \Delta r_j(t)^2 \rangle)^{1/2}} ,$$

where $\Delta r_i(t) = r_i(t) - \langle \Delta r_i(t) \rangle$ and $r_i(t)$ is the position of the atom corresponding to the ith node. The resulting network was used to determine the optimal path between two given nodes (a source and a sink), using The Floyd–Warshall algorithm [https://doi.org/10.1145/367766.368168]. The optimal path is defined to be the connecting route between the two nodes (residues), which

minimizes the number of intermediate nodes and maximizes the sum of edge betweenness of the connecting route.

Results

Computational prediction

During the development of iNicSnFR3, the ligand binding pocket between the two lobes of the PBP Venus flytrap domain was previously adapted to nicotine by site-saturated mutagenesis (SSM) on "first-shell" residues that lie within 7 Å of the bound ligand as well as on several "second-shell" residues that have interactions with the first-shell residues [1, 6]. To predict mutation hotspots outside the "first shell", we applied computational design.

We first used AutoDock Vina to dock nicotine into iNicSnFR3a [24]. The structure of iNicSnFR3a was extracted from the crystal structure of varenicline-bound iNicSnFR3a (PDB ID: 7S7T) and the varenicline molecule was deleted from the structure to provide an empty binding pocket for nicotine docking. The structure was prepared in AutoDockTools by removing waters, adding polar hydrogens, and assigning Gasteiger charges. To verify the docking procedure, we first docked varenicline into the prepared structure, where the highest scoring pose showed ~ 0.5 Å RMSD from the original varenicline molecule in 7S7T. Then, we docked nicotine into the structure and the conformation with the highest score was selected for computational design.

The nicotine-docked iNicSnFR3a structure was first standardized by the computational protein design suite TRIAD. Then we used the single mutation stability module from TRIAD to rank every possible amino acid substitution at every residue position in the PBP moiety of iNicSnFR3a, one at a time, by predicting the change in energy of folding upon mutation. The top 100 variants were selected (Table S1), from which we selected 34 positions for subsequent sequence design: S9, N20, E24, E27, R36, N39, N46, K51, R52, E64, E78, S325, K336, R341, K342, K362, L384, T413, K414, M418, E429, Q431, D434, D439, D452, D453, K454, R467, E476, N497, K499, D501, E513, E517, excluding all prolines and positions that have been optimized in previous evolution [1].

Finally, we applied the sequence design module from TRIAD to these selected positions individually. All 20 amino acid residues were simulated at each position and the resulting energy changes were evaluated. This is essentially the same prediction strategy as used in the TRIAD single mutation stability module, but the overall energy was calculated more comprehensively and reliably using the Rosetta algorithm. Sequence design on each selected position generated 20 entries of energy scores, from which we identified the positions that could generate the largest energy change when mutated from their wild-type sequences (termed "energy gain", Table S2). The top 10 positions (N20, E27, N46, K51, R52, K342, Q431, K454, R467, and E476), which gave the largest energy changes from wild-type sequence upon mutation, plus N11 which was mutated from iNicSnFR3a to iNicSnFR3b, were selected for SSM screening.

Directed evolution

The evolution path from iNicSnFR3a [1] to iNicSnFR12 is shown in **Figure 1**. In this study, we performed the first round of SSM on positions N20, K342, Q431, and R467 on iNicSnFR3a individually; but the mutations (N20R, K342Y, Q431A, and R467S) with little to no improved sensitivity. Therefore, we tested various combinations of these mutations via SDM. The double mutant (iNicSnFR3a Q431A R467S) showed slightly improved sensitivity.

Starting from iNicSnFR3a Q431A R467S, we performed the second round of SSM on N20 and K342 again, as well as on other positions from the list including N11, E27, N46, K51, R52, K454, E476. We identified N11E as a beneficial mutation, yielding iNicSnFR3a N11E Q431A R467S, which we termed iNicSnFR11.

However, the third round of SSM on iNicSnFR11 yielded no beneficial mutations. We reasoned that iNicSnFR11-nicotine binding affinity might be trapped at a local maximum [25]. We therefore constructed and screened variants that combined several of the previously identified beneficial mutations. We made the first such "combinatorial library" of iNicSnFR11 incorporating some or all of the following mutations: K342, Y342, Q431, A431, G431, D431, E431, R467, G467, S467, H467. Screening this library yielded iNicSnFR11 A431D with improved nicotine sensitivity. Then we made the second combinatorial library of iNicSnFR11 A431D incorporating K51, L51, Q51, R51, R52, V52, T52, E476, K476, M476, V476. We obtained iNicSnFR11 A431D

E476K with further improved nicotine sensitivity, which we termed iNicSnFR12. Therefore. iNicSnFR12 is iNicSnFR3a with mutations: N11E, Q431D, R467S, and E476K.

Characterization of iNicSnFR11 and iNicSnFR12

We used excitation at 485 nm, performed emission measurements at 510 nm, and generated dose-response relations for purified iNicSnFR11 and iNicSnFR12 against four nicotinic agonists: nicotine, ACh, choline, and varenicline. For nicotine at iNicSnFR11, we found a value of 10.6 μM (Figure 2a), or a 2.5-fold improvement over iNicSnFR3a [1]. The iNicSnFR11 data show roughly equal EC_{50} values for varenicline, ACh, and nicotine (Figure 2a), which was also true for iNicSnFR3a [1].

In contrast, the EC_{50} for choline is > 17-fold greater than for nicotine (the fitted value of 170 μM should be considered approximate because the fluorescence did not approach a maximum at 1 mM, the highest value tested). Because we intend to apply iNicSnFR11 and iNicSnFR12 in the sub-μM range, we emphasize measurements in this range. As expected, the dose-response relation of iNicSnFR11 was linear in this range. The proportionality constant (which we term the S-slope) was 1.5 μM^{-1} (Figure 2b), representing a ~ five-fold improvement over the value of 0.3 μM^{-1} estimated for iNicSnFR3a [1].

For iNicSnFR12, the EC_{50} for nicotine was 8.7 μM, the lowest value we have measured in the iNicSnFR series (**Figure 3a**). Varenicline and ACh had EC_{50} were similar (within two-fold) to the value for nicotine (**Figure 3a**); this similarity resembles the pattern for iNicSnFR3a and iNicSnFR11. Again in contrast, EC_{50} for choline was > 12-fold greater than for nicotine (because the dose-response relation for choline did not approach saturation at the highest [choline] tested, we provide no fitted EC_5).

As expected, in the sub-μM range, nicotine displays a linear dose-response relation at iNicSnFR12. The S-slope for nicotine is 2.6 μM^{-1} (Figure 3b). Therefore, iNicSnFR12 responds to nicotine with sensitivity representing a 6.5-fold improvement over iNicSnFR3a [1] and a 1.6-fold improvement over iNicSnF11.

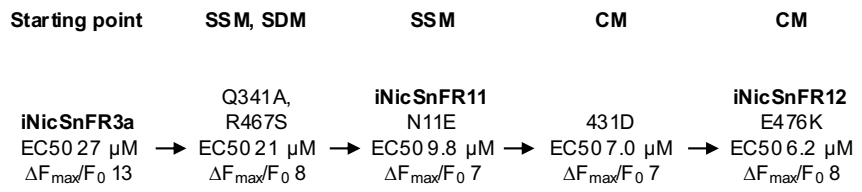


Figure 1. Directed evolution of iNicSnFR11 and iNicSnFR12. The path of protein engineering of iNicSnFR constructs. Data for iNicSnFR3a are the average of two measurements reported previously for the purified protein [1]. Other measurements reported in this figure were made on bacterial lysates, which usually have lower $\Delta F_{max}/F_0$ values (compare with **Figures 2a, 3a** below). SSM, site saturation mutagenesis; SDM, site-directed mutagenesis; CM, combinatorial mutagenesis.

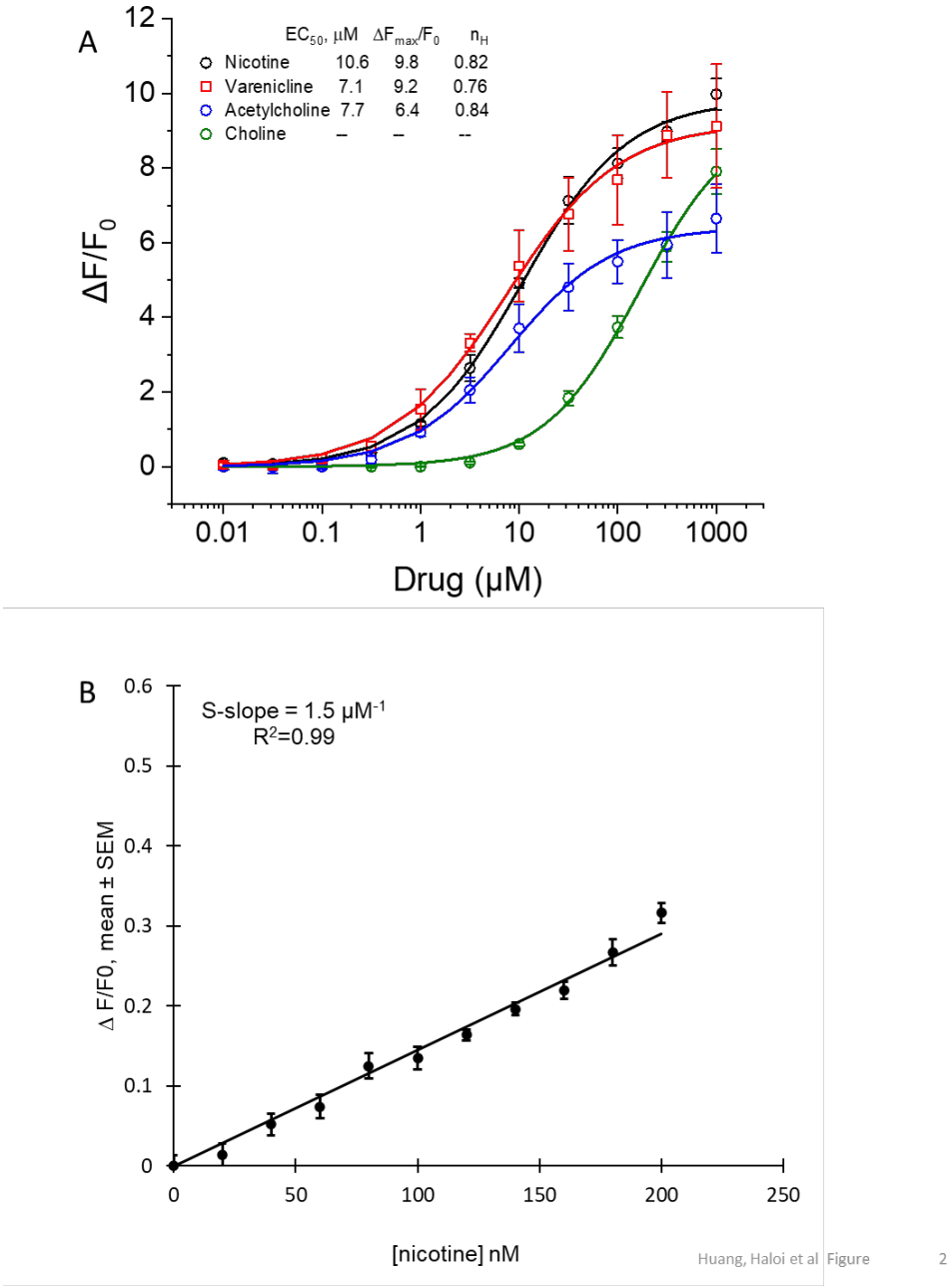
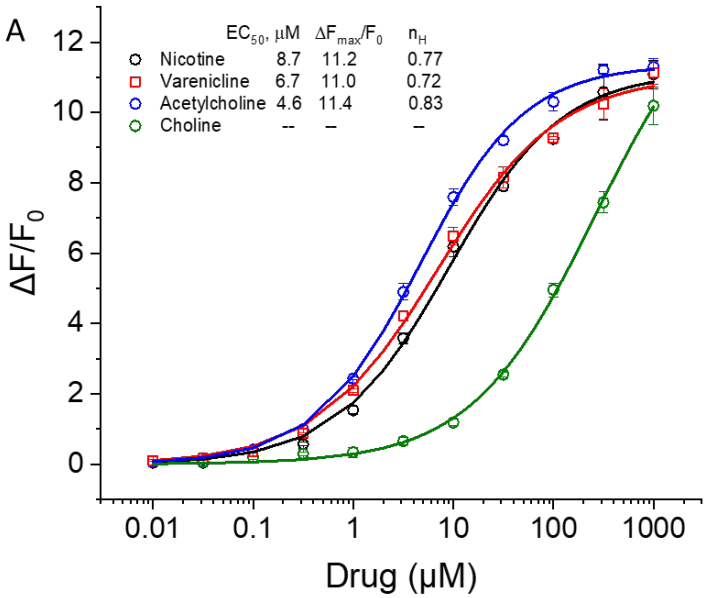
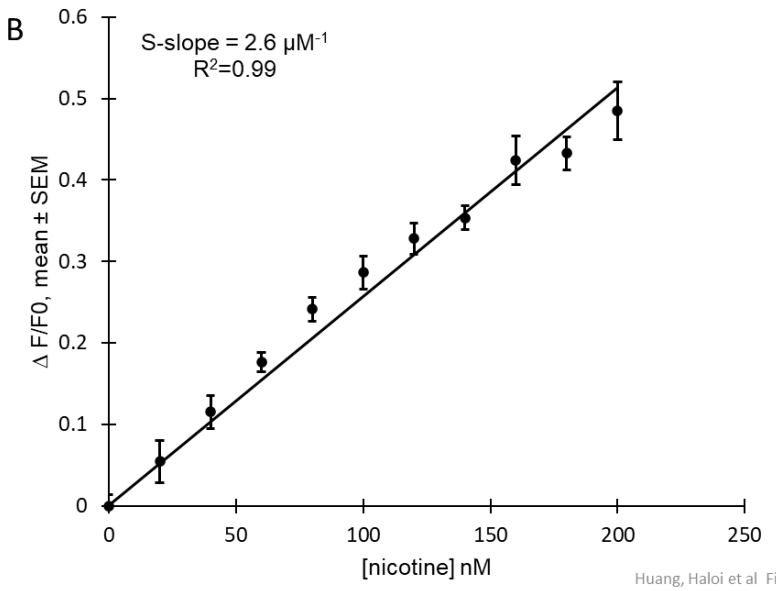


Figure 2. Dose-response relations for purified iNicSnFR11.

A, Dose-response relations for several nicotinic drugs. Purified iNicSnFR11 was measured at 100 nM in 3×PBS, pH 7.40. Data are mean ± SEM (n = 3). B, Dose-response relation for nicotine at sub-µM concentrations. Purified iNicSnFR11 was measured at 100 nM in 3×PBS, pH 7.40. Data are mean ± SEM (n = 16).





Huang, Haloi et al Figure 3

Figure 3. Dose-response relations for purified iNicSnFR12.

A, Dose-response relations for several nicotinic drugs. Purified iNicSnFR12 was measured at 100 nM in 3×PBS, pH 7.40. Data are mean ± SEM (n = 3). B, Dose-response relation for nicotine at sub-μM concentrations. Purified iNicSnFR12 was measured at 100 nM in 3×PBS, pH 7.40. Data are mean ± SEM (n = 8).

Modeling of nicotine within the diffuse electron density

In our previous study [6], the X-ray structure of iDrugSnFR3a saturated with nicotine (PDB ID: 7S7U) was determined with diffuse electron density for the nicotine, which could not be definitively built in the deposited model. To better understand the structural basis for nicotine binding and selectivity of this structure, we sought a more precise ligand-bound model. We developed a computational workflow that involves 1) exhaustive exploration of drug conformations, 2) clustering of the energy minimized conformations, 3) launching of parallel MD simulations from each cluster, 4) re-clustering the MD data, and 5) analysis of cluster populations and cross-correlation calculations with the X-ray map (details in Methods) (Figure 4A). In our analysis, the most highly populated cluster (cluster 1) showed the greatest cross-correlations with the experimental map. Therefore, we chose the pose closest to the centroid of cluster 1 (pose 1) as the representative ligand-bound model, for further simulations and analysis. The existence of multiple sparsely-populated clusters might arise because nicotine is inherently dynamic in the binding pocket, a factor that might contribute to the diffuse electron density.

Next, we analyzed detailed interactions of nicotine within the binding pocket using all the members of cluster 1 (**Figure 4B**). The protonated amine at the methylpyrrolidine of nicotine formed several cation-π interactions with Y65, Y357 and Y460, and hydrogen bond interactions with the sidechain of D397 and the backbone carbonyl of N355. Additional π - π like interactions occurred between the pyridine group and F12, F391 and W436.

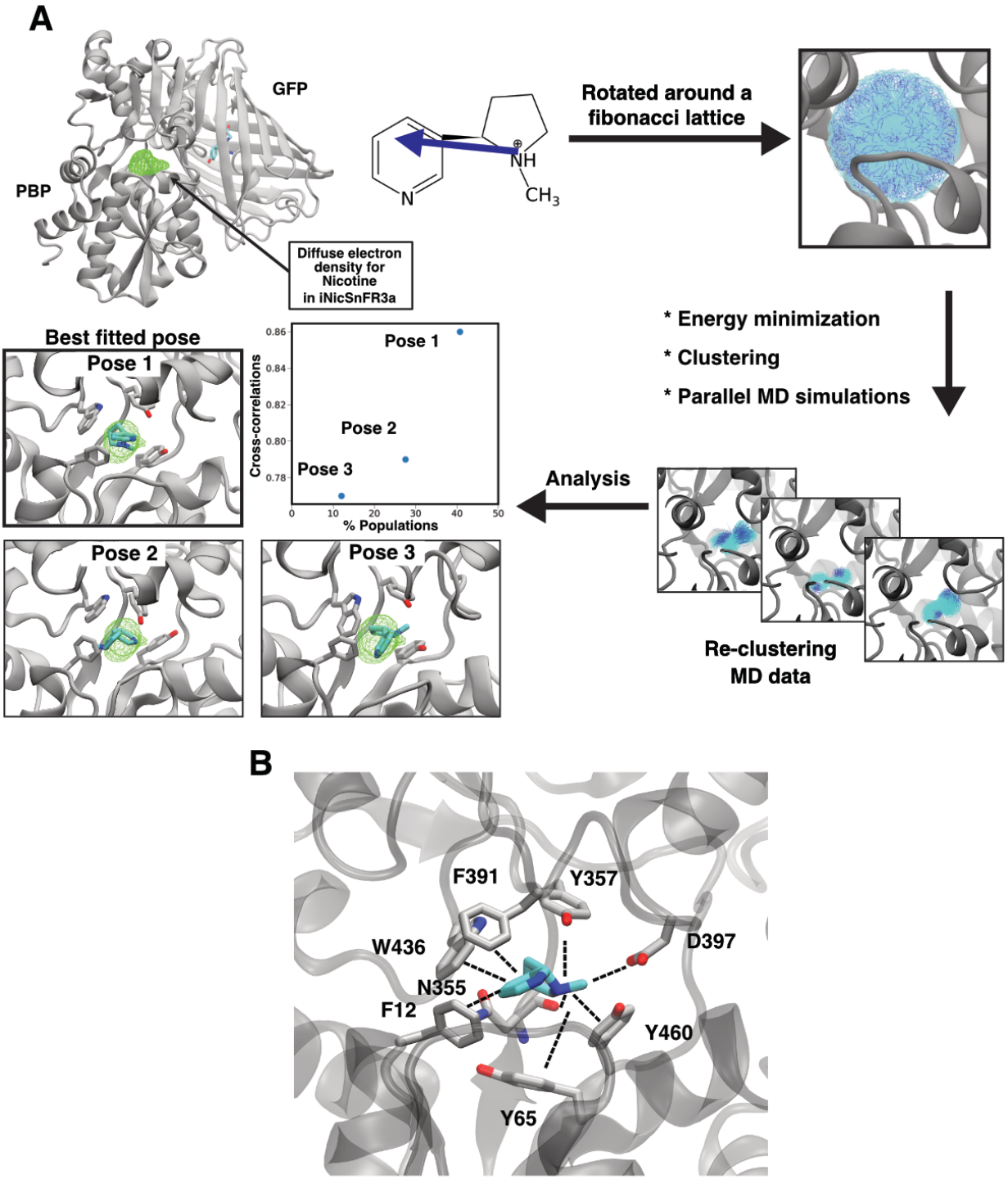


Figure 4: An MD simulation-based workflow to assign a nicotine-bound pose in the X-ray map of nicotine-bound iNicSnFR3a (PDB: 7S7U). (A) The PBP and the GFP (chromophore shown as sticks). The diffuse X-ray density of nicotine is shown in green. To search for possible nicotine poses in the binding pocket, a vector within the ligand was defined and the vector was rotated

along a Fibonacci lattice, including both internal conformations and self-rotation. All the poses were then energy minimized, followed by clustering, and parallel MD simulations. The simulation data set was then re-clustered. The molecular view of the top 3 poses (defined by the member, closest to the centroid of a particular cluster) is shown. Pose 1, a member of the most highly populated cluster, overlaps best with the X-ray map as quantified by the map to model cross-correlation calculations. (B) Molecular view of the detailed ligand-protein interaction in Pose 1. A contact was considered to be present if any two atoms from nicotine and protein lie within 3.5 Å in > 30% of the total members of cluster 1.

Nicotine binding and structural changes

To visualize dynamic nicotine binding in both iNicSnFR12 and iNiCSnFR3a, we performed twelve replicate MD simulations for the two sensors, each for 1 μs, totaling 12 μs for each system, using the ligand-bound structure determined above. In the majority of replicates (8 out of 12 for each system), nicotine remained bound, indicating that the starting structure represented a comparably stable bound state for both sensors. We analyzed these replicates further to understand structural consequences of the differentiating mutations. The N-terminus of the TDPEGAYETVKKEYKR sequence comprises helix 4 in the nomenclature of Fukami et al. [26, 27] and connects the PBP to GFP. Interestingly, in NicSNFr12, helix 4 was shifted ~4 Å away from the centroid of the PBP **(Figure 5**A**)**, compared to NicSNFr3a. One measure of this shift is the probability distribution of the distance between the Cα atom of a terminal residue D329 and the Cα atom Y461, the terminal of a β-sheet that locates nearest to helix 4, residue Y461 (**Figure 5A**). Next, we sought to investigate 1) the basis for this striking change, and 2) whether this shift could be related to increased fluorescence response to nicotine in iNicSnFR12.

Four mutations distinguish iNicSnF12 from the earlier iNicSnFR3a. Of these, R467S is located nearest to helix 4 and exhibited a differential hydrogen bond pattern in the two sensors **(Figure 5B)**. In iNicSnFR3a, R467 reaches out to helix 2 located just "above" it and forms a helix capping interaction with the backbone of the two C-terminal residues I50 and K51 (in addition to a salt-bridge interaction with D469 which locates just "below" R467). These interactions are, however, absent in iNicSnFR12; instead, S467 interacts with the side-chain and backbone of helix 4 C-terminal residues N344 and W343 of helix 4, respectively. We quantified this observation by measuring the distance from the sidechain of residue 467 to N344 **(Figure 5B)**. The probability distribution of this distance metric clearly showed a closer contact between S467-N344 in iNicSnFR12 compared to in iNicSnFR3a. This interaction slightly unwinds and pulls the C-terminus of helix 4 towards the centroid of the PBP **(Figure S2)**; this action could help to force the other end of the helix to tilt away from the PBP as a balancing effect.

We next focused our analysis on the hypothesis that the helix shift modifies sensitivity to nicotine in iNicSnFR12 by remodeling allosteric connectivity to or from the ligand binding site. We presume this connectivity can be partially evaluated in terms of the pathways of residues that mediate force propagation among functional sites in the protein [21]. To compare these pathways,

we performed dynamic network analysis on both the iNicSnFR3a and iNicSnFR12 systems using the Network-View plugin in VMD (see Methods) [21-23]. The source was selected as Y460, a key aromatic residue coordinating the cationic amine of nicotine as described above (**Figure 4B**). The sink was selected as D329, located at the N-terminal tip of the tilted helix 4. Our analysis showed a dramatic change in the allosteric pathways between these two residues. In iNicSnFR3a, the pathway traversed the short helix adjacent to linker 1 (orange). Linker 1 connects the PBP to the GFP and includes the critical residue E78. In the apo state, E78 lies within ~ 2.5 Å of the fluorophore, acting as a "candle snuffer" that drastically decreases fluorescence. The conformational change upon ligand binding moves E78 ~ 14 Å away from the fluorophore, allowing fluorescence [6]. In the bound state, E78 forms salt bridges with Lys97 and Arg99 [6]. In the MD simulations of the iNicSnFR12 bound state, these salt bridges remained present; and the "candle snuffer" mechanism therefore remained intact. However, the mechanical pathway between Y460 and D329 traversed the PBP strand terminating in the mutated position R467S.

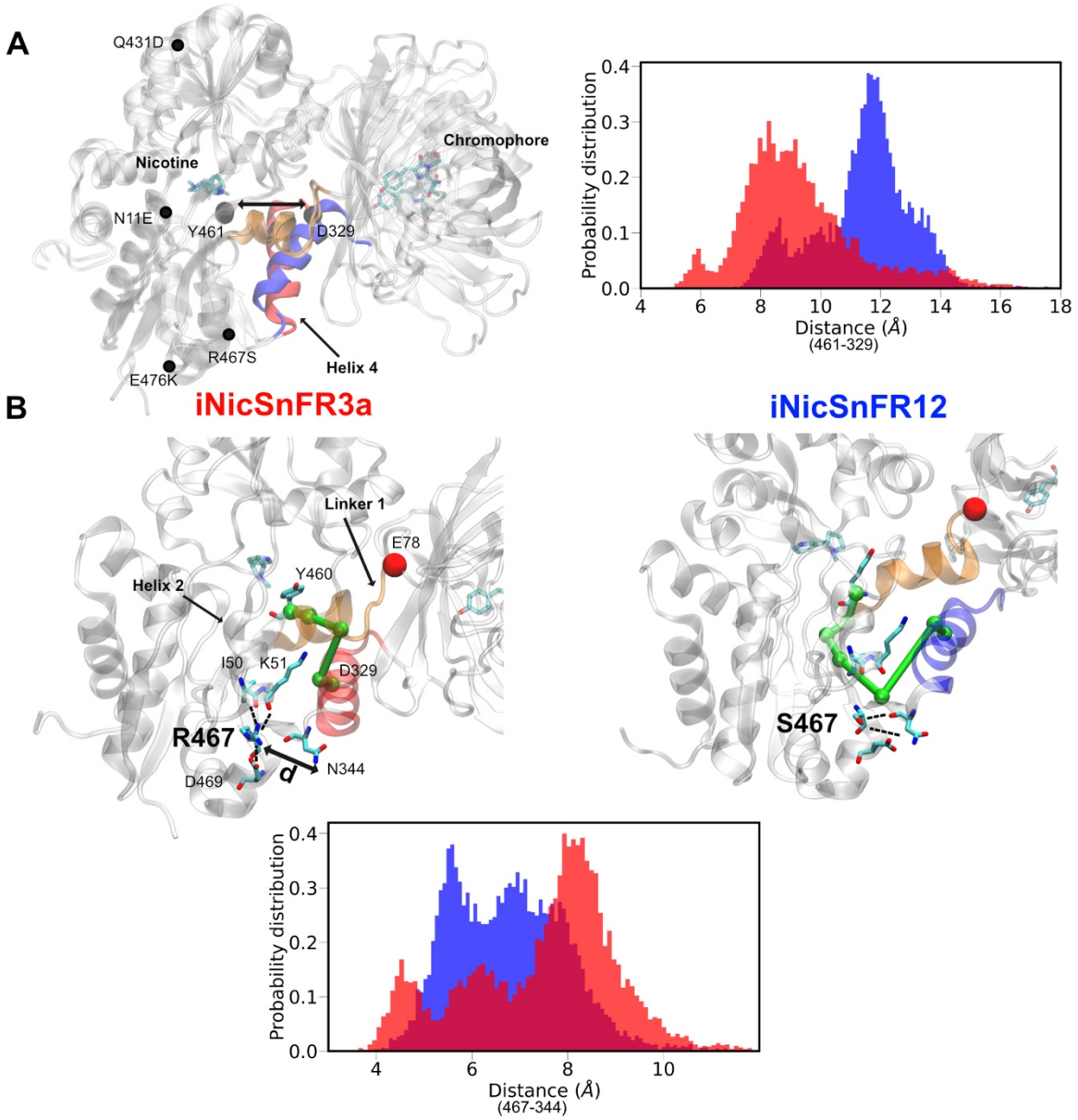


Figure 5. Tilting of linker helix 4 in iNicSnFR12 might modulate ligand binding

(A) All the mutations present in iNicSnFR12 from iNicSnFR3a are labeled. In iNicSnFR12, helix 4 (colored in blue) shifts towards the GFP. To measure this process, we analyzed the probability distribution of the distance between the Cα atom of a terminal residue D329 and the Cα atom Y461, the terminal of a β-sheet that locates nearest to helix 4, residue Y461. (B) The allosteric network from the nicotine binding site residue, Y460, to the N-terminal residue of helix 4, D329, is depicted in green. The probability distribution shows hydrogen bond interactions of residue 467 and N344, measured by calculating the distance between the side chains of these two residues in the two systems.

iNicSnFR12 detects nicotine at 100 nM concentrations in diluted mouse serum and human plasma

Figure 3 shows that solutions containing iNicSnFR12 detect nicotine at < 100 nM. We tested whether this capability continues in solutions containing 1:4 dilutions of mouse and human biofluids. In experiments with mouse serum, Table 2 shows that with only 3 wells per each [nicotine], iNicSnFR distinguished 67 nM nicotine in four-fold diluted mouse serum. However, $\Delta F/F_0$ was somewhat less than in PBS alone. We tested whether this decrease arises from an increase in F_0 due to 10 μM choline, which is thought to be present in human plasma and serum [28-30]. A correction for this level, guided by the choline responses in Figure 4, increased $\Delta F/F_0$, and the corrected $\Delta F/F_0$ was two-fold lower than in PBS (Fig 2A). This two-fold lower response resembled that found for methadone and iMethadoneSnFR in mouse serum [31], presumably for the same reason: the increased salinity. These preliminary data strongly suggest that iNicSnFR12 responds as expected in biofluids; the expected choline response elevates the baseline but adds little or no noise to the data.

In experiments on 1:4 dilutions of commercially obtained human plasma, we found surprisingly high fluorescence in the absence of nicotine. We have not systemically studied the origin of this fluorescence, which seems too high to arise from the presence of varenicline, nicotine, or choline (Figure 6). Despite this background signal, we found it possible to measure added nicotine at concentrations as low as 100 nM; and the fluorescence increase was linear with added [nicotine].

[Nicotine] nM	$\Delta F/F_0$	SEM	Choline-corrected $\Delta F/F_0$
67	0.079	0.021	0.117
200	0.169	0.009	0.252

Table 2. Fluorescence measurements with iNicSnFR12 in mouse serum. Each 100 μL well contains 25% plasma, 75% 4X PBS for adequate buffering, 100 nM iNicSnFR12, and varying concentrations of nicotine tartrate from 20 to 200 nM. N = 3 wells.

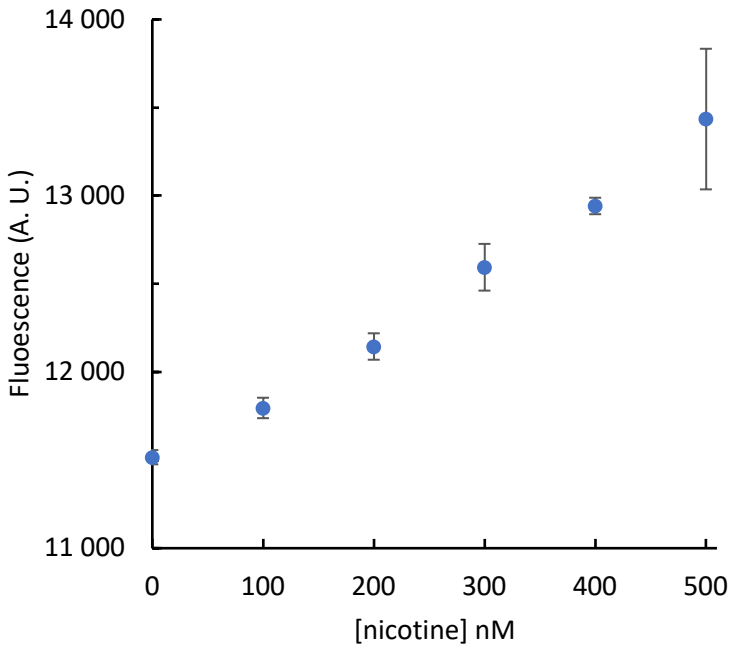


Figure 6 Dose-response data for nicotine at iNicSnFR12 in 25% human serum.

Data were measured 30 min after addition of nicotine. Mean ± SEM, N = 3 samples per concentration.

Activation of iNicSnFR12 in associated with increased absorption at 470-500 nm

Figure 7 presents optical absorbance data for iNicSnF12 in the presence of various nominal [nicotine] between zero (blue trace) and the nearly saturating value of 280 μM. Typical fluorescence experiments are conducted at excitation wavelengths between 470 and 500 nm. At the absorption peak (496 nM), the absorbance of nearly saturated iNicSnFR12 is at least 11-fold larger than in the apo state.

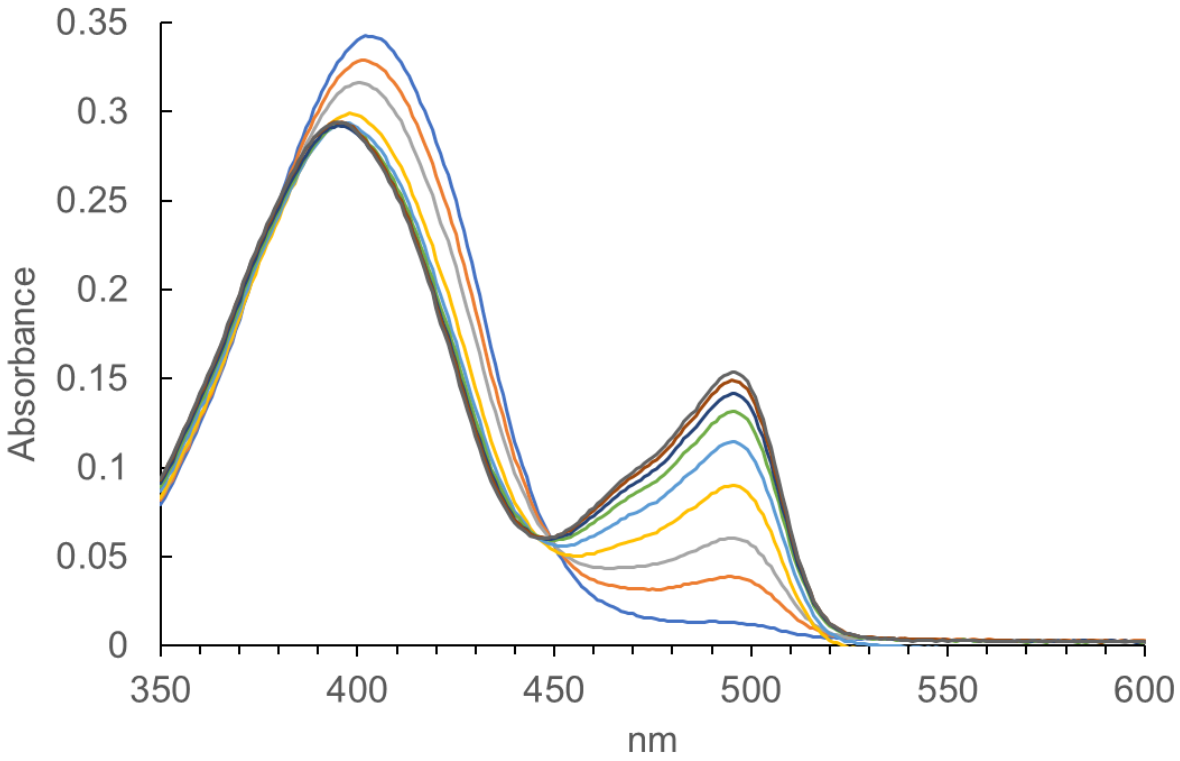


Figure 7. iNicSnFR12 shows increased absorption upon nicotine binding.

Concentrated nicotine solution was added to the iNicSnFR solution to produce the following nominal [nicotine] (μM): blue, zero; red, 20; gray, 40 orange, 80; blue, 120; green, 160; black, 200; purple, 240; dark gray, 280.

Discussion

Characteristics of iNicSnFR12

We show an iterative approach to designing and testing improved fluorescent sensors for nicotine. The approach includes computational protein design and modest-throughput mutagenesis screening using iNicSnFR fluorescence to identify successful variants. We also show a systematic, internally consistent description of nicotine interaction with designed protein sensors, and molecular dynamics simulations to visualize structural consequences of sensitizing mutations on allosteric connectivity involving the ligand binding site.

Computational approaches in this work enabled the identification not only of potentially sensitizing mutations, but also of an experimental nicotine binding pose, and of the structural impact of mutations lacking structural data. These results, however, should be interpreted with appropriate caution, given that large-scale structural impact of engineered mutations may be insufficiently sampled on the timescale of an atomistic simulation, and mechanisms of activation cannot be conclusively determined in absence of experimental structures for both variants in multiple functional states. Classical molecular dynamics force fields may be particularly limited in modeling cation-π interactions, as they do not explicitly consider quantum mechanical terms [32, 33]. Still, substantial remodeling of allosteric pathways involving ligand binding is consistent with a dramatically altered conformational landscape for the nicotine-bound state of iNicSnFR12 versus iNicSnFR3a (Figure 5).

Our measurements (Figure 7) represent the first attempt to compare the absorption characteristics of an iDrugSnFR between the apo and fully liganded state. For iNicSnfFR12, the absorption ratio shows reasonable agreement with the data of Figure 3a, which implies an F_{max}/F_0 ratio of ~ 11. These data suggest that the fluorescence increase when iNicSnFR12 binds nicotine is mostly due to the increased probability of absorbing a photon, rather than from major changes in the quantum yield for fluorescence. We cannot yet rule out the possibility that activation also involves smaller (< 1.5-fold) changes in the quantum yield for fluorescence. Similar conclusions arise from previous studies on other cpGFP-based sensors, such as the GCaMP series [34]. In the 400-450 nM range, iNicSnFR12 shows only minor changes in absorption, like other cpGFP-based sensors [34]. Further measurements are required with higher-quality instrumentation and without the necessity for correcting volumes due to addition of the ligand.

Contemplated uses for iNicSnF12

Nicotine sensors are evolved/designed specifically to bind nicotine, not ACh. However, the constructs do retain sensitivity to varenicline and acetylcholine. Neither of these molecules are expected to interfere with most continuous measurements in human biofluids such as sweat and interstitial fluid.

In healthy adults, the concentration of choline in plasma is ~ 10 μM [28-30]. One expects similar levels in interstitial fluid. This is expected to produce ~ 1% activation of iNicSnFR12. This activation is expected to produce an increased baseline but little degradation of the signal due to nicotine. Nonetheless, decreasing the sensitivity of iNicSnFRs to choline remains a goal of further improvements in the iNicSnFR series.

An additional goal is to improve the sensitivity to nicotine so that ~ 10 nM can be detected. This would allow an iNicSnFR to become the sensor protein in minimally invasive or non-invasive devices that continuously monitor nicotine concentration in biofluids during smoking or vaping sessions, as well as during the decay phase that operates for several hours after a session.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. The raw MD simulation trajectories can be found in: 10.5281/zenodo.7529914

Acknowledgements

A.L.N. was supported by California Tobacco-Related Disease Research Program (TRDRP) Grant 27FT-0022.

H.A.L. was supported by California TRDRP Grant 27IP-0057, National Institute on Drug Abuse Grant DA049140, and National Institute of General Medical Sciences (NIGMS) Grant GM-123582.

D.A.D. was supported by California TRDRP Grant T29IR0455.

S. M. was supported by a grant from the Rosen Bioengineering Center at Caltech.

H. A. L. and S. M. were supported by a grant from the Merkin Institute for Translational Research at Caltech.

E. J. F. was supported by a Caltech Summer Undergraduate Research Fellowship from Paraskeva N. Danailov.

MD simulations were performed using the computing facilities of Karolina Supercomputer and were supported by the Knut and Alice Wallenberg Foundation, the Swedish Research Council (2019-02433, 2021-05806), the Swedish e-Science Research Centre (SeRC), and the BioExcel Center of Excellence (EU-823830), with compute resources provided by the Swedish National Infrastructure for Computing (SNIC) and Euro-HPC (EHPC-REG-2021R0074).

We thank Zoe Beatty, Kallol Bera, Nicholas Friesenhahn, Heather Lukas, and Anand Muthusamy for advice and guidance. We thank Purnima Deshpande for lab management.

[1] A.V. Shivange, P.M. Borden, A.K. Muthusamy, A.L. Nichols, K. Bera, H. Bao, I. Bishara, J. Jeon, M.J. Mulcahy, B. Cohen, S.L. O'Riordan, C. Kim, D.A. Dougherty, E.R. Chapman, J.S. Marvin, L.L. Looger, H.A. Lester, Determining the pharmacokinetics of nicotinic drugs in the endoplasmic reticulum using biosensors, J Gen Physiol 151(4) (2019) 738-757.
[2] B.J. Henderson, H.A. Lester, Inside-out neuropharmacology of nicotinic drugs, Neuropharmacol 96(Pt B) (2015) 178-93.
[3] N.L. Benowitz, J. Hukkanen, P. Jacob, 3rd, Nicotine chemistry, metabolism, kinetics and biomarkers, Handb Exp Pharmacol (192) (2009) 29-60.
[4] E.K. Unger, J.P. Keller, M. Altermatt, R. Liang, A. Matsui, C. Dong, O.J. Hon, Z. Yao, J. Sun, S. Banala, M.E. Flanigan, D.A. Jaffe, S. Hartanto, J. Carlen, G.O. Mizuno, P.M. Borden, A.V. Shivange, L.P. Cameron, S. Sinning, S.M. Underhill, D.E. Olson, S.G. Amara, D. Temple Lang, G. Rudnick, J.S. Marvin, L.D. Lavis, H.A. Lester, V.A. Alvarez, A.J. Fisher, J.A. Prescher, T.L. Kash, V. Yarov-Yarovoy, V. Gradinaru, L.L. Looger, L. Tian, Directed Evolution of a Selective and Sensitive Serotonin Sensor via Machine Learning, Cell 183(7) (2020) 1986-2002.e26.
[5] K. Bera, A. Kamajaya, A.V. Shivange, A.K. Muthusamy, A.L. Nichols, P.M. Borden, S. Grant, J. Jeon, E. Lin, I. Bishara, T.M. Chin, B.N. Cohen, C.H. Kim, E.K. Unger, L. Tian, J.S. Marvin, L.L. Looger, H.A. Lester, Biosensors Show the Pharmacokinetics of S-Ketamine in the Endoplasmic Reticulum, Frontiers in cellular neuroscience 13 (2019) 499-499.
[6] A.L. Nichols, Z. Blumenfeld, C. Fan, L. Luebbert, A.E.M. Blom, B.N. Cohen, J.S. Marvin, P.M. Borden, C.H. Kim, A.K. Muthusamy, A.V. Shivange, H.J. Knox, H.R. Campello, J.H. Wang, D.A. Dougherty, L.L. Looger, T. Gallagher, D.C. Rees, H.A. Lester, Fluorescence activation mechanism and imaging of drug permeation with new sensors for smoking-cessation ligands, eLife 11 (2022).
[7] A. Nichols, Z. Blumenfeld, J Neurosci in press (2023).
[8] S. Kille, C.G. Acevedo-Rocha, L.P. Parra, Z.G. Zhang, D.J. Opperman, M.T. Reetz, J.P. Acevedo, Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis, ACS synthetic biology 2(2) (2013) 83-92.
[9] F.W. Studier, Protein production by auto-induction in high-density shaking cultures, Protein Expression and Purification 41(1) (2005) 207-234.
[10] N. Haloi, A.K. Vasan, E.J. Geddes, A. Prasanna, P.-C. Wen, W.W. Metcalf, P.J. Hergenrother, E. Tajkhorshid, Rationalizing the generation of broad spectrum antibiotics with the addition of a positive charge, Chemical Science 12(45) (2021) 15028-15044.
[11] R. Marques, C. Bouville, M. Ribardière, L.P. Santos, K. Bouatouch, Spherical Fibonacci Point Sets for Illumination Integrals, Computer Graphics Forum 32(8) (2013) 134-143.
[12] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, K. Schulten, Scalable molecular dynamics with NAMD, Journal of computational chemistry 26(16) (2005) 1781-802.
[13] J.C. Phillips, D.J. Hardy, J.D.C. Maia, J.E. Stone, J.V. Ribeiro, R.C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M.C.R. Melo, B.K. Radak, R.D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L.V. Kalé, K. Schulten, C. Chipot, E. Tajkhorshid, Scalable molecular dynamics on CPU and GPU architectures with NAMD, J Chem Phys 153(4) (2020) 044130.

[14] P.D. Adams, P.V. Afonine, G. Bunkoczi, V.B. Chen, I.W. Davis, N. Echols, J.J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N.W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T.C. Terwilliger, P.H. Zwart, PHENIX: A comprehensive Python-based system for macromolecular structure solution, Acta Crystallographica D Biological Crystallography 66(Pt 2) (2010) 213-21.
[15] S. Páll, A. Zhmurov, P. Bauer, M. Abraham, M. Lundborg, A. Gray, B. Hess, E. Lindahl, Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS, J Chem Phys 153(13) (2020) 134110.
[16] C. Tian, K. Kasavajhala, K.A.A. Belfon, L. Raguette, H. Huang, A.N. Migues, J. Bickel, Y. Wang, J. Pincay, Q. Wu, C. Simmerling, ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution, J Chem Theory Comput 16(1) (2020) 528-552.
[17] X. He, V.H. Man, W. Yang, T.S. Lee, J. Wang, A fast and high-quality charge model for the next generation general AMBER force field, J Chem Phys 153(11) (2020) 114502.
[18] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, The Journal of Chemical Physics 98(12) (1993) 10089-10092.
[19] M. Parrinello, A. Rahman, Crystal Structure and Pair Potentials: A Molecular-Dynamics Study, Physical Review Letters 45(14) (1980) 1196-1199.
[20] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, J Chem Phys 126(1) (2007) 014101.
[21] A. Sethi, J. Eargle, A.A. Black, Z. Luthey-Schulten, Dynamical networks in tRNA:protein complexes, Proc Natl Acad Sci U S A 106(16) (2009) 6620-5.
[22] J. Eargle, Z. Luthey-Schulten, NetworkView: 3D display and analysis of protein·RNA interaction networks, Bioinformatics 28(22) (2012) 3000-1.
[23] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J Mol Graph 14(1) (1996) 33-8, 27-8.
[24] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, Journal of computational chemistry 31(2) (2010) 455-61.
[25] M.S. Packer, D.R. Liu, Methods for the directed evolution of proteins, Nat Rev Genet 16(7) (2015) 379-94.
[26] K. Fukami-Kobayashi, Y. Tateno, K. Nishikawa, Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history, J Mol Biol 286(1) (1999) 279-90.
[27] R.P. Berntsson, S.H. Smits, L. Schmitt, D.J. Slotboom, B. Poolman, A structural classification of substrate-binding proteins, FEBS Lett 584(12) (2010) 2606-17.
[28] J. Klein, A. Koppen, K. Loffelholz, J. Schmitthenner, Uptake and Metabolism of Choline By Rat-Brain After Acute Choline Administration, Journal of Neurochemistry 58(3) (1992) 870-876.
[29] J. Bligh, The level of free choline in plasma, J Physiol 117(2) (1952) 234-40.
[30] S.H. Zeisel, M.F. Epstein, R.J. Wurtman, Elevated choline concentration in neonatal plasma, Life Sci 26(21) (1980) 1827-31.
[31] A.K. Muthusamy, C.H. Kim, S.C. Virgil, H.J. Knox, J.S. Marvin, A.L. Nichols, B.N. Cohen, D.A. Dougherty, L.L. Looger, H.A. Lester, Three Mutations Convert the Selectivity of a Protein Sensor from Nicotinic Agonists to S-Methadone for Use in Cells, Organelles, and Biofluids, J Am Chem Soc (2022).

[32] H.M. Khan, A.D. MacKerell, Jr., N. Reuter, Cation-π Interactions between Methylated Ammonium Groups and Tryptophan in the CHARMM36 Additive Force Field, J Chem Theory Comput 15(1) (2019) 7-12.
[33] H. Liu, H. Fu, X. Shao, W. Cai, C. Chipot, Accurate Description of Cation−π Interactions in Proteins with a Nonpolarizable Force Field at No Additional Cost, Journal of Chemical Theory and Computation 16(10) (2020) 6397-6407.
[34] L.M. Barnett, T.E. Hughes, M. Drobizhev, Deciphering the molecular mechanism responsible for GCaMP6m's Ca^{2+}-dependent change in fluorescence, PLoS One 12(2) (2017) e0170934.