

Computational Design of Protein-Based Biosensors for Nicotine

Thesis by
Nicholas James Friesenhahn

In Partial Fulfillment of the Requirements for the
degree of
Doctor of Philosophy

The Caltech logo, featuring the word "Caltech" in a bold, orange, sans-serif font, centered within a light orange rectangular background.

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2026
(Defended May 29, 2026)

© 2026

Nicholas James Friesenhahn
ORCID: 0000-0002-8129-0171

ACKNOWLEDGEMENTS

I would like to acknowledge my advisor Prof. Steve Mayo for his unwavering support and guidance throughout my PhD. In addition to advising me on the technical aspects of my projects, he also contributed to my professional development by enabling me to participate in the Biotechnology Leadership Program at Caltech, complete a summer internship at Protomer Technologies (a wholly owned subsidiary of Eli Lilly and Company) in my third year as a graduate student, and present my work on nicotine biosensors at the Biophysical Society annual meeting in 2026. He has also been invaluable in connecting me with Mayo lab alumni to explore career opportunities after my PhD. I would also like to thank all Mayo Lab members with whom I overlapped for providing feedback during my group meeting presentations and informal chats. These include Monica Breckow, Jingzhou Wang, Aiden Aceves, Shan Huang, Lucas Schaus, Sarah Gillespie, Blade Olson, Andres Orta, Adrian Bahn, Luis Caldera, Mikhail Hameedi, Olive Cheng, Amanda Zhou, and Galina Shevchenko.

Another professor to whom I am indebted for his support and willingness to discuss research is Prof. Henry Lester, whose family of nicotine biosensors I was fortunate enough to work on. He provided me with all the equipment, reagents, and lab space I could need to conduct biophysical assays on biosensor variants, and the weekly group meetings he held were conducive to my steady research progress. Other research mentors who provided me with feedback at these meetings include Prof. Dennis Dougherty, Bruce Cohen, Aaron Nichols, Chris Marotta, Anand Muthusamy, Kallol Bera, and Shan Huang.

In terms of experimental support, several people mentored me and worked alongside me during my PhD. From the Mayo lab, I would like to thank Jingzhou Wang and Monica Breckow for teaching me molecular biology and protein purification techniques. I was also fortunate to work with such dedicated undergraduate students in the lab. Sophie Dalfonzo (Caltech BS '25) was instrumental in identifying biosensor variants from the consensus design approach and assisting in biophysical method development with an emphasis on assay reproducibility. Additionally, Alexis Kim, an undergraduate at Colgate University visiting at Caltech, was a great help with some of my final key experiments. From the Lester lab, I

would like to thank Aaron Nichols and Anand Muthusamy for training me on several laboratory techniques. I would also like to thank the Lester Lab manager Purnima Deshpande for her operational support.

I would like to share deep gratitude to my parents and sisters for their love and support. Although I expressed interest in becoming a medical doctor until partway through college, deep down, my mom always suspected that I would become a scientist. In addition to providing me with the best education I could have hoped for, my parents instilled me with a strong work ethic and resolute integrity.

Finally, I would like to express my heartfelt gratitude to my loving wife Mairead. We began dating shortly before I passed my candidacy exam, so she got to share in many of my ups and downs in graduate school. We got married during my fourth year of graduate school and welcomed our daughter Josephine in my fifth. Mairead is a wonderful wife and mother, and I am so grateful to have her by my side as I finish my academic training and look forward to what is next for our life together.

R. Ch. v.

ABSTRACT

It has been observed that among individual users of nicotine-containing products and between nicotine delivery systems (smoking, vaping, patches, etc.), there is a large degree of variability in the pharmacokinetic profile of nicotine. The time course of [nicotine]_t, is crucial to understanding why some nicotine-containing products are addictive, while others aid in smoking cessation via nicotine-replacement therapy. The current methods for measuring [nicotine]_t via blood draws and LC-MS analysis are cost-prohibitive for the number of nicotine users, estimated at 1.2 billion globally. To enable [nicotine]_t measurements at the necessary scale, we have developed a series of protein-based intensity-based nicotine sensing fluorescent reporters (iNicSnFRs). These sensors are comprised of a nicotine-binding moiety engineered from the OpuBC periplasmic binding protein of *Thermoanaerobacter sp. X513* and circularly permuted superfolderGFP. Through a variety of computational protein design and experimental techniques, we discovered mutations that improved the biosensor's sensitivity, as measured by EC_{50} , 61-fold, from iNicSnFR3a with a 27 μM EC_{50} to iNicSnFR12-D72A with a 440 nM EC_{50} . Notably, iNicSnFR12-D72A, with its series-best 11.6 μM^{-1} fluorescence response in the linear regime of the dose-response relation, is able to resolve 10 nM nicotine, the lowest [nicotine] measured between smoking and vaping sessions in plasma via the blood draw method. With unparalleled sensitivity and robust thermostability profile, iNicSnFR12-D72A meets many requirements for a sensor to be used in biofluids as a continuous nicotine monitor, an important tool in reducing nicotine addiction and associated morbidity and mortality from tobacco-related diseases.

PUBLISHED CONTENT AND CONTRIBUTIONS

All reproduced with permission from Oxford University Press.

Chapter 1:

Nichols AL, Marotta CB, Lukas H, Friesenhahn NJ, Wagenaar DA, Mayo SL, Dougherty DA, Benowitz NL, Gao W, Muthusamy AK, Lester HA. Continuous Nicotine Monitors for Personal Nicotine Pharmacokinetics: A Receptor-Aware Research Agenda. *Nicotine & Tobacco Research*. 2025;27(9):1499-512. 10.1093/ntr/ntaf062

N.J.F contributed to the investigation of the feasibility of continuous nicotine monitors.

Chapter 2:

Haloi N, Huang S, Nichols AL, Fine EJ, Friesenhahn NJ, Marotta CB, Dougherty DA, Lindahl E, Howard RJ, Mayo SL, Lester HA. Interactive computational and experimental approaches improve the sensitivity of periplasmic binding protein-based nicotine biosensors for measurements in biofluids. *Protein Engineering, Design and Selection*. 2024;37:gzae003. 10.1093/protein/gzae003

N.J.F. designed experiments, generated reagents, performed experiments, produced figures, and contributed to the writing of the manuscript.

Appendix B:

Wang J, Aceves AJ, Friesenhahn NJ, Mayo SL. CDRxAbs: antibody small-molecule conjugates with computationally designed target-binding synergy. *Protein Engineering, Design and Selection*. 2025;38:gzaf004. 10.1093/protein/gzaf004

N.J.F designed experiments, generated reagents, performed experiments, and produced figures for the manuscript.

TABLE OF CONTENTS

Acknowledgements.....	iii
Abstract	v
Published Content and Contributions.....	vi
Table of Contents.....	vii
List of Illustrations and/or Tables.....	viii
Nomenclature	xii
Chapter 1: Introduction—Continuous Nicotine Monitors for Personal Nicotine Pharmacokinetics: A Receptor-Aware Research Agenda	1
Introduction	2
Three Motivations for the Proposed Continuous Nicotine Monitor	3
The CNM Concept is Directly Relevant to Behavioral Studies.....	9
Relationship to Conventional Pharmacological Concepts	11
Interstitial Fluid is the Most Relevant Compartment	12
Specifications for a CNM	14
Available Molecular Technology for a CNM	17
A Not-for-Profit, Collaborative Research Agenda.....	26
Bibliography.....	29
Chapter 2: Interactive Computational and Experimental Approaches Improve the Sensitivity of Periplasmic Binding Protein-Based Nicotine Biosensors for Measurements in Biofluids.....	39
Introduction	41
Results and Discussion.....	42
Methods	71
Bibliography.....	81
Chapter 3: Computational Design of Periplasmic Binding Protein-Based Biosensor Enables Measurement of Full Range of Pharmacologically Relevant Nicotine Concentrations	85
Introduction	86
Results and Discussion.....	87
Methods	95
Bibliography.....	103
Chapter 4: Conclusion.....	107
Appendix A: Supplemental Information on the Design and Characterization of Nicotine Biosensors	108
Appendix B: CDRxAbs: Antibody Small-Molecule Conjugates with Computationally Designed Target-Binding Synergy	115
Introduction	116
Results and Discussion.....	119
Methods	132
Bibliography.....	165

LIST OF ILLUSTRATIONS AND/OR TABLES

<i>Number</i>	<i>Page</i>
Chapter 1	
Figure 1.1 Collected data on time course of plasma [nicotine] _t , using the intravenous (IV) blood-draw method.....	3
Figure 1.2 Molecular basis of the iNicSnFR approach to a CNM.....	15
Figure 1.3 Design for a miniaturized fiber photometric CNM, based on initial progress with iNicSnFR12 entrapped in a hydrogel.....	19
Figure 1.4 Electrochemical approaches to a CNM.....	21
Figure 1.5 Synchronized averaging will reduce noise in CNM recordings	25
Figure 1.6 Graphical Summary	26
Chapter 2	
Figure 2.1 Graphical Abstract	40
Table 2.S1 Top 100 mutation outputs from single mutant stability analysis on iNicSnFR3a ranked by energy score	43
Table 2.S2: Mutation outputs from sequence design on iNicSnFR3a on 34 previously identified positions ranked by energy gain	46
Figure 2.2 Directed Evolution and Fluorescence Characteristics of iNicSnFR11 and iNicSnFR12.....	48
Figure 2.3 iNicSnFR12 shows increased absorption at 450–520 nm upon nicotine binding	52
Figure 2.4 An MD simulation-based workflow to assign a nicotine-bound pose in the X-ray map of nicotine-bound iNicSnFR3a (PDB: 7S7U)	53
Figure 2.5 Tilting of linker helix 4 in iNicSnFR12 might modulate ligand binding	55
Figure 2.S1 Partial unwinding of Helix 4 in iNicSnFR12	57

Figure 2.6 S467R mutation to iNicSnFR12 restores enhanced stability inherent to iNicSnFR3a	60
Figure 2.S2 Thermal melt data in the absence of nicotine	61
Figure 2.S3 Thermal melt data in the presence of nicotine	63
Figure 2.7 Dose-response data for nicotine at iNicSnFR12 in biofluids	66
Figure 2.S4 Simulations of decreased S-slope caused by partial activation	67
Table 2.S3: Protocol for gradually releasing the restraints during MD simulations of the 10 nicotine-iDrugSnFR3a complexes, during the search for nicotine docking poses	76
Figure 2.S5 Stability of PBP complexes in simulations	77

Chapter 3

Figure 3.1 Single mutants of iNicSnFR12 with enhanced sensitivity for nicotine	88
Table 3.1: Nicotine sensitivity metrics from fitted dose-response data	89
Figure 3.2 Structural rationale for improved nicotine sensitivity conferred by D72A	90
Figure 3.3 Dose-response relations for nicotine, varenicline, and choline	91
Figure 3.4 Thermostability assessment of iNicSnFR variants via DSF	93
Figure 3.5 Assessments of iNicSnFR12 and iNicSnFR12-D72A resilience to thermal stress	94
Table 3.2: Summary of correction factors of fluorescence quantum yield measurements	102

Appendix A

Table A.1: Consensus design mutations for positions in iNicSnFR12 with above-average <i>RE</i>	108
Table A.2: Single mutants of iNicSnFR12 from α -helix stabilization calculations	109

Figure A.1 Dose-response relations of iNicSnFR12 and the most sensitive single mutants from lysate screen	110
Figure A.2 Dose-response relations of multi-mutants of iNicSnFR12 compared to iNicSnFR12-D72A	111
Figure A.3 Comparison of dose-response relations of sensors in different equivalents of PBS buffer.....	112
Figure A.4 Dose-response time-course experiments to determine appropriate incubation time and BSA equivalents	113
Figure A.5 Reduction in intra-replicate variability in δ -slope experiment by inclusion of BSA in samples and optimization of pipetting by liquid-handling robot	114

Appendix B

Figure B.1 Graphical Abstract.....	116
Figure B.2 Computationally designed nanobody-biotin conjugates.....	119
Figure B.3 Design models and SPR binding for CDR sequence designs v119 and v149	124
Figure B.4 Framework design for reduced aggregation	125
Figure B.5 MD simulation analysis of various models.....	128
Figure B.6 Summary of results and design workflow	130
Table B.S1 4NBX.B-biotin103 amino acid sets for site-saturation mutation scans	135
Figure B.S1 Searching for optimal CDR conformation against monomeric streptavidin model.....	145
Figure B.S2 Identification of optimal conjugation strategy and finalized conjugate models	146
Figure B.S3 Additional Supporting SEC Traces	148
Figure B.S4 Summary of MD simulations performed for 4NBX.B-biotin103 v186 and v186_Fr against mSA _{WT}	149
Figure B.S5 SPR measurements from the intermediate design variants	150

Figure B.S6 Intact-protein mass spectrometry (MS) confirmed mono-conjugated materials.....	151
Figure B.S7 Designing and testing of a nanobody scaffold obtained by directly docking against mSA	152
Figure B.S8 CDR3 loop modeling for 12 docked poses	153
Table B.S2 Top 20 sequence outputs from CDR design of site 31, 32, 104, and 105 on 4NBX.B-biotin103 WT.....	153
Table B.S3 Top 20 sequence outputs from framework design on 4NBX.B-biotin103 v186.....	155
Table B.S4: Top 20 sequence outputs from framework design on 4NBX.B-biotin103 v149.....	156
Table B.S5: Summary of energy scores for proteins experimentally evaluated in this study.....	157
Table B.S6 Amino acid sequences for the protein templates used in this study	157

NOMENCLATURE

ACh. Acetylcholine

ADC. Antibody-drug conjugate.

Apo. In the absence of ligand.

AUC. Area under the curve of drug concentration plotted against time.

BSA. Bovine serum albumin.

CDR. Complementarity-determining region.

CDR_xAb. Complementarity-determining region-extended antibody.

CGM. Continuous glucose monitor.

CM. Combinatorial mutagenesis.

CNM. Continuous nicotine monitor.

cpGFP. Circularly permuted green fluorescent protein.

cp_{sf}GFP. Circularly permuted superfolder green fluorescent protein.

CSF. Cerebrospinal fluid.

DSF. Differential scanning fluorimetry; also known as the thermal shift assay or the ThermoFluor assay.

EC₅₀. Half maximal effective concentration.

EGFP. Enhanced green fluorescent protein.

ENDS. Electronic nicotine delivery system.

ER. Endoplasmic reticulum.

FRET. Förster resonance energy transfer.

GCaMP. Genetically encoded fluorescent calcium indicator.

GFP. Green fluorescent protein.

HPLC. High-performance liquid chromatography.

iDrugSnFR. Intensity-based drug sensing fluorescent reporter.

iNicSnFR. Intensity-based nicotine sensing fluorescent reporter.

ISF. Interstitial fluid.

K_D . Equilibrium dissociation constant.

LC-MS. Liquid chromatography-mass spectrometry.

MD. Molecular dynamics.

MI. Mutual information.

MS. Mass spectrometry.

MSA. Multiple sequence alignment.

mSA. Monomeric streptavidin.

nAChR. Nicotinic acetylcholine receptor.

NRT. Nicotine replacement therapy.

ONP. Oral nicotine product.

PBP. Periplasmic binding protein.

PBS. Phosphate-buffered saline.

PCR. Polymerase chain reaction.

PET. Positron emission tomography.

RE. Relative entropy—a measure of amino acid conservation within an MSA.

RFU. Relative fluorescence units.

RRP. Reduced risk product.

RMSD. Root mean square deviation.

S-slope. Metric for evaluating the sensitivity of a sensor for a ligand calculated as $\frac{\Delta F_{\max}/F_0}{EC_{50}}$.

SDM. Site-directed mutagenesis.

SDS-PAGE. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

SEC. Size-exclusion chromatography.

SEM. Standard error of the mean.

SPR. Surface plasmon resonance.

SSM. Site-saturation mutagenesis.

T_m . Melting temperature.

WT. Wild type.

δ -slope. Slope of the linear portion of the dose-response relation for a given ligand and sensor. Units are inverse concentration.

Φ . Fluorescence quantum yield.

$[x]$. Concentration of x (nicotine, choline, etc.).

$[x]_t$. Concentration of x (nicotine, choline, etc.) as a function of time t .

Chapter 1

INTRODUCTION

Adapted from:

Nichols AL, Marotta CB, Lukas H, Friesenhahn NJ, Wagenaar DA, Mayo SL, Dougherty DA, Benowitz NL, Gao W, Muthusamy AK, Lester HA. Continuous Nicotine Monitors for Personal Nicotine Pharmacokinetics: A Receptor-Aware Research Agenda. *Nicotine & Tobacco Research*. 2025;27(9):1499-512. 10.1093/ntr/ntaf062

Reproduced with permission from Oxford University Press.

Abstract

A minimally invasive “continuous nicotine monitor” (CNM) would resolve the dynamic nicotine concentration, $[\text{nicotine}]_t$, faced by high-sensitivity nicotinic acetylcholine receptors (nAChRs) during and after nicotine intake by individual subjects. Motivations: “Know the potential enemy at an individual level.” Smoking or vaping produces an initial “bolus” of nicotine in the blood and brain, lasting ~5 minutes with a peak concentration of ~100–200 nM. The bolus largely governs reinforcement, reward, and cognitive enhancement. A prolonged declining phase of $[\text{nicotine}]_t$, with a half-time of 1–4 hours, largely suppresses withdrawal symptoms and governs the known cell biology of addiction. Next, “Know the potential therapy,” because individual $[\text{nicotine}]_t$ records will be useful during research on the effectiveness of nicotine replacement therapy. Finally, “Know the physiology.” The only three known effects on nAChRs in cerebrospinal fluid at the relevant $[\text{nicotine}]_t$ are activation, desensitization, and chaperoning/upregulation. Therefore, additional mechanistic insights will arise from correlating $[\text{nicotine}]_t$ with readily measurable physiological data on those effects in molecular, cellular, and brain slice systems and animal models. Interstitial fluid is the appropriate compartment for a CNM. The molecular sensor technology could employ fluorescence, as shown by

progress on measuring $[\text{nicotine}]_t$ with improved variants of intensity-based nicotine-sensing fluorescent reporters (iNicSnFRs). Electrochemical measurements of $[\text{nicotine}]_t$ may also be possible. Studies like the Population Assessment of Tobacco Health would contextualize $[\text{nicotine}]_t$ measurements during each subject's ad libitum nicotine intake, hopefully at a cost $< \$100$ for a 24-hour record.

Introduction

Ever since Columbus's crew sampled tobacco in 1492, nicotine research has led the way in new concepts and techniques for neuroscience. With that history in mind, we intend this article to inform two communities. The nicotine and tobacco research community may wish to learn about the prospects for new generations of wearable, minimally invasive or noninvasive devices to measure the pharmacokinetics of nicotine at an individual level.

The wearable device community may wish to learn about the specific opportunities and challenges associated with nicotine. The opportunities are great: at any given time, some 60% of the world's >1 billion smokers would like to quit. The challenge: nicotine is present at concentrations ~ 1 million times smaller than glucose, the target for the most successful continuous, minimally invasive monitor, yet the required time resolution resembles that for glucose!

We wish to measure nicotine concentration $[\text{nicotine}]$ versus time, t , near the relevant receptors, or $[\text{nicotine}]_t$. Since tobacco produces primarily the S-enantiomer and only the charged species binds to nicotinic acetylcholine receptors (nAChRs), one would more precisely write $[(S)\text{-nicotine}^+]_t$.

Three Motivations for the Proposed Continuous Nicotine Monitor

1. “Know the Potential Enemy”: Personal Pharmacokinetics

It is already known that cigarettes and electronic nicotine delivery systems (ENDS) produce two phases of $[\text{nicotine}]_t$. The first phase occurs while nicotine is entering the blood. It is the relatively high, initial peak or “bolus” of $[\text{nicotine}]$ (100–200 nM for 5–10 minutes, **Figure 1.1**, area highlighted mins 0–10). This phase appears to provide reinforcement via a sense of well-being, stress relief, or cognitive boost. The bolus results from the efficiency of smoking/vaping: (1) the large area (50–75 m²) of the lungs and (2) the high membrane permeability of the nicotine free base. During a puff, nicotine crosses the apical and basolateral membranes of alveolar endothelial cells, reaching arterial blood within a few seconds, then the brain a few seconds later [1-4]. When nicotine is no longer present in the alveoli, $[\text{nicotine}]_t$ begins to decline.

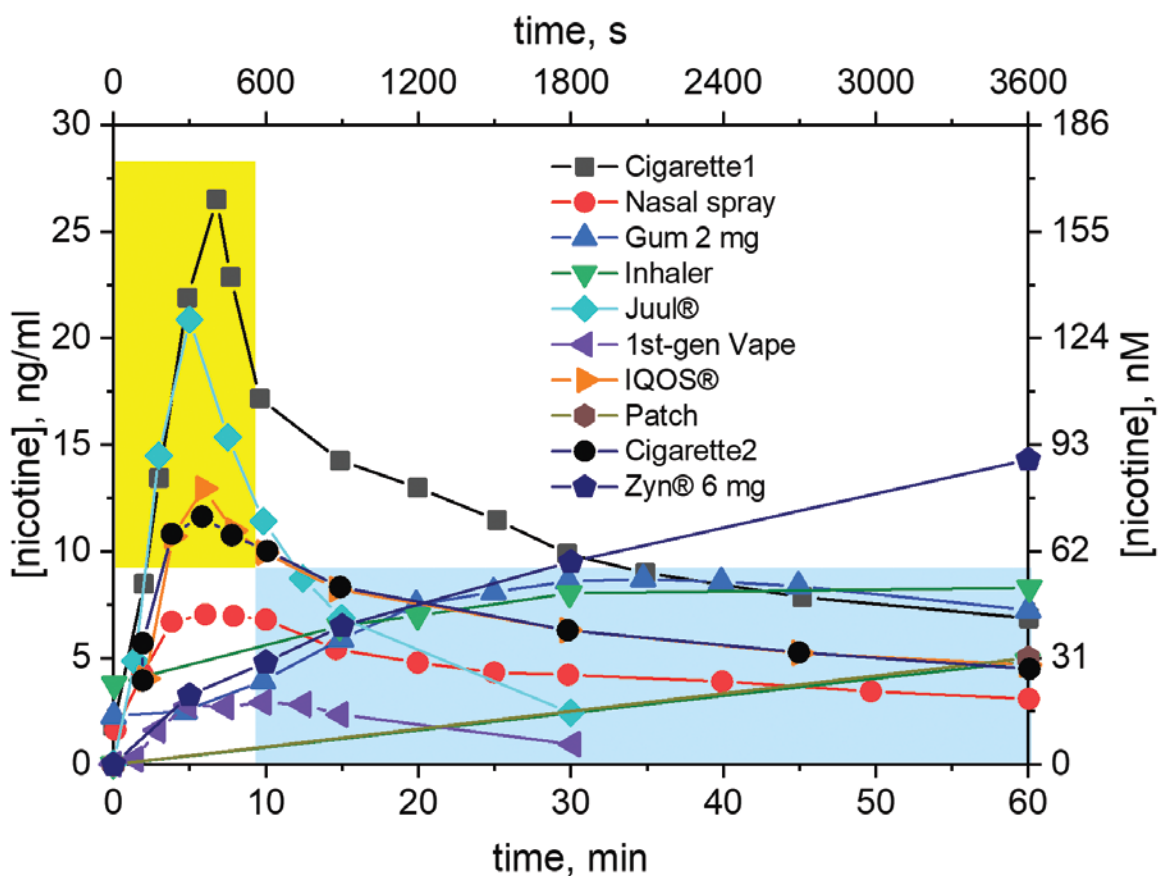


Figure 1.1 Collected data on time course of plasma [nicotine]_t, using the intravenous (IV) blood-draw method. Measurements on cigarette smoking cover a wide range, mostly falling between the two examples given (Cigarette1 [5] and Cigarette2 [6]). Juul, Cigarette1, first-generation (pre-2015) vape [5]; IQOS, Cigarette2 [6]; Nasal spray, Gum, Inhaler, Patch [7]; Zyn [8]. Time zero is the first puff, insertion of the pouch, or attachment of the patch. Measures of variability have been omitted for clarity. All studies comprised at least 24 subjects at least 19 y of age. In one study [6], the subjects had an average age of 34 y and were 52.5% male. Study [7] is a review that compiled several other studies. Two studies [5, 8] did not report on average age or % male.

Oral nicotine products (ONPs) include nicotine pouches, lozenges, gums, and inhalers and involve permeation through the buccal tissues rather than the alveoli. Transdermal patches eponymously use transfer through the skin.

The prolonged, declining phase of [nicotine]_t (**Figure 1.1**, area highlighted mins 10–60) occurs with a half-time of 1–4 hours after a single cigarette. The prolonged phase both suppresses withdrawal symptoms and maintains the cellular biological aspects of addiction [9]. For a given individual, the prolonged declining phase probably differs less than the bolus among methods of intake, because the flux of nicotine into the blood and cerebrospinal fluid (CSF) has nearly stopped. An exception is the transdermal patch; after removal, there may be a plateau of 1–2 hours before the decline begins.

Within a given study (15–30 subjects), [nicotine]_t measured with a single method of nicotine intake shows great variability among subjects; a typical coefficient of variation (CV) is at least 0.5 [10]. When nicotine is administered intravenously in controlled doses (mg/kg), the CV is usually much smaller [10]. The variability in [nicotine]_t among users is a strong reason to study individual [nicotine]_t records. Experiments will require orders of magnitude more subjects than is possible with existing pharmacokinetic methods to address open problems in addiction and therapy:

- Pioneering studies suggest that menthol prolongs the lifetime of [nicotine] in mice [11] and humans [12], increases conditioned place preference to nicotine in mice [13], and potentiates upregulation of nAChRs in mice and humans (see “Know the Physiology,” below) [14]. Menthol might exert these effects via [1] its inhibition of cytochrome P450 2A6 (CYP2A6) [15], [2] its inhibition of TrpA1 channels in respiratory pathways [16], or [3] its action as a chemical chaperone for nAChRs [13]. Mechanism [1], but neither [2] nor [3], would increase and prolong [nicotine]_t.
- On average, women experience more difficulty in smoking cessation than men [17-19]. Experiments with deuterium-labeled nicotine show that the prolonged declining phase is faster in women [20]. Men seem to value the reinforcing effects of the bolus more than women [18]; Does this arise from sex differences in the bolus? How extensive are the individual variations? Does estrogen accelerate nicotine metabolism solely by inducing CYP2A6 [20]?
- Do the varied effects of nicotine across the lifespan [21] correlate with changes in nicotine pharmacokinetics, for instance slowed pharmacokinetics in elderly people [22, 23]?
- Do smokers titrate the bolus to achieve 100–150 nM nicotine, but no more?
- How does the bolus change with reduced-nicotine cigarettes [24-26] or adjustable e-cigarettes [27]?
- How do individual vapers control [nicotine]_t during the day? Do some “seek the peak” and others “avoid the trough?”

2. “Know the Potential Therapy”: Personal Nicotine Replacement Therapy

Nicotine replacement therapy (NRT) will play a role in smoking cessation for the foreseeable future. Various national regulatory agencies (for instance, the US Food and

Drug Administration [FDA]) have *approved* nicotine for smoking cessation via NRT in the form of transdermal patches, inhalers, and gum.

This *approval* does not yet extend to ENDS, which now receive *authorization* for marketing from the FDA's Center for Tobacco Products (CTP) as "reduced risk products" (RRPs) under the Tobacco Control Act of 2009. Like many previous authors, we acknowledge the potential abuse liability of RRP, including the possibility that RRP could cause increases in dual use (vaping and smoking) [28].

However, the FDA's Center for Drug Evaluation and Research (CDER) has now cleared at least one "investigative new drug" (IND) application of an ENDS for investigation as a prescription-only device leading to smoking cessation. In an early iteration of the rationale for such devices, previous CTP and FDA leaders stated [29], "We are examining possible steps the agency could take to address the pharmacokinetic performance of FDA-approved medicinal nicotine products to help more smokers quit. Factors for consideration may include the speed with which nicotine is delivered... [from] products capable of delivering nicotine without having to set tobacco on fire." In a promising technology, piezoelectrically driven ultrasonic mesh inhalers produce controllable droplet size, require no carrier molecules such as propylene glycol and vegetable glycerin, and do not heat the liquid at all [30, 31]. This could represent a step forward in individually controllable NRT [32-34]. The mesh nebulizer that received the IND clearance showed [nicotine]_t kinetics approaching that of a cigarette and much faster than a conventional inhaler [35].

Several metrics have been proposed as guidelines for effective NRT. One suggestion is to measure flux from the device to the subject [36]; this can be measured in the device itself or by collecting nicotine with an impactor. A continuous nicotine monitor (CNM) measuring [nicotine]_t near the relevant nAChRs would be much more useful by capturing the best time resolution. It has been suggested [37] that the most successful means of smoking cessation involves the greatest rate of rise of [nicotine]_t, $\frac{d[\text{nicotine}]_t}{dt}$, presumably

duplicating the “bolus.” An aspirational CNM must resolve the bolus, on an individual basis.

Newer techniques for delivering NRT also raise the specter of abuse liability in two ways. First, it must be noted that many individuals fail to quit smoking with buccal or transdermal NRT because they miss the bolus from a cigarette. If a prescription-only device, such as an ultrasonic nebulizer, perfectly copies the bolus of a user’s favorite cigarette, after a prescribed time course of NRT, a smoker might simply switch to a nicotine salt ENDS, which does optimize the bolus. We acknowledge that vaping manufacturers could use a CNM to enhance the bolus and, therefore, increase abuse liability.

Second, nicotine may well remain in the CSF for the prolonged declining phase, desensitizing nAChRs and provoking the cell biology of nicotine addiction. Newer ONPs also have abuse liability.

Despite these concerns about the abuse liability of NRT, one cannot deduce either the bolus or the prolonged declining phase during therapeutic NRT from first principles; we must perform dynamic, quantitative, individual measurements of $[\text{nicotine}]_t$. Individual measurements of $[\text{nicotine}]_t$ are a highly appropriate component of research to test hypotheses that various forms of NRT benefit distinct subgroups of people defined by “tobacco abuse” [38].

3. “Know the Physiology”: Activation, Desensitization, and Chaperoning/Upregulation of Receptors

The CNM concept interfaces directly with nearly a century of research on time-resolved nAChR responses. Nicotine acts on nAChRs in only three ways, according to present knowledge. Nicotine activates nAChRs, producing ion flux (including Ca^{2+} flux) and depolarization. After a few minutes at the relevant nAChRs, nicotine desensitizes nAChRs. Slowest of all—hours to days—nicotine upregulates nAChRs via one or more

post-translational processes, primarily chaperoning of nascent nAChRs within the endoplasmic reticulum (ER). The first two processes have received intense study since the identification of nAChRs, including experiments on isolated cells, brain slices, in vivo recordings on animal models, imaging of nAChR availability, functional magnetic resonance imaging, and heterologous expression, mostly via electrophysiological and Ca^{2+} recording.

Chaperoning/upregulation are both more complex and less well studied [9, 39-42]. How is the prolonged phase crucial for nicotine dependence? Nicotine enters the ER, beginning a pathway that we term “inside-out” [9]. This entry occurs within a few seconds after nicotine appears near cells, and at a concentration within 2-fold of the extracellular value [43]. Importantly, even [nicotine] as low as 10 nM activates the “inside-out” pathway [41], and in some people, [nicotine] continues to exceed 10 nM for ~6 hours after smoking/vaping [43-47]. A major effect of the inside-out pathway is to produce upregulation of plasma membrane nAChRs, and this upregulation is selective at every level studied—brain region, cell type, axonal versus somatic versus dendritic localization, and subunit stoichiometry of the upregulation nAChRs [9].

Important simplifications are now in hand: It is generally agreed that the relevant nAChRs are $\alpha 4\beta 2^*$, $\alpha 6\beta 2^*$, and possibly $\alpha 2\beta 2^*$ (the * denotes the possible presence of an additional subunit, such as $\alpha 5$, in the assembled pentameric receptor). Also, the relevant peak [nicotine]_t values are those achieved by smoking, <200 nM, in part because higher [nicotine] activates other nAChRs in aversive pathways. In another major simplification, measurements in CSF are unnecessary: nicotine reaches the CSF within seconds from brain capillaries because of its high $\log D_{\text{pH } 7.4}$ [48].

Some aspects of nicotine pharmacokinetics are shaped by the effects of inhaled nicotine in the mouth, airways, and the alveolar epithelium; the nicotine concentrations in these compartments are much higher than considered above, and they take place at additional targets. (1) Deprotonated (free base) nicotine activates TrpA1, a channel abundantly expressed in airways [16, 49]. This irritation could induce “braking” of respiration

[50]. (2) Aerosolized nicotine in the alveoli equilibrates with the cytosol of alveolar epithelial cells, and this could lead to cytosolic concentrations of hundreds of millimolar (mM; Henry's law). In alveolar epithelial cells, these concentrations may be high enough to chaperone and upregulate additional nAChRs, such as $\alpha 3\beta 4^*$ [51, 52]. These complications are additional reasons to measure $[\text{nicotine}]_t$ in CSF-like compartments, downstream from measurements on nicotine flux from devices.

The CNM Concept is Directly Relevant to Social and Behavioral Studies of Nicotine Intake

Because nicotine consumption—certainly for the abusive aspects, and probably also for the smoking cessation aspects—involves many cues, social support, and other activities such as eating and drinking, flavorings, and diurnal rhythms, it is crucial that measurements of $[\text{nicotine}]_t$ occur during ad libitum consumption in daily life. The CNM we envision would satisfy this requirement.

The Population Assessment of Tobacco Health (PATH) study has resulted in 737 research papers as of mid-2024. The PATH study's biomarker and biospecimen components have varied. PATH-like studies would be enhanced by 24-hour CNM records of $[\text{nicotine}]_t$, and we argue below that such records would be more useful than the currently employed urinary nicotine equivalents. Distributing and supporting the CNM could become as straightforward as the activities associated with continuous glucose monitors (CGMs; see below). If a study also collects tissue, blood, or saliva, the DNA analyses could further define additional components of nicotine metabolism, including CYP2B6 [53] and CYP2E1 [54].

Present Measurements of $[\text{Nicotine}]_t$

The classical ideas introduced so far have motivated measurements of $[\text{nicotine}]_t$ on hundreds of subjects gathered in several recent reviews [10, 37, 55]. Two methods now provide “gold standard” measurements of nicotine pharmacokinetics. The intravenous

blood-draw method is performed in a clinic on a subject with an intravenous catheter [5, 6, 56-58]. It costs thousands of dollars per subject, including tens of dollars for each sample in a typical series of 10–25 samples repeated at 5-minute intervals. The positron emission tomography (PET) method, which mixes ^{11}C -nicotine with inhaled smoking or vaping, has a better temporal resolution (a few seconds) [59, 60]. For both methods, a study typically includes 15–30 subjects. Both methods are too tedious and expensive for routine personal use. Neither can be routinely applied to youths.

Slower methods have been especially useful in showing that the prolonged, declining phase varies up to 10-fold among individuals, partially due to polymorphisms in cytochrome P450 2A6 (CYP2A6). CYP2A6 partially governs the conversion of nicotine to (nearly inactive) cotinine [61]. The “nicotine metabolite ratio” (NMR) measures CYP2A6 activity. NMR reveals that ~15% of the population are “slow” metabolizers: individuals with defective CYP2A6 have up to a severalfold longer [nicotine]_t half-life, and the increase depends on the genotype [61]. Slow metabolizers smoke fewer cigarettes per day; but especially for moderate to heaving smokers, they score just as strongly “dependent” on Fagerstrom-like tests [62].

NMR does not measure the key data stream, [nicotine]_t. Furthermore, genetic risk scores capture only 34.5% of the variability in NMR [63, 64]. Proxy measures for [nicotine]_t in blood include urinary nicotine metabolites. The only presently available physiological proxy for [nicotine]_t, increased heart rate [46], is complicated for ad libitum intake because heart rate is influenced by several activities of daily living: exercise, simultaneous consumption of caffeinated beverages and nicotine, and tolerance to subsequent cigarettes [65-67]. Some ENDS measure puff strength, duration, and frequency (“smoking topography”) [68].

Thus, [nicotine]_t is the relevant metric and conventional intravenous (IV) and PET methods have too few subjects to resolve most confounding factors. In contrast, we envision a CNM that costs <100 USD (2024) to scalably study tens of thousands of

subjects, transforming “confounding factors” for either therapy or abuse into “identifiable risk factors” or “personal smoking cessation therapy.”

Relationship to Conventional Pharmacological Concepts

Unbound Concentrations

All the techniques we envision will monitor the unbound concentration of nicotine⁺. This is the form sensed by the binding site of nAChRs, the active site of enzymes such as CYP2A6, and the binding site of biosensor proteins described below. However, the TrpA1 channel probably binds the deprotonated (free base) form [16].

Volume of Distribution

The unbound concentration of any drug should be distinguished from the total amount of the drug in the organism [69]. This concept is usually associated with apparent volume of distribution. Volume of distribution in humans is conventionally inferred from measurements of the so-called clearance [69]. Most drugs have an apparent volume of distribution >1 L/kg, thought to arise either via drug binding to proteins, accumulation within lipids, or accumulation within acidic organelles (“acid trapping”). Nicotine has a relatively low volume of distribution, 1.8–4.2 L/kg [70], with a consensus value of 2.6 L/kg [71]. That is, nicotine becomes bound or otherwise sequestered to the extent that the total amount of nicotine in the body would be appropriate to the volume of water equal to ~2.6 times the body’s weight (“corporal water volume”).

Most of these possibilities and mechanisms can influence [nicotine]_t. Therefore, some of these mechanisms would shape the data stream produced by a CNM—but none would challenge the validity of the [nicotine]_t data as they apply to actions on nAChRs.

Area Under the Curve of Concentration $\times$ Time

Area under the curve (AUC) is often invoked to compare actions of therapeutic drugs. For nicotine, abuse potential is sometimes assumed to increase with AUC [55, 72, 73], but this idea has little experimental proof. More generally, the goal to “know the physiology” will transcend the AUC metric by showing how the bolus and the prolonged phase affect activation, desensitization, or chaperoning of nAChRs.

Interstitial Fluid is the Most Relevant Compartment

CSF bathes highly nicotine-sensitive nAChRs in several distinct brain regions; therefore, $[\text{nicotine}]_t$ in CSF governs activation, desensitization, and upregulation of nAChRs. However, no minimally invasive or noninvasive, nonradioactive technique is available for measuring $[\text{nicotine}]_t$ in CSF. For all means of nicotine intake, the blood–brain barrier, formed by tight junctions in cerebral capillaries and by the end-feet of astrocytes, provides a delay of at most a few seconds in achieving a CSF $[\text{nicotine}]_t$ that essentially equals blood $[\text{nicotine}]_t$ [10, 48].

As in most organs, brain cells lie within 50 μm of capillaries. Nicotine can diffuse passively over 50 μm within 5 seconds (assuming a diffusion constant of 0.5 $\mu\text{m}^2/\text{ms}$). Smokers and vapers feel a “buzz” within ~ 20 seconds, showing that nicotine has appeared in the CSF and reached nAChRs. Rat measurements confirm that $[\text{nicotine}]_t$ in CSF is close to that in plasma [74]. Special intra-arterial ^{11}C nicotine PET measurements in humans achieve a resolution of ~ 5 seconds [2, 4, 59], which is unnecessary for most studies of $[\text{nicotine}]_t$.

Interstitial fluid (ISF) is minimally invasive. Nicotine is predicted to be slightly more permeable to capillaries in peripheral tissue than to brain capillaries; but the difference is likely to be just a few seconds. In this context, time-resolved measurements of $[\text{nicotine}]_t$ in ISF are presumably more relevant than blood to $[\text{nicotine}]_t$ in CSF.

Commercial tobacco manufacturers modify the taste and and/or pH of cigars and cigarettes; and this may also modify $[\text{nicotine}]_i$ [75, 76]. In nicotine salt ENDS, the pH of the e-liquid and of the aerosol is lower than cigarettes; but the higher solubility of nicotine^+ than of free base nicotine allows increased $[\text{nicotine}^+]$ in the e-liquid and in the inhaled aerosol. Manufacturers of nicotine salt ENDS may also choose an anion that affects airway flow [4, 77]. All NRT strategies also manipulate nicotine pharmacokinetics [10, 37].

On the one hand, the various effects on nicotine entry into the blood cannot yet be predicted simply either from the amount of e-liquid left in the reservoir, from the volume of each puff, or from the nicotine deposited on an impactor after a controlled number of mechanical puffs. On the other hand, once the nicotine enters the blood, its concentration ($<1 \mu\text{M}$) is so low that buffers in the blood maintain its pH at ~ 7.4 . If a nicotine salt is ingested, the blood similarly dilutes the anion, which therefore need not be considered further. As noted above, the nicotine passes within seconds from the blood to the CSF and other compartments such as ISF. Therefore, $[\text{nicotine}]_i$ in the blood or ISF is an excellent measurement of the delivery system's total pharmacologically relevant, dynamic result—whether the delivery system is a combustible cigarette, an ENDS, an ONP, or NRT.

In summary, we strongly favor interstitial fluid as the compartment of choice for measurements of personal $[\text{nicotine}]_i$.

Analogy to CGMs

All FDA-approved CGMs reside in ISF, are inserted with minimally invasive techniques, transmit their data, $[\text{glucose}]_i$, wirelessly, and have a temporal resolution of ~ 5 minutes. One FDA-approved glucose monitor uses fluorescent detection [78, 79], although via a different molecular mechanism than the one described below.

Most CGMs now in use employ electrochemical detection via glucose oxidase positioned at the intradermal tip of a 5-mm-long gold wire. The latest electrochemical CGMs last ≥ 10 days. CGMs have become commodity items, available over the counter (\$50) for use by prediabetic patients and by nondiabetic consumers to satisfy their curiosity about their [glucose]. Several labs are now attempting to develop CGMs in essentially noninvasive form: arrays of microneedles a few hundred micrometers long and tens of micrometers wide, with their tips in ISF.

Sweat

Although sweat measurements have the great advantage of being noninvasive, other aspects lead to uncertainties that sweat can provide quantitative, time-resolved measurements relevant to CSF. Sweat is secreted through a complex duct at varying pressures. According to a biofluidic model of sweat glands, a biomolecule secreted by the gland reaches the skin surface within <5 minutes, fulfilling the temporal resolution criterion of a CNM [80]. Active microfluidics may then be required in the sensor device to transport the sweat to the sensing surface or chamber. This would introduce a delay; but if the flow is laminar, the time of actual secretion can be determined [81]. Primary sweat (in the sweat coil) is quite acidic [80], leading to concerns about acid trapping, which may explain initial measurements that [nicotine] in sweat is >10 times higher than in blood [82]. It is not understood whether the walls of the sweat duct would then allow reabsorption of nicotine. Researchers in the sweat sensor field are working to eliminate these complications, with the goal of realistic measurements on a known time scale [83-86].

Specifications for a CNM

Concentration Resolution: ~ 10 nM

Although the EC_{50} for nicotine activation of nAChRs is typically several hundred nanomolar (nM), much lower concentrations desensitize or upregulate nAChRs [9]. For instance, the

EC_{50} for upregulation is 37 nM [41]. We therefore consider it necessary to measure $[\text{nicotine}]_t$ as it declines to values as low as ~ 10 nM.

Absolute [Nicotine] Calibration

The signal from all copies of the CNM should have an invariant, simple relationship to $[\text{nicotine}]_t$. In **Figure 1.2**, we show that a candidate fluorescent sensor molecule displays a linear response in the relevant range of $[\text{nicotine}]$, and we refer to the sensitivity as δ -slope. We would accept an absolute accuracy of $\pm 20\%$. Modern electrochemical CGMs do have absolute calibrations, in part because a few CGMs from each production batch are tested at varying $[\text{glucose}]$; the resulting calibrations are then embedded in all CGMs from the batch [87].

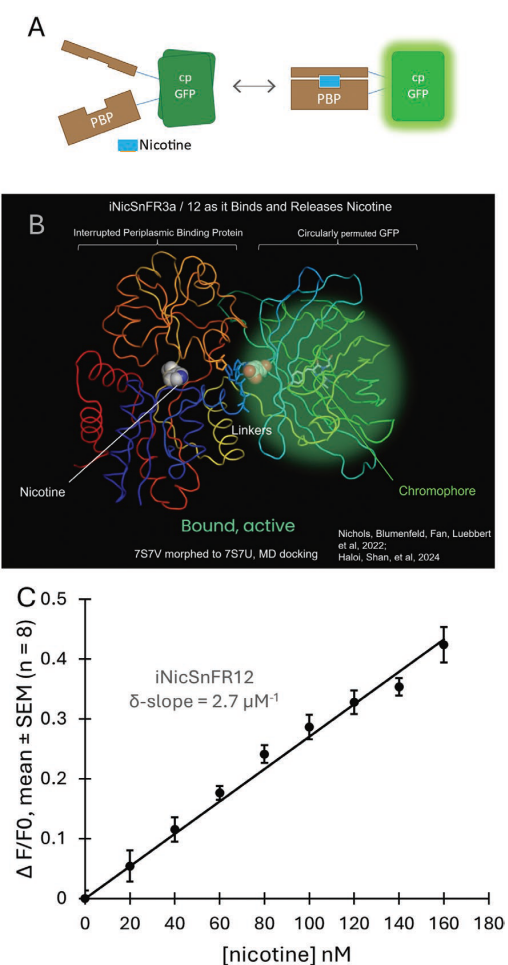


Figure 1.2 Molecular basis of the iNicSnFR approach to a CNM. (A) Schematic view of the single-chain fluorescent sensing molecule such as iNicSnFR12. (B) A single frame from **Supplementary Video 1.S1** (available at *Nicotine and Tobacco Research* online), depicting the details of the conformational change that underlies sensing of nicotine by iNicSnFR12. (C) Concentration–response curve at [nicotine] in the range 0–180 nM, the expected plasma, CSF, or ISF range during nicotine intake via smoking, vaping, transdermal patches, inhalers, or oral nicotine products [10]. Measurements were conducted in 96-well microtiter plates, and each well contained 100 nM purified iNicSnFR12 in 100 μ L of solution. F_0 and ΔF are the resting fluorescence and additional nicotine-induced fluorescence, respectively. The excitation and emissions wavelengths are 496 and 535 nm, respectively. From Haloi et al. [88] iNicSnFR: intensity-based nicotine-sensing fluorescent reporter.

Temporal Resolution: Five Minutes

The CNM must have the temporal resolution to measure the bolus of [nicotine]_t (~100 nM for ~5 minutes, **Figure 1.1**, area highlighted mins 0–10). This implies a temporal resolution of ~300 seconds: the typical 10-puff period for a single cigarette or a typical ENDS “vaping session.” Juul and IQOS achieve 60–150 nM nicotine during this 300- to 600-second time frame (**Figure 1.1**).

Selectivity: At Least 100-Fold for Nicotine

The ideal CNM would have zero response to other molecules in the ISF. We aim to render the sensing molecules 100-fold more sensitive to nicotine than to all other ligands. According to this specification, even if an endogenous interfering molecule has a 100-fold higher concentration higher than nicotine, this would not markedly distort the [nicotine]_t signal.

Durability: At Least 24 Hours

A suitable first-generation CNM would measure [nicotine]_t for roughly 24 hours. This would allow 40 or more cigarettes or vaping sessions, providing an excellent view of the prolonged phase. Including a full 24 hours has the additional advantage that [nicotine]_t is expected to fall to nearly zero during the night, providing an important set of control readings.

Measurements of [nicotine]_t during research on smoking cessation would require much longer records. By analogy with CGMs, a 10-day period seems approachable. Mice with brain expression of cpGFP-based biosensors provide signals for as long as 1 year [89]. Mice expressing an adeno-associated virus encoding iFentanylSnFR, which differs by only a dozen amino acids from intensity-based nicotine-sensing fluorescent reports (iNicSnFRs), continue to give useful responses to fentanyl for at least 3 weeks [90].

Available Molecular Technology for a CNM

Nicotine-Binding Protein-Based Approaches: The iNicSnFR Family

The iNicSnFR family [43, 88, 91] is based on ~50 years of research on periplasmic binding proteins (PBPs), comprising some 20,000 genes in bacterial and archaeal species. The PBP binds a small molecule (ligand) of importance to the microorganism (in an intact bacterium, the PBP would then present the ligand to other proteins). More than 450 atomic-scale structures for PBPs now exist in the protein data bank RCSB. In all known cases, the PBP responds to the ligand binding by undergoing a conserved “Venus flytrap” or “clamshell” conformational change [92]. Most known PBPs exist as monomers, and the conformational change depends on the ligand conformation with a Hill coefficient near 1. This well-understood change requires <1 second in most cases, rendering the PBP suitable to become the molecular basis of a real-time, continuous, reversible, reagent-less sensor (**Figure 1.2**).

The PBP of most interest for a CNM is a mutated variant of OpuBC from a hyperthermophilic bacterial species which uses choline or betaine as an osmolyte. We mutated OpuBC to favor binding of acetylcholine [93] or nicotine [43]. Each member of the iNicSnFR family is a single protein chain. The chain represents a nearly modular merger of two moieties: the PBP binding moiety and a fluorescent readout moiety. We use the term “merged” to avoid implying that these two moieties are simply concatenated, C-terminal of the first connected to N-terminal of the second. The GFP moiety is circularly permuted (cpGFP), as in GCaMP sensors for Ca^{2+} . The circularly permuted cpGFP is then inserted into the PBP sequence, near the hinge region. Recently, the iNicSnFR project was aided by adding computational predictions [94-96] to the directed evolution pipeline. iNicSnFR12 achieves nearly the desired sensitivity in 96-well saline-based lab tests (**Figure 1.2C**) [88]. The concentration–response relation is linear in the range shown, because the EC_{50} for the entire relation is $>1\ \mu\text{M}$ [88].

The iNicSnFR approach exploits the amplification produced by modulating the fluorescence on the basis of nicotine binding to the PBP moiety. During each second that the single nicotine molecule is bound to the PBP, the fluorophore absorbs as many as 10^8 blue photons, then emits ~ 0.7 times as many green photons [88]. The optical elements resemble those of a modern fluorescence microscope, simplified further by measuring with a single photodetector (“fiber photometry”). **Figure 1.3** presents a schematic for our planned next-generation CNM.

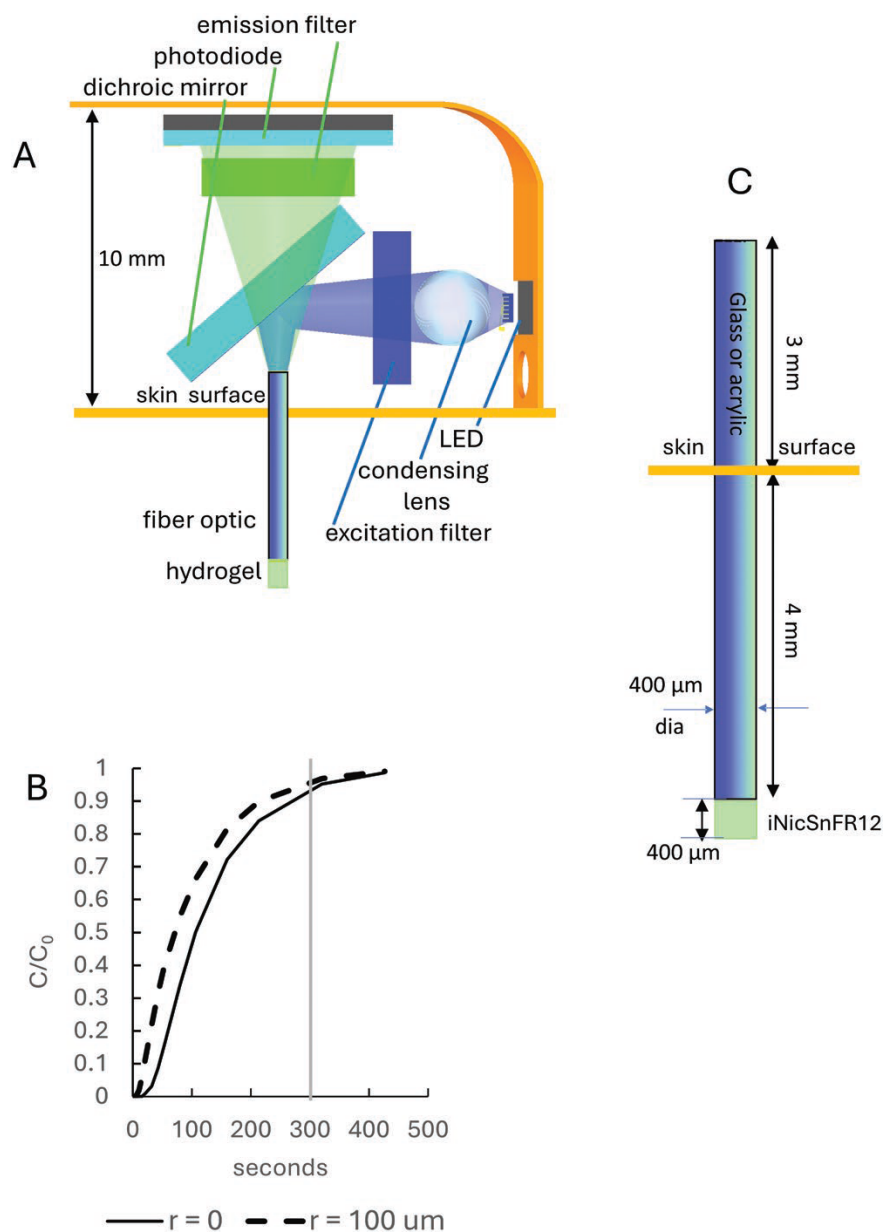


Figure 1.3 Design for a miniaturized fiber photometric CNM, based on initial progress with iNicSnFR12 entrapped in a hydrogel. [97] (A) The iNicSnFR-containing hydrogel is placed in ISF, at the end of a fiber optic inserted by a spring apparatus like those in some CGMs [98]. The miniature fiber photometer is 14 mm tall, including the protruding fiber optic. There are two time-multiplexed excitation LEDs. Excitation at 470 and 405 nm yield nicotine-dependent and nicotine-insensitive fluorescence, respectively [43, 99], a common tactic to control for movement artifacts

and bleaching. The SolidWorks file is available from the authors. **(B)** Simulations of radial diffusion in a cylindrical hydrogel [100, 101]. A hydrogel 400 μm in diameter contains 20 μM iNicSnFR. We assume that nicotine has a diffusion constant, D , of 0.3 $\mu\text{m}^2/\text{ms}$ in the hydrogel containing no iNicSnFR ($\sim 2/3$ of the free-solution D). Effective D is further reduced by rebinding to iNicSnFR molecules with a K_D of 10 μM . Following a jump from $[\text{nicotine}] = 0$ to C_0 at time 0 in the external solution, $[\text{nicotine}]_r$ within the gel approaches 90% of C_0 within 300 s, satisfying the criterion for temporal resolution. **(C)** In the design, the hydrogel contains 6.28 pmol of iNicSnFR, 20 μM . This is close to the 10 pmol of our standard assay condition in which a 100 μL microtiter well contains 100 nM iNicSnFR12 (see **Figure 1.2**).

We presented preliminary data for an approach in which the purified iNicSnFR molecules are trapped within a hydrogel, at the tip of a fiber optic [97]. The hydrogel would be surrounded by ISF. The hydrogel approach (1) prevents other proteins from interacting with sensor molecules, distorting F_0 or ΔF , (2) avoids increasing F_0 by endogenous cellular fluorescence, and (3) minimizes contact with immune cells. The purified protein-hydrogel approach differs from most cpGFP sensors, which are expressed in animal models via viral vectors or transgenes. Points (1) and (2) also allow us to plan ways for absolute calibrations of δ -slope (see **Figure 1.2C**).

The Redox-Modified PBP Method

This method (**Figure 1.4A**) was suggested as early as 2001 [102], then lay fallow until 2021. A recent report confirmed the concept, for continuously sensing glutamine with the eponymously named glutamine binding protein [103].

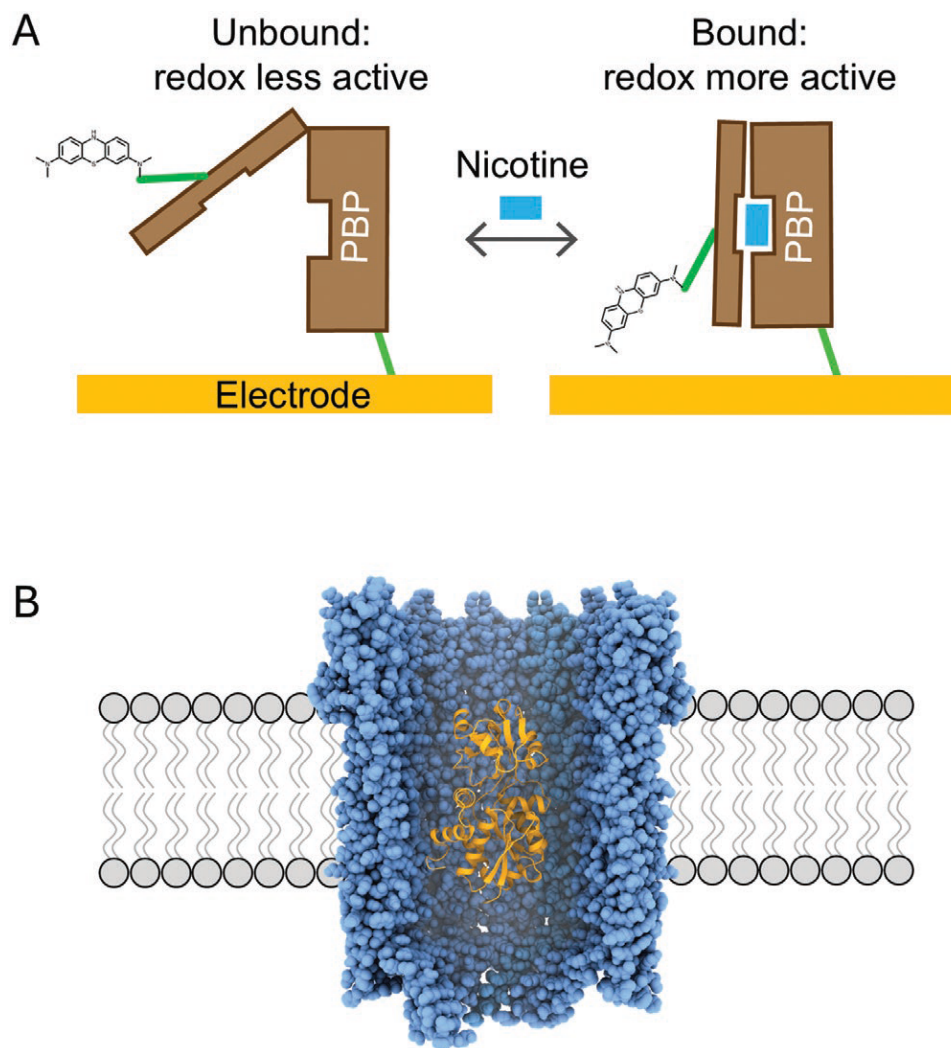


Figure 1.4 Electrochemical approaches to a CNM. (A) PBP bearing a redox group. The exemplar group is methylene blue. (B) Conceptual view of a PBP in a nanopore. The apo-PBP from iNiCSnFR3a (excised from pdb file 7S7W) was manually positioned within the vestibule of hemolysin E from *E. Coli K12* (pdb 2WCD), using ChimeraX. The complex was conceptually positioned in a bilayer membrane.

Site-selective modification would attach the redox probe to the nicotine-binding PBP at a sequence position that experiences an acceptably large movement (10–20 Å) upon ligand binding (see **Figures 1.2B** and **1.4A**). We have defined several candidate residues, based on liganded and unliganded structures of iNicSnFR12. The original previous study,

with other ligand–PBP pairs, coupled redox-active groups to introduced cysteine residues [102, 103]; and we have begun with this tactic. Our preliminary experiments have utilized Cys variants of iNicSnFR12 and also the PBP alone, without the cpGFP moiety (“PBP12”).

For the redox probe, one can choose among more than a dozen molecules [104]. We have begun experiments with methylene blue. Phenazine derivatives [103] and ruthenium (Ru(II)) [102] have also been used with PBPs. In the most general tactic, one would site-selectively introduce noncanonical azido amino acids. Our preliminary studies have produced signals that show dose dependence on nicotine. However, we are still experimenting with conditions that yield satisfactory control data and also yield reversibility suitable for continuous sensing.

The electrodes for the intradermal version of the electrochemical CNM would be subcutaneous gold wires (0.5 mm diameter, 5 mm length) resembling those in CGMs—though with the more sophisticated analog processing required by pulsed techniques such as square-wave voltammetry. In another variation, the PBP could be coupled to a graphene monolayer that acts as the gate of a field-effect transistor [105].

The PBP-Gated Nanopore Approach

A single nicotine binding event could, in principle, be amplified electrically if the binding gates a channel or a nanopore that passes 10^8 ions/s (a typical single-channel current) while the nicotine is bound (**Figure 1.4B**) [106, 107]. In a pioneering report, *Salmonella typhi* nanopores were embedded in planar bilayers. Conventional electrophysiological amplifiers and software were used in single-channel mode to analyze the nanopore current. Each of 13 different metabolite-binding PBPs proved suitable for ligand detection [108]. These were expressed in bacterial systems, purified, and added to the chambers separated by the membrane. The PBPs produced a range of fluctuations in the nanopore current, typically on a time scale of seconds. Adding the cognate ligand for the PBP then produced additional current fluctuations. Usually the liganded (“closed”)

conformation of the PBP blocked the current more completely, suggesting that “as the structure of the protein becomes more compacted, the protein penetrates deeper inside the nanopore, resulting in more current being blocked.” Importantly, the EC_{50} of the ligand-induced current fluctuations was similar in all cases to the ligand–PBP interaction in solution.

Thus, the PBP-gated nanopore approach is biophysically tractable and pharmacologically understandable. Although no PBP was studied in the class F family which contains OpuBC (the molecule that was evolved to yield PBP12) [92], these data are encouraging.

A deployable CNM need not be constrained by fragile lipid bilayers and single-channel measurements. The membrane would be a synthetic polymer, as used in pore sequencing; and many PBP12–nanopore complexes would be embedded in parallel, providing a macroscopic signal proportional to [nicotine]. This would utilize presently available high-throughput planar electrophysiological methods, for assaying candidate variants in a protein engineering campaign that develops and optimizes the PBP–nanopore complex. The supported membrane and electrodes would then be fixed to the tip of CGM-like probe.

Nicotine-Binding Aptamer-Based Approaches

In the present context, an aptamer (a synthetic single-stranded oligonucleotide) would be attached to an electrode. The aptamer would be evolved or selected to bind a nicotine molecule, changing shape. This would change the distance between an attached redox-active group and the electrode, resulting in a signal analogous to the redox-modified PBP [109]. Electrochemical aptamer-based sensors have been developed for metabolites, therapeutic drugs, and abused drugs; but unfortunately, there is not yet a reported aptamer with selective affinity for nicotine. Furthermore, we expect a poorer dynamic range and lifetime from aptamers compared to fluorescent hydrogels [110].

Nicotine-Oxidizing Enzymes

In part because amperometric glucose sensors have found widespread use, investigators have considered oxidoreductase enzymes that already use nicotine as a substrate. The major challenge arises because we have specified that $[\text{nicotine}]_t$ must be measured at 10^6 -fold lower levels than $[\text{glucose}]_t$, and each nicotine molecule that becomes oxidized by the enzyme contributes only a single elementary charge to the current.

Nicotine oxidase from *Pseudomonas putida* NicA2, was studied with amperometry [111]. Selectivity for nicotine is excellent. A mutated version became the basis for an issued patent [112], aimed primarily at measurements in sweat. The limit of detection was 1 μM , yielding a current of 0.01 μA for an electrode area that was probably 3 mm^2 . An intradermal electrode would presumably have a 100-fold smaller area. This would give a signal to 10 nM nicotine of just 1 pA, a challenge to measure. Studies on *Shinella* sp. HZN7 NctB, another nicotine-oxidizing enzyme, show that mutations can increase turnover number only ~ 2 -fold from the wild-type enzyme [113].

The study of Tai et al. [82] claimed to measure the current associated with CYP2B6 oxidation of nicotine [114]. There was no control for the additional oxidative currents; it is likely that the enzyme simply increased previously measured faradaic currents [114] by ~ 2 -fold, leaving the sensitivity at ~ 100 -fold less than our goals.

Data Analysis

Pioneering instruments generally produce suboptimal data. It may not be possible for a CNM to detect when a subject takes a puff or otherwise ingests nicotine. We plan to equip the initial iNicSnFR12-based CNMs with three pushbuttons. Buttons 1, 2, and 3 “time-stamp” the signal at the beginning of a cigarette, a vaping session, or a nicotine pouch, respectively. In simulations, vast reductions in noise can result (**Figure 1.5**).

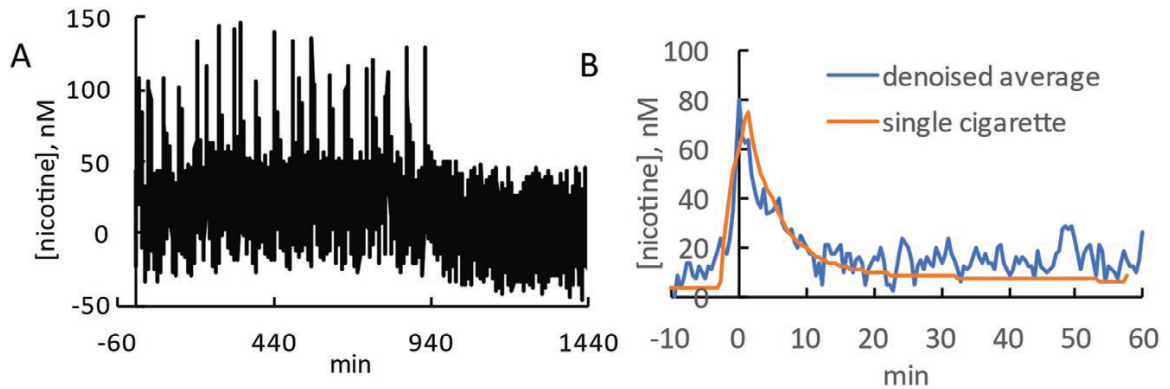


Figure 1.5 Synchronized averaging will reduce noise in CNM recordings. (A) A simulated noisy CNM data stream. The dose table for the simulator [73] contained observed ad libitum times for one smoker [68] (21 cigarettes, total; we assumed 1 mg doses). We added white noise, single-pole filtered at $\tau = 5$ min (the target temporal resolution) with an RMS deviation of 29 nM (roughly the present resolution of iNicSnFR12). The resulting $[\text{nicotine}]_t$ signal is poorly discernible. (B) Using the dose table, we retrospectively synchronized and averaged the traces during 10 min before and 60 min following the beginning of each cigarette. As expected, this decreased the noise by a factor $\sqrt{21}$. The resulting denoised average trace approximates the peak and waveform of the noiseless simulated $[\text{nicotine}]_t$ for a single cigarette. RMS: root mean square.

It would be desirable to automatically detect smoking/vaping episodes, so that subjects can ingest ad libitum without the added burden of time-stamping the dataset [115]. Machine learning algorithms are being developed to identify meal times and predict individual patients' glucose trajectory from present CGM measurements [115, 116]. How much of this effort is useful for the 24-hour measurements we envision for the CNM? We suggest that ~ 500 subject-days could provide an initial time-stamped dataset analogous to the glucose sensor time series in the OhioT1DM dataset, which contain 56 days $\times$ 8 type 1 diabetic patients [117]. This dataset could then provide a benchmark for developing new smoothing and curve-fitting algorithms.

The Caltech authors include a group with expertise in developing wearable devices. Like several other groups, we have reported on-device analog and digital signal processing, Bluetooth-based wireless communication, and battery power. These seem suitable for a CNM [118, 119].

A Not-for-Profit, Collaborative Research Agenda: Summary, Economics, and Support

Figure 1.6 summarizes this review in a cartoon form. Developing a CNM is a research agenda; and the outcome will be primarily a wearable instrument for research on the neuroscience of human nicotine use.

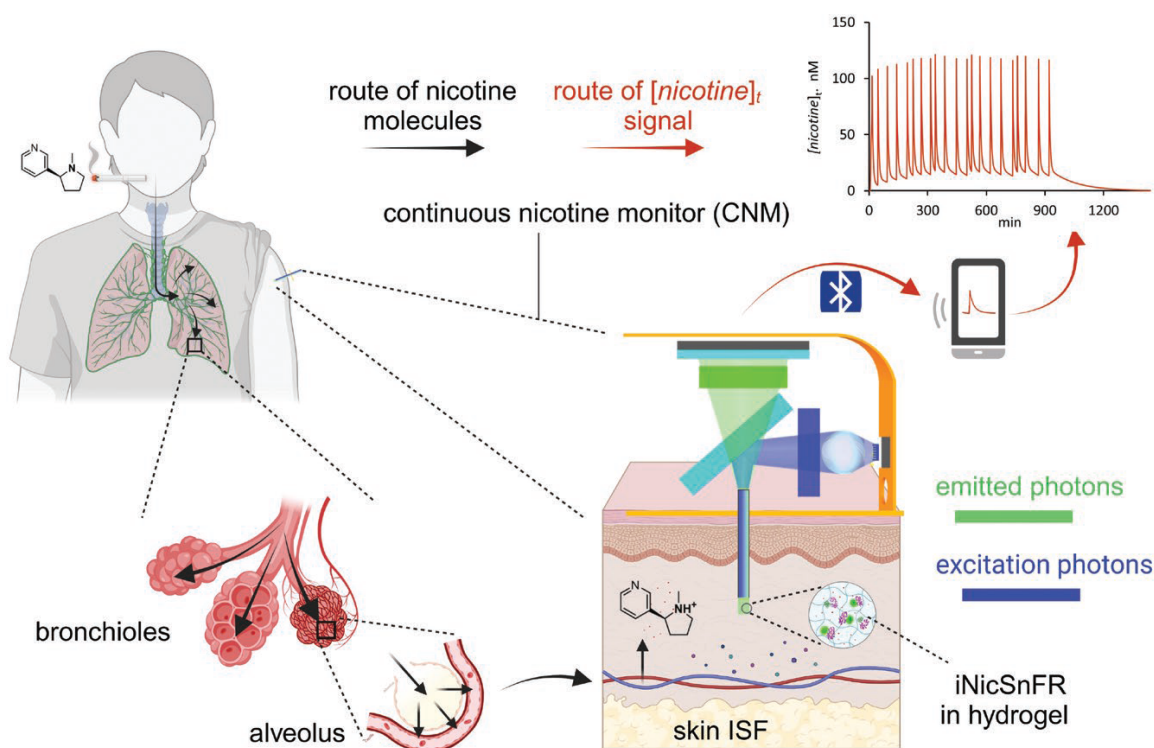


Figure 1.6 Graphical Summary. The figure illustrates this review’s stated motivation for measurements of individual $[\text{nicotine}]_t$, our present opinions on the most likely path to achieve a wearable CNM, and our suggestions for further analysis of $[\text{nicotine}]_t$. The lungs, especially the bronchioles and alveoli, are the most frequent route for nicotine to reach the bloodstream. Because ISF $[\text{nicotine}]_t$ strongly resembles blood and

cerebrospinal fluid (CSF) $[\text{nicotine}]_t$ (see text), a minimally invasive intradermal CNM provides appropriate measurements of $[\text{nicotine}]_t$. Entrapping a purified protein of the iNicSnFR family in a hydrogel has promise as a molecular sensing strategy with the stated specifications. During ad libitum smoking or vaping, a photometric CNM can also incorporate compensation and normalization algorithms (not shown in detail) to result in the $[\text{nicotine}]_t$ signal. The wearable CNM communicates $[\text{nicotine}]_t$ via Bluetooth to a portable device. The $[\text{nicotine}]_t$ trace shown is simulated from the smoking times observed for ad libitum smoking by one exemplar subject over the time course of one day [68, 73]. In the future, further time series analysis of many actual individual $[\text{nicotine}]_t$ records can enlighten research on nicotine and tobacco. Created partially in BioRender.

The example we have given, PATH-like research programs, might call for 30 000 CNMs/year. At these quantities, in the design of **Figure 1.3**, each thin-film wafer would be cut to yield 75 optical components (excitation filter, dichroic mirror, or emission filter). The total cost of these three components would be ~\$18 per CNM. The electronics, including a Bluetooth chip with sufficient processing power to normalize and correct the signal, would cost ~\$22 per CNM. Other components, including purified iNicSnFR protein, plastic fiber optic, lens, 3D-printed body, and battery, would have a total cost of ~\$20. Therefore, we envision a cost of <\$100 per CNM. If each subject in a PATH-like study wears a CNM for 24 hours in association with each annual interview, the additional cost will be \$3 M/year—a single-digit percentage of the PATH budget.

We advocate a research agenda analogous to the ongoing, highly collaborative, international development of two other neuroscience instruments: miniature fluorescent microscopes [120] and high-density multielectrode probes [121]. Initial experiments can and must use rodent models; but unlike the two examples given, the research agenda must progress to measurements in humans with all deliberate speed. The medical device industry is largely driven by clinician uptake and insurance reimbursement; however, no reimbursement path for a CNM exists today. Therefore, the tobacco control, smoking

cessation, and regulatory communities must decide whether public sources or a nonprofit startup [122] provide the most appropriate route for supporting the development of a CNM.

Acknowledgements

H.A.L. thanks the Society for Research on Nicotine and Tobacco for the invitation to prepare this material for the 2024 Langley Lecture. We thank Mario Danek, Eric Donny, Ryan Drenan, Jed Rose, Rachel Tyndale, and Mitchell Zeller for helpful discussions. The Environmental Exposure and Toxicology Study Section of the California Tobacco-Related Disease Research Program (TRDRP) and the Neurobiology of Motivated Behavior Study Section of the National Institutes of Health provided written critiques of the motivation for a continuous nicotine monitor.

Bibliography

1. Henningfield JE, Stapleton JM, Benowitz NL, Grayson RF, London ED. Higher levels of nicotine in arterial than in venous blood after cigarette smoking. *Drug Alcohol Depend.* 1993;33(1):23-9. 10.1016/0376-8716(93)90030-t
2. Rose JE, Behm FM, Westman EC, Coleman RE. Arterial nicotine kinetics during cigarette smoking and intravenous nicotine administration: implications for addiction. *Drug Alcohol Depend.* 1999;56(2):99-107. 10.1016/s0376-8716(99)00025-3
3. Hukkanen J, Jacob P, 3rd, Benowitz NL. Metabolism and disposition kinetics of nicotine. *Pharmacol Rev.* 2005;57(1):79-115.
4. Wall A, Roslin S, Borg B, McDermott S, Walele T, Nahde T, O'Connell G, Thompson J, Lubberink M, Antoni G. E-Cigarette Aerosol Deposition and Disposition of [¹¹C]Nicotine Using Positron Emission Tomography: A Comparison of Nicotine Uptake in Lungs and Brain Using Two Different Nicotine Formulations. *Pharmaceuticals (Basel).* 2022;15(3). 10.3390/ph15030367
5. Bowen A, Xing C, inventors Nicotine salt formulations for aerosol devices and methods thereof. United States patent 9,215,895. 2015.
6. Brossard P, Weitkunat R, Poux V, Lama N, Haziza C, Picavet P, Baker G, Ludicke F. Nicotine pharmacokinetic profiles of the Tobacco Heating System 2.2, cigarettes and nicotine gum in Japanese smokers. *Regul Toxicol Pharmacol.* 2017;89:193-9. 10.1016/j.yrtph.2017.07.032
7. Royal College of Physicians. Nicotine without smoke: Tobacco harm reduction. London: ECP; 2016.
8. Lunell E, Fagerstrom K, Hughes J, Pendrill R. Pharmacokinetic Comparison of a Novel Non-tobacco-Based Nicotine Pouch (ZYN) With Conventional, Tobacco-Based Swedish Snus and American Moist Snuff. *Nicotine Tob Res.* 2020;22(10):1757-63. 10.1093/ntr/ntaa068
9. Henderson BJ, Lester HA. Inside-out neuropharmacology of nicotinic drugs. *Neuropharmacol.* 2015;96(Pt B):178-93. 10.1016/j.neuropharm.2015.01.022
10. Kovar L, Selzer D, Britz H, Benowitz N, St Helen G, Kohl Y, Bals R, Lehr T. Comprehensive Parent-Metabolite PBPK/PD Modeling Insights into Nicotine Replacement Therapy Strategies. *Clin Pharmacokinet.* 2020;59(9):1119-34. 10.1007/s40262-020-00880-4
11. Alsharari SD, King JR, Nordman JC, Muldoon PP, Jackson A, Zhu AZ, Tyndale RF, Kabbani N, Damaj MI. Effects of Menthol on Nicotine Pharmacokinetic, Pharmacology and Dependence in Mice. *PLoS One.* 2015;10(9):e0137070. 10.1371/journal.pone.0137070
12. Benowitz NL, Herrera B, Jacob P, 3rd. Mentholated cigarette smoking inhibits nicotine metabolism. *J Pharmacol Exp Ther.* 2004;310(3):1208-15. 10.1124/jpet.104.066902
13. Henderson BJ, Wall TR, Henley BM, Kim CH, Nichols WA, Moaddel R, Xiao C, Lester HA. Menthol Alone Upregulates Midbrain nAChRs, Alters nAChR Subtype

- Stoichiometry, Alters Dopamine Neuron Firing Frequency, and Prevents Nicotine Reward. *J Neurosci*. 2016;36(10):2957-74. 10.1523/jneurosci.4194-15.2016
14. Brody A, Mukhin A, La Charite J, Ta K, Farahi J, Sugar C, Mamoun M, Vellios E, Archie M, Kozman M, Phuong J, Arlorio F, Mandelkern M. Up-regulation of nicotinic acetylcholine receptors in menthol cigarette smokers. *Int J Neuropsychopharmacol*. 2013;16:957-66.
 15. Benowitz NL, Hukkanen J, Jacob P, 3rd. Nicotine chemistry, metabolism, kinetics and biomarkers. *Handb Exp Pharmacol*. 2009(192):29-60. 10.1007/978-3-540-69248-5_2
 16. Schreiner BS, Lehmann R, Thiel U, Ziemba PM, Beltran LR, Sherkheli MA, Jeanbourquin P, Hugli A, Werner M, Gisselmann G, Hatt H. Direct action and modulating effect of (+)- and (-)-nicotine on ion channels expressed in trigeminal sensory neurons. *Eur J Pharmacol*. 2014;728:48-58. 10.1016/j.ejphar.2014.01.060
 17. McKee SA, McRae-Clark AL. Consideration of sex and gender differences in addiction medication response. *Biol Sex Differ*. 2022;13(1):34. 10.1186/s13293-022-00441-3
 18. Maher EE, Strzelecki AM, Weafer JJ, Gipson CD. The importance of translationally evaluating steroid hormone contributions to substance use. *Front Neuroendocrinol*. 2023;69:101059. 10.1016/j.yfrne.2023.101059
 19. Klemperer EM, Kock L, Feinstein MJP, Coleman SRM, Gaalema DE, Higgins ST. Sex differences in tobacco use, attempts to quit smoking, and cessation among dual users of cigarettes and e-cigarettes: Longitudinal findings from the US Population Assessment of Tobacco and Health Study. *Prev Med*. 2024;185:108024. 10.1016/j.ypmed.2024.108024
 20. Benowitz NL, Lessov-Schlaggar CN, Swan GE, Jacob P, 3rd. Female sex and oral contraceptive use accelerate nicotine metabolism. *Clin Pharmacol Ther*. 2006;79(5):480-8. 10.1016/j.clpt.2006.01.008
 21. Ren M, Lotfipour S, Leslie F. Unique effects of nicotine across the lifespan. *Pharmacol Biochem Behav*. 2022;214:173343. 10.1016/j.pbb.2022.173343
 22. Molander L, Hansson A, Lunell E. Pharmacokinetics of nicotine in healthy elderly people. *Clinical Pharmacology & Therapeutics*. 2001;69(1):57-65. 10.1067/mcp.2001.113181
 23. Pérez-Martín H, Lidón-Moyano C, González-Marrón A, Fu M, Pérez-Ortuño R, Ballbè M, Martín-Sánchez JC, Pascual JA, Fernández E, Martínez-Sánchez JM. Variation in Nicotine Metabolization According to Biological Factors and Type of Nicotine Consumer. *Healthcare [Internet]*. 2023; 11(2):[179 p.].
 24. Hatsukami DK, Luo X, Jensen JA, al'Absi M, Allen SS, Carmella SG, Chen M, Cinciripini PM, Denlinger-Apte R, Drobles DJ, Koopmeiners JS, Lane T, Le CT, Leischow S, Luo K, McClernon FJ, Murphy SE, Paiano V, Robinson JD, Severson H, Sipe C, Strasser AA, Strayer LG, Tang MK, Vandrey R, Hecht SS, Benowitz NL, Donny EC. Effect of Immediate vs Gradual Reduction in Nicotine Content of Cigarettes on Biomarkers of Smoke Exposure: A Randomized Clinical Trial. *Jama*. 2018;320(9):880-91. 10.1001/jama.2018.11473

25. Benowitz NL, Donny EC, Edwards KC, Hatsukami D, Smith TT. The Role of Compensation in Nicotine Reduction. *Nicotine Tob Res.* 2019;21(Suppl 1):S16-s8. 10.1093/ntr/ntz120
26. Krebs NM, Zhu J, Wasserman E, Kuprewicz R, Martinez DJ, Veldheer S, Livelsberger C, Modesto J, Reinhart L, Trushin N, Reilly SM, Liao J, Fazzi A, Bascom R, Richie JP, Foulds J, Horn K, Muscat JE. Switching to Progressively Reduced Nicotine Content Cigarettes in Smokers With Low Socioeconomic Status: A Double-Blind Randomized Clinical Trial. *Nicotine Tob Res.* 2021;23(6):992-1001. 10.1093/ntr/ntaa247
27. Cox S, Goniewicz ML, Kosmider L, McRobbie H, Kimber C, Dawkins L. The Time Course of Compensatory Puffing With an Electronic Cigarette: Secondary Analysis of Real-World Puffing Data With High and Low Nicotine Concentration Under Fixed and Adjustable Power Settings. *Nicotine Tob Res.* 2021;23(7):1153-9. 10.1093/ntr/ntab013
28. Glantz SA, Nguyen N, Oliveira da Silva AL. Population-Based Disease Odds for E-Cigarettes and Dual Use versus Cigarettes. *NEJM Evid.* 2024;3(3):EVIDoa2300229. 10.1056/EVIDoa2300229
29. Gottlieb S, Zeller M. A Nicotine-Focused Framework for Public Health. *N Engl J Med.* 2017;377(12):1111-4. 10.1056/NEJMp1707409
30. Fink JB, Stapleton KW. Nebulizers. *J Aerosol Med Pulm Drug Deliv.* 2024;37(3):140-56. 10.1089/jamp.2024.29110.jbf
31. Knoch M. New Generation Nebulizers. *J Aerosol Med Pulm Drug Deliv.* 2024;37(3):157-65. 10.1089/jamp.2024.29116.mk
32. Danek M, inventor; Respira Technologies, Inc, assignee. Apparatus for producing an aerosol for inhalation by a person patent US 10,888,117 B2. 2021.
33. Danek M, inventor; Qnovia, Inc, assignee. Apparatus for producing an aerosol for inhalation by a person patent US11690963B2. 2024.
34. Danek M, Betts K, Kovacevich I, BANSANSINE J, Walsh J, Kar-Heng Cheng C, Breen C, RIGBERG J, Granger T, RAWASHDEH M, Hall R, CHARLES T, COKER J, inventors; Qnovia, Inc, assignee. Electronic devices for aerosolizing and inhaling liquid having diaphragm and a pressure sensor patent US11925207B2. 2024.
35. Businesswire. Qnovia, Inc. announces positive results from first-in-human pharmacokinetic and safety study of lead asset QN-01, a prescription inhaled smoking cessation therapy 2023 [Available from: <https://www.businesswire.com/news/home/20231129757219/en/>].
36. El Hourani M, Shihadeh A, Talih S, Eissenberg T. Comparison of nicotine emissions rate, 'nicotine flux', from heated, electronic and combustible tobacco products: data, trends and recommendations for regulation. *Tobacco Control.* 2023;32(e2):e180-e3. 10.1136/tobaccocontrol-2021-056850
37. Kolli AR, Veljkovic E, Calvino-Martin F, Esposito M, Kuczaj AK, Koumal O, Rose JE, Peitsch MC. Nicotine flux and pharmacokinetics-based considerations for early assessment of nicotine delivery systems. *Drug Alcohol Depend Rep.* 2024;11:100245. 10.1016/j.dadr.2024.100245
38. Jill M. Williams MD, Jonathan Foulds PD. Treating Addiction to Tobacco and Nicotine Products: American Psychiatric Association Publishing; 2025.

39. Schwartz RD, Kellar KJ. *In vivo* regulation of [³H]acetylcholine recognition sites in brain by nicotinic cholinergic drugs. *J Neurochem.* 1985;45(2):427-33.
40. Bencherif M, Fowler K, Lukas RJ, Lippiello PM. Mechanisms of up-regulation of neuronal nicotinic acetylcholine receptors in clonal cell lines and primary cultures of fetal rat brain. *J Pharmacol Exp Ther.* 1995;275(2):987-94.
41. Kuryatov A, Luo J, Cooper J, Lindstrom J. Nicotine acts as a pharmacological chaperone to up-regulate human $\alpha 4\beta 2$ acetylcholine receptors. *Mol Pharmacol.* 2005;68(6):1839-51.
42. Nashmi R, Xiao C, Deshpande P, McKinney S, Grady SR, Whiteaker P, Huang Q, McClure-Begley T, Lindstrom JM, Labarca C, Collins AC, Marks MJ, Lester HA. Chronic nicotine cell specifically upregulates functional $\alpha 4^*$ nicotinic receptors: basis for both tolerance in midbrain and enhanced long-term potentiation in perforant path. *J Neurosci.* 2007;27(31):8202-18. 10.1523/JNEUROSCI.2199-07.2007
43. Shivange AV, Borden PM, Muthusamy AK, Nichols AL, Bera K, Bao H, Bishara I, Jeon J, Mulcahy MJ, Cohen B, O'Riordan SL, Kim C, Dougherty DA, Chapman ER, Marvin JS, Looger LL, Lester HA. Determining the pharmacokinetics of nicotinic drugs in the endoplasmic reticulum using biosensors. *J Gen Physiol* 2019;151(4):738-57. 10.1085/jgp.201812201
44. Benowitz NL, Jacob P, 3rd, Denaro C, Jenkins R. Stable isotope studies of nicotine kinetics and bioavailability. *Clin Pharmacol Ther.* 1991;49(3):270-7.
45. Levi M, Dempsey DA, Benowitz NL, Sheiner LB. Population pharmacokinetics of nicotine and its metabolites I. Model development. *J Pharmacokinet Pharmacodyn.* 2007;34(1):5-21. 10.1007/s10928-006-9027-z
46. Teeguarden JG, Housand CJ, Smith JN, Hinderliter PM, Gunawan R, Timchalk CA. A multi-route model of nicotine-cotinine pharmacokinetics, pharmacodynamics and brain nicotinic acetylcholine receptor binding in humans. *Regul Toxicol Pharmacol.* 2013;65(1):12-28. 10.1016/j.yrtph.2012.10.007
47. Marchand M, Brossard P, Merdjan H, Lama N, Weitkunat R, Ludicke F. Nicotine Population Pharmacokinetics in Healthy Adult Smokers: A Retrospective Analysis. *Eur J Drug Metab Pharmacokinet.* 2017;42(6):943-54. 10.1007/s13318-017-0405-2
48. Matta SG, Balfour DJ, Benowitz NL, Boyd RT, Buccafusco JJ, Caggiula AR, Craig CR, Collins AC, Damaj MI, Donny EC, Gardiner PS, Grady SR, Heberlein U, Leonard SS, Levin ED, Lukas RJ, Markou A, Marks MJ, McCallum SE, Parameswaran N, Perkins KA, Picciotto MR, Quik M, Rose JE, Rothenfluh A, Schafer WR, Stolerman IP, Tyndale RF, Wehner JM, Zirger JM. Guidelines on nicotine dose selection for *in vivo* research. *Psychopharm.* 2007;190(3):269-319.
49. Hansen E, Arendt-Nielsen L, Boudreau SA. A Comparison of Oral Sensory Effects of Three TRPA1 Agonists in Young Adult Smokers and Non-smokers. *Front Physiol.* 2017;8:663. 10.3389/fphys.2017.00663
50. Carstens E, Carstens MI. Sensory Effects of Nicotine and Tobacco. *Nicotine Tob Res.* 2022;24(3):306-15. 10.1093/ntr/ntab086
51. Fu XW, Lindstrom J, Spindel ER. Nicotine activates and up-regulates nicotinic acetylcholine receptors in bronchial epithelial cells. *Am J Respir Cell Mol Biol.* 2009;41(1):93-9.

52. Mazzo F, Pistillo F, Grazioso G, Clementi F, Borgese N, Gotti C, Colombo SF. Nicotine-Modulated Subunit Stoichiometry Affects Stability and Trafficking of $\alpha 3\beta 4$ Nicotinic Receptor. *J Neurosci*. 2013;33(30):12316-28. 10.1523/jneurosci.2393-13.2013
53. Claw KG, Beans JA, Lee SB, Avey JP, Stapleton PA, Scherer SE, El-Boraie A, Tyndale RF, Nickerson DA, Dillard DA, Thummel KE, Robinson RF. Pharmacogenomics of nicotine metabolism: novel CYP2A6 and CYP2B6 genetic variation patterns in Alaska Native and American Indian populations. *Nicotine Tob Res*. 2019. 10.1093/ntr/ntz105
54. Howard LA, Ahluwalia JS, Lin SK, Sellers EM, Tyndale RF. CYP2E1*1D regulatory polymorphism: association with alcohol and nicotine dependence. *Pharmacogenetics*. 2003;13(6):321-8. 10.1097/01.fpc.0000054090.48725.a2
55. Cao Y, Liu X, Hu Z, Li J, Chen X, Xiong Y, Zheng F, Zhang J, Zhang L, Liu X. Assessing Nicotine Pharmacokinetics of New-Generation Tobacco Products and Conventional Cigarettes: A Systematic Review and Meta-analysis. *Nicotine & Tobacco Research*. 2024. 10.1093/ntr/ntae199
56. Benowitz NL, Jacob P, 3rd. Metabolism of nicotine to cotinine studied by a dual stable isotope method. *Clin Pharmacol Ther*. 1994;56(5):483-93.
57. Yingst JM, Foulds J, Veldheer S, Hrabovsky S, Trushin N, Eissenberg TT, Williams J, Richie JP, Nichols TT, Wilson SJ, Hobkirk AL. Nicotine absorption during electronic cigarette use among regular users. *PLoS One*. 2019;14(7):e0220300. 10.1371/journal.pone.0220300
58. Yingst JM, Hrabovsky S, Hobkirk A, Trushin N, Richie JP, Jr., Foulds J. Nicotine Absorption Profile Among Regular Users of a Pod-Based Electronic Nicotine Delivery System. *JAMA Netw Open*. 2019;2(11):e1915494. 10.1001/jamanetworkopen.2019.15494
59. Solingapuram Sai KK, Zuo Y, Rose JE, Garg PK, Garg S, Nazih R, Mintz A, Mukhin AG. Rapid Brain Nicotine Uptake from Electronic Cigarettes. *J Nucl Med*. 2020;61(6):928-30. 10.2967/jnumed.119.230748
60. Zuo Y, Mukhin AG, Berg H, Morgan JD, Mintz A, Rose JE, Solingapuram Sai KK. Comparison of brain nicotine uptake from electronic cigarettes and combustible cigarettes. *Neuropsychopharmacology*. 2022;47(11):1939-44. 10.1038/s41386-022-01410-5
61. Dempsey D, Tutka P, Jacob P, 3rd, Allen F, Schoedel K, Tyndale RF, Benowitz NL. Nicotine metabolite ratio as an index of cytochrome P450 2A6 metabolic activity. *Clin Pharmacol Ther*. 2004;76(1):64-72. 10.1016/j.clpt.2004.02.011
62. Tanner JA, Tyndale RF. Variation in CYP2A6 Activity and Personalized Medicine. *J Pers Med*. 2017;7(4). 10.3390/jpm7040018
63. El-Boraie A, Taghavi T, Chenoweth MJ, Fukunaga K, Mushiroda T, Kubo M, Lerman C, Nollen NL, Benowitz NL, Tyndale RF. Evaluation of a weighted genetic risk score for the prediction of biomarkers of CYP2A6 activity. *Addict Biol*. 2020;25(1):e12741. 10.1111/adb.12741
64. Chenoweth MJ, Lerman C, Knight J, Tyndale RF. Influence of CYP2A6 Genetic Variation, Nicotine Dependence Severity, and Treatment on Smoking Cessation Success. *Nicotine Tob Res*. 2023;25(6):1207-11. 10.1093/ntr/ntac268

65. Benowitz NL. Cigarette smoking and cardiovascular disease: pathophysiology and implications for treatment. *Prog Cardiovasc Dis*. 2003;46(1):91-111. 10.1016/s0033-0620(03)00087-2
66. Mendelson JH, Goletiani N, Sholar MB, Siegel AJ, Mello NK. Effects of smoking successive low- and high-nicotine cigarettes on hypothalamic-pituitary-adrenal axis hormones and mood in men. *Neuropsychopharmacology*. 2008;33(4):749-60. 10.1038/sj.npp.1301455
67. St Helen G, Ross KC, Dempsey DA, Havel CM, Jacob P, 3rd, Benowitz NL. Nicotine Delivery and Vaping Behavior During ad Libitum E-cigarette Access. *Tob Regul Sci*. 2016;2(4):363-76. 10.18001/TRS.2.4.8
68. Pauwels CGGM, Boots AW, Visser WF, Pennings JLA, Talhout R, Van Schooten F-J, Opperhuizen A. Characteristic Human Individual Puffing Profiles Can Generate More TNCO than ISO and Health Canada Regimes on Smoking Machine When the Same Brand Is Smoked. *International Journal of Environmental Research and Public Health*. 2020;17(9):erph3225.
69. Smith D, Allerton C, Kalgutkar A, van de Waterbeemd H, Walker D. *Pharmacokinetics and Metabolism in Drug Design*. 3rd ed. Mannhold R, editor. Weinheim: Wiley; 2012.
70. Olsson Gisleskog PO, Perez Ruixo JJ, Westin A, Hansson AC, Soons PA. Nicotine Population Pharmacokinetics in Healthy Smokers After Intravenous, Oral, Buccal and Transdermal Administration. *Clin Pharmacokinet*. 2021;60(4):541-61. 10.1007/s40262-020-00960-5
71. NCBI. PubChem Nicotine [CID=89594]. Available from: <https://pubchem.ncbi.nlm.nih.gov/compound/89594>
72. Miwa JM, Freedman R, Lester HA. Neural systems governed by nicotinic acetylcholine receptors: emerging hypotheses. *Neuron*. 2011;70(1):20-33. 10.1016/j.neuron.2011.03.014
73. Wang AZ, Jeon J, Drenan RM, Lester HA. Simulating early steps in nicotine dependence: Pharmacokinetics, activation, and chaperoning of nicotinic receptors. *Molecular Pharmacology*. 2025;107(4):100016. 10.1016/j.molpha.2025.100016
74. Rollema H, Shrikhande A, Ward KM, Tingley FD, 3rd, Coe JW, O'Neill BT, Tseng E, Wang EQ, Mather RJ, Hurst RS, Williams KE, de Vries M, Cremers T, Bertrand S, Bertrand D. Pre-clinical properties of the $\alpha 4\beta 2$ nicotinic acetylcholine receptor partial agonists varenicline, cytisine and dianicline translate to clinical efficacy for nicotine dependence. *Br J Pharmacol*. 2010;160(2):334-45. 10.1111/j.1476-5381.2010.00682.x
75. Pankow JF. A consideration of the role of gas/particle partitioning in the deposition of nicotine and other tobacco smoke compounds in the respiratory tract. *Chem Res Toxicol*. 2001;14(11):1465-81. 10.1021/tx0100901
76. Benowitz NL. The Central Role of pH in the Clinical Pharmacology of Nicotine: Implications for Abuse Liability, Cigarette Harm Reduction and FDA Regulation. *Clin Pharmacol Ther*. 2022;111(5):1004-6. 10.1002/cpt.2555
77. Zuo Y, Solingapuram Sai KK, Jazic A, Bansode AH, Rose JE, Mukhin AG. Comparison of brain nicotine accumulation from traditional combustible cigarettes

- and electronic cigarettes with different formulations. *Neuropsychopharmacology*. 2024;49(4):740-6. 10.1038/s41386-024-01800-x
78. Mortellaro M, DeHennis A. Performance characterization of an abiotic and fluorescent-based continuous glucose monitoring system in patients with type 1 diabetes. *Biosens Bioelectron*. 2014;61:227-31. 10.1016/j.bios.2014.05.022
 79. Irace C, Cutruzzola A, Tweden K, Kaufman FR. Device profile of the everSense continuous glucose monitoring system for glycemic control in type-1 diabetes: overview of its safety and efficacy. *Expert Rev Med Devices*. 2021;18(10):909-14. 10.1080/17434440.2021.1982380
 80. Sonner Z, Wilder E, Heikenfeld J, Kasting G, Beyette F, Swaile D, Sherman F, Joyce J, Hagen J, Kelley-Loughnane N, Naik R. The microfluidics of the eccrine sweat gland, including biomarker partitioning, transport, and biosensing implications. *Biomicrofluidics*. 2015;9(3):031301. 10.1063/1.4921039
 81. Lin H, Tan J, Zhu J, Lin S, Zhao Y, Yu W, Hojaiji H, Wang B, Yang S, Cheng X, Wang Z, Tang E, Yeung C, Emaminejad S. A programmable epidermal microfluidic valving system for wearable biofluid management and contextual biomarker analysis. *Nature communications*. 2020;11(1):4405. 10.1038/s41467-020-18238-6
 82. Tai LC, Ahn CH, Nyein HYY, Ji W, Bariya M, Lin Y, Li L, Javey A. Nicotine Monitoring with a Wearable Sweat Band. *ACS sensors*. 2020;5(6):1831-7. 10.1021/acssensors.0c00791
 83. Yang Y, Gao W. Wearable and flexible electronics for continuous molecular monitoring. *Chem Soc Rev*. 2019;48(6):1465-91. 10.1039/c7cs00730b
 84. Choi J, Chen S, Deng Y, Xue Y, Reeder JT, Franklin D, Oh YS, Model JB, Aranyosi AJ, Lee SP, Ghaffari R, Huang Y, Rogers JA. Skin-Interfaced Microfluidic Systems that Combine Hard and Soft Materials for Demanding Applications in Sweat Capture and Analysis. *Adv Healthc Mater*. 2021;10(4):e2000722. 10.1002/adhm.202000722
 85. Harshman SW, Strayer KE, Davidson CN, Pitsch RL, Narayanan L, Scott AM, Schaeublin NM, Wiens TL, Phelps MS, O'Connor ML, Mackowski NS, Barrett KN, Leyh SM, Eckerle JJ, Strang AJ, Martin JA. Rate normalization for sweat metabolomics biomarker discovery. *Talanta*. 2021;223(Pt 1):121797. 10.1016/j.talanta.2020.121797
 86. Min J, Tu J, Xu C, Lukas H, Shin S, Yang Y, Solomon SA, Mukasa D, Gao W. Skin-Interfaced Wearable Sweat Sensors for Precision Medicine. *Chem Rev*. 2023;123(8):5049-138. 10.1021/acs.chemrev.2c00823
 87. Hoss U, Budiman ES. Factory-Calibrated Continuous Glucose Sensors: The Science Behind the Technology. *Diabetes Technol Ther*. 2017;19(S2):S44-s50. 10.1089/dia.2017.0025
 88. Haloi N, Huang S, Nichols AL, Fine EJ, Friesenhahn NJ, Marotta CB, Dougherty DA, Lindahl E, Howard RJ, Mayo SL, Lester HA. Interactive computational and experimental approaches improve the sensitivity of periplasmic binding protein-based nicotine biosensors for measurements in biofluids. *Protein Engineering, Design and Selection*. 2024;37:gzae003. 10.1093/protein/gzae003
 89. Gonzalez WG, Zhang H, Harutyunyan A, Lois C. Persistence of neuronal representations through time and damage in the hippocampus. *Science*. 2019;365(6455):821-5. 10.1126/science.aav9199

90. Muthusamy A, Rosenberg M, Kim C, Wang A, Ebisu H, TM C, A K, Marvin J, Cohen B, L L, Y O, Meister M, Lester H. Correspondence of fentanyl brain pharmacokinetics and behavior measured via engineering opioids biosensors and computational ethology. bioRxiv. 2024:2024.03.15.584894. 10.1101/2024.03.15.584894
91. Nichols AL, Blumenfeld Z, Fan C, Luebbert L, Blom AEM, Cohen BN, Marvin JS, Borden PM, Kim CH, Muthusamy AK, Shivange AV, Knox HJ, Campello HR, Wang JH, Dougherty DA, Looger LL, Gallagher T, Rees DC, Lester HA. Fluorescence activation mechanism and imaging of drug permeation with new sensors for smoking-cessation ligands. eLife. 2022;11:e74648. 10.7554/eLife.74648
92. Scheepers GH, Lycklama ANJA, Poolman B. An updated structural classification of substrate-binding proteins. FEBS Lett. 2016;590(23):4393-401. 10.1002/1873-3468.12445
93. Borden P, Shivange AV, Marvin JS, Cichon J, Dan C, Podgorski K, Novak O, Tanimoto M, Lobas M, Khakh B, Dittman J, Gan W, Koyama M, Jayaraman V, Zhu J, Lester HA, Looger LL. A genetically encoded fluorescent sensor for in vivo acetylcholine detection. bioRxiv. 2019.
94. Privett HK, Kiss G, Lee TM, Blomberg R, Chica RA, Thomas LM, Hilvert D, Houk KN, Mayo SL. Iterative approach to computational enzyme design. Proc Natl Acad Sci U S A. 2012;109(10):3790-5. 10.1073/pnas.1118082108
95. Blomberg R, Kries H, Pinkas DM, Mittl PR, Grutter MG, Privett HK, Mayo SL, Hilvert D. Precision is essential for efficient catalysis in an evolved Kemp eliminase. Nature. 2013;503(7476):418-21. 10.1038/nature12623
96. Unger EK, Keller JP, Altermatt M, Liang R, Matsui A, Dong C, Hon OJ, Yao Z, Sun J, Banala S, Flanigan ME, Jaffe DA, Hartanto S, Carlen J, Mizuno GO, Borden PM, Shivange AV, Cameron LP, Sinning S, Underhill SM, Olson DE, Amara SG, Temple Lang D, Rudnick G, Marvin JS, Lavis LD, Lester HA, Alvarez VA, Fisher AJ, Prescher JA, Kash TL, Yarov-Yarovoy V, Gradinaru V, Looger LL, Tian L. Directed Evolution of a Selective and Sensitive Serotonin Sensor via Machine Learning. Cell. 2020;183(7):1986-2002.e26. 10.1016/j.cell.2020.11.040
97. Nichols AL, Marotta CB, Frisenhahn NJ, Dalfonzo SR, Muthusamy A, Wagenaar DA, Mayo SL, Dougherty DA, Lester HA. Hydrogel Encapsulation of a Designed Fluorescent Protein Biosensor for Continuous Measurements of Sub-100 nM Nicotine. ACS Applied Bio Materials. 2026;9(2):765-74. 10.1021/acsabm.5c01418
98. Halac J, Gray JM, Johnston ND, England JD, Simpson PC, Neale PV, Blackwell J, Wells MNB, Pirondini K, Reinhardt AM, Kempkey MD, inventorsTranscutaneous analyte sensor systems and methods. United States patent 11,602,291. 2023.
99. Bera K, Kamajaya A, Shivange AV, Muthusamy AK, Nichols AL, Borden PM, Grant S, Jeon J, Lin E, Bishara I, Chin TM, Cohen BN, Kim CH, Unger EK, Tian L, Marvin JS, Looger LL, Lester HA. Biosensors Show the Pharmacokinetics of S-Ketamine in the Endoplasmic Reticulum. Frontiers in cellular neuroscience. 2019;13:499-. 10.3389/fncel.2019.00499
100. Carslaw H, Jaeger J. Conduction of heat in solids, Clarendon: Clarendon Press, Oxford; 1959.
101. Crank J. The Mathematics of Diffusion. Second ed. Oxford: Clarendon Press; 1975.

102. Benson DE, Conrad DW, de Lorimier RM, Trammell SA, Hellinga HW. Design of bioelectronic interfaces by exploiting hinge-bending motions in proteins. *Science*. 2001;293(5535):1641-4. 10.1126/science.1062461
103. Takamatsu S, Lee J, Asano R, Tsugawa W, Ikebukuro K, Sode K. Continuous electrochemical monitoring of L-glutamine using redox-probe-modified L-glutamine-binding protein based on intermittent pulse amperometry. *Sensors and Actuators B: Chemical*. 2021;346. 10.1016/j.snb.2021.130554
104. Kang D, Ricci F, White RJ, Plaxco KW. Survey of Redox-Active Moieties for Application in Multiplexed Electrochemical Biosensors. *Anal Chem*. 2016;88(21):10452-8. 10.1021/acs.analchem.6b02376
105. Krishnan SK, Nataraj N, Meyyappan M, Pal U. Graphene-Based Field-Effect Transistors in Biosensing and Neural Interfacing Applications: Recent Advances and Prospects. *Analytical Chemistry*. 2023;95(5):2590-622. 10.1021/acs.analchem.2c03399
106. Tanimoto IMF, Cressiot B, Greive SJ, Le Pioufle B, Bacri L, Pelta J. Focus on using nanopore technology for societal health, environmental, and energy challenges. *Nano Res*. 2022;15(11):9906-20. 10.1007/s12274-022-4379-2
107. Ying YL, Hu ZL, Zhang S, Qing Y, Fragasso A, Maglia G, Meller A, Bayley H, Dekker C, Long YT. Nanopore-based technologies beyond DNA sequencing. *Nat Nanotechnol*. 2022;17(11):1136-46. 10.1038/s41565-022-01193-2
108. Zernia S, van der Heide NJ, Galenkamp NS, Gouridis G, Maglia G. Current Blockades of Proteins inside Nanopores for Real-Time Metabolome Analysis. *ACS Nano*. 2020;14(2):2296-307. 10.1021/acsnano.9b09434
109. Dauphin-Ducharme P, Ploense KL, Arroyo-Currás N, Kippin TE, Plaxco KW. Electrochemical Aptamer-Based Sensors: A Platform Approach to High-Frequency Molecular Monitoring In Situ in the Living Body. *Methods Mol Biol*. 2022;2393:479-92. 10.1007/978-1-0716-1803-5_25
110. Arroyo-Currás N. Aptamers Can Be Effective Affinity Receptors for Biosensing. *ECS Sensors Plus*. 2024;3(3):030001. 10.1149/2754-2726/ad7bc7
111. Tararina MA, Xue S, Smith LC, Muellers SN, Miranda PO, Janda KD, Allen KN. Crystallography Coupled with Kinetic Analysis Provides Mechanistic Underpinnings of a Nicotine-Degrading Enzyme. *Biochemistry*. 2018;57(26):3741-51. 10.1021/acs.biochem.8b00384
112. Galagan J, Grinstaff M, Kuzmanovic U, Chen M, Alexandrovna M, Allen K, inventors; Trustees of Boston University,, assignee. Enzyme-Based Electrochemical Nicotine Biosensor. US patent 11,331,020. 2022.
113. Deay DO, 3rd, Seibold S, Battaile KP, Lovell S, Richter ML, Petillo PA. Improving the kinetic parameters of nicotine oxidizing enzymes by homologous structure comparison and rational design. *Arch Biochem Biophys*. 2022;718:109122. 10.1016/j.abb.2022.109122
114. Mehmeti E, Kilic T, Laur C, Carrara S. Electrochemical determination of nicotine in smokers' sweat. *Microchemical Journal*. 2020;158. 10.1016/j.microc.2020.105155
115. Zheng M, Ni B, Kleinberg S. Automated meal detection from continuous glucose monitor data through simulation and explanation. *J Am Med Inform Assoc*. 2019;26(12):1592-9. 10.1093/jamia/ocz159

116. Vettoretti M, Cappon G, Facchinetti A, Sparacino G. Advanced Diabetes Management Using Artificial Intelligence and Continuous Glucose Monitoring Sensors. *Sensors (Basel)*. 2020;20(14). 10.3390/s20143870
117. Marling C, Bunesco R. The OhioT1D Dataset [Available from: <http://smarthealth.cs.ohio.edu/OhioT1DM-dataset.html>].
118. Gao W, Emaminejad S, Nyein HYY, Challa S, Chen K, Peck A, Fahad HM, Ota H, Shiraki H, Kiriya D, Lien DH, Brooks GA, Davis RW, Javey A. Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis. *Nature*. 2016;529(7587):509-14. 10.1038/nature16521
119. Yang Y, Song Y, Bo X, Min J, Pak OS, Zhu L, Wang M, Tu J, Kogan A, Zhang H, Hsiai TK, Li Z, Gao W. A laser-engraved wearable sensor for sensitive detection of uric acid and tyrosine in sweat. *Nat Biotechnol*. 2020;38:217-24. 10.1038/s41587-019-0321-x
120. Aharoni D, Khakh BS, Silva AJ, Golshani P. All the light that we can see: a new era in miniaturized microscopy. *Nat Methods*. 2019;16(1):11-3. 10.1038/s41592-018-0266-x
121. Steinmetz NA, Aydin C, Lebedeva A, Okun M, Pachitariu M, Bauza M, Beau M, Bhagat J, Böhm C, Broux M, Chen S, Colonell J, Gardner RJ, Karsh B, Kloosterman F, Kostadinov D, Mora-Lopez C, O'Callaghan J, Park J, Putzeys J, Sauerbrei B, van Daal RJJ, Vollan AZ, Wang S, Welkenhuysen M, Ye Z, Dudman JT, Dutta B, Hantman AW, Harris KD, Lee AK, Moser EI, O'Keefe J, Renart A, Svoboda K, Häusser M, Haesler S, Carandini M, Harris TD. Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science*. 2021;372(6539). 10.1126/science.abf4588
122. Marblestone A, Gamick A, Kalil T, Martin C, Cvitkovic M, Rodrigues SG. Unblock research bottlenecks with non-profit start-ups. *Nature*. 2022;601(7892):188-90. 10.1038/d41586-022-00018-5

Chapter 2

INTERACTIVE COMPUTATIONAL AND EXPERIMENTAL APPROACHES IMPROVE THE SENSITIVITY OF PERIPLASMIC BINDING PROTEIN-BASED NICOTINE BIOSENSORS FOR MEASUREMENTS IN BIOFLUIDS

Adapted from:

Haloi N, Huang S, Nichols AL, Fine EJ, Friesenhahn NJ, Marotta CB, Dougherty DA, Lindahl E, Howard RJ, Mayo SL, Lester HA. Interactive computational and experimental approaches improve the sensitivity of periplasmic binding protein-based nicotine biosensors for measurements in biofluids. *Protein Engineering, Design and Selection*. 2024;37:gzae003. 10.1093/protein/gzae003

Reproduced with permission from Oxford University Press.

Abstract

We developed fluorescent protein sensors for nicotine with improved sensitivity. For iNicSnFR12 at pH 7.4, the proportionality constant for $\Delta F/F_0$ vs [nicotine] (δ -slope, $2.7 \mu\text{M}^{-1}$) is 6.1-fold higher than the previously reported iNicSnFR3a. The activated state of iNicSnFR12 has a fluorescence quantum yield of at least 0.6. We measured similar dose-response relations for the nicotine-induced absorbance increase and fluorescence increase, suggesting that the absorbance increase leads to the fluorescence increase *via* the previously described nicotine-induced conformational change, the ‘candle snuffer’ mechanism. Molecular dynamics (MD) simulations identified a binding pose for nicotine, previously indeterminate from experimental data. MD simulations also showed that Helix 4 of the periplasmic binding protein (PBP) domain appears tilted in iNicSnFR12 relative to iNicSnFR3a, likely altering allosteric network(s) that link the ligand binding site to the fluorophore. In thermal melt experiments, nicotine stabilized the PBP of the tested

iNicSnFR variants. iNicSnFR12 resolved nicotine in diluted mouse and human serum at 100 nM, the peak [nicotine] that occurs during smoking or vaping, and possibly at the decreasing levels during intervals between sessions. NicSnFR12 was also partially activated by unidentified endogenous ligand(s) in biofluids. Improved iNicSnFR12 variants could become the molecular sensors in continuous nicotine monitors for animal and human biofluids.

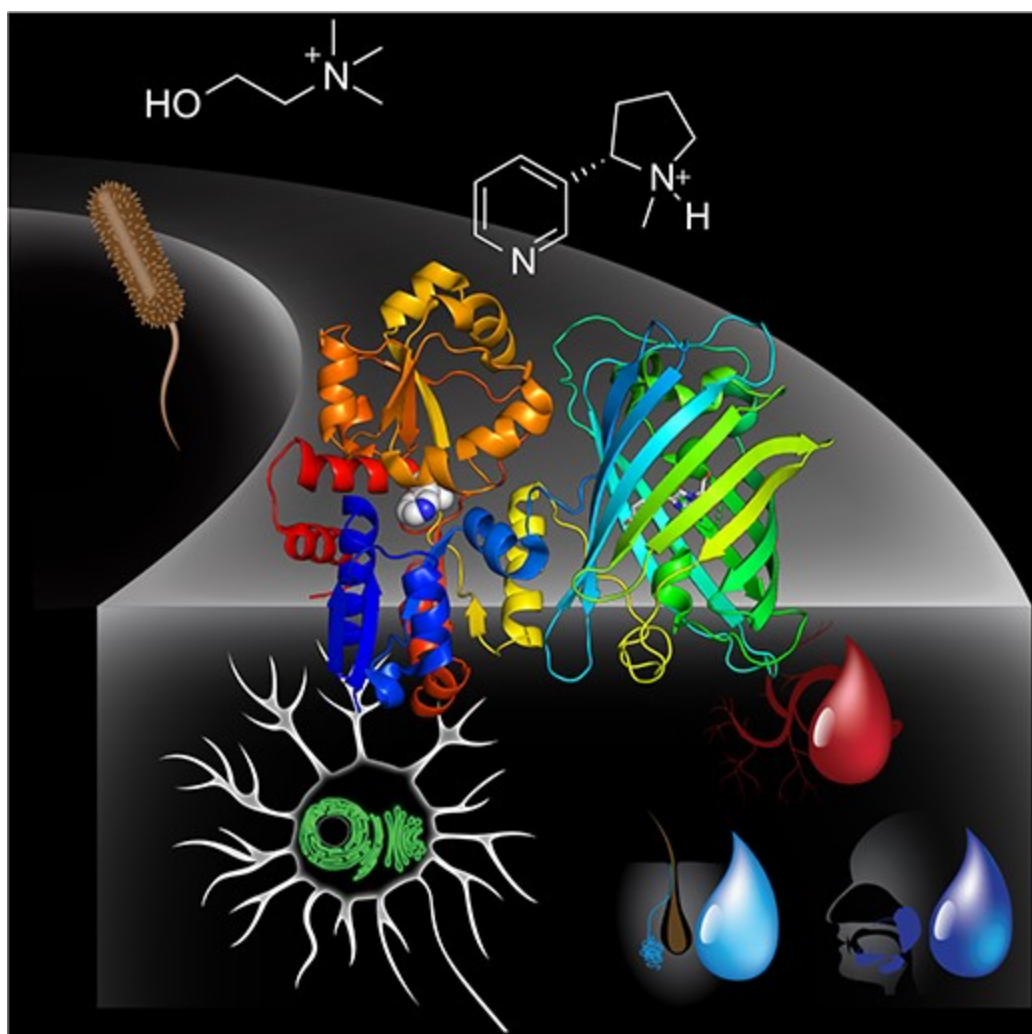


Figure 2.1 Graphical Abstract. iNicotineSnFR12 results from directed evolution of a merger between a bacterial periplasmic choline-binding protein and a circularly permuted GFP. The new biosensor, suitable for experiments in biofluids and within cells, has the sensitivity to measure [nicotine] at blood levels during smoking and vaping.

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 wearable, 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 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 has 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 ($6.5 < \text{pK}_a < 10$), 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 states [6], it is now possible to test structure-based hypotheses to improve the sensitivity.

Here we combined computational protein design with mutagenesis, fluorescence, and absorbance assays. The resulting sensor, iNicSnFR12, is ~6.1-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 a ligand pose consistent with experiments, 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 serve in minimally invasive devices to continuously monitor nicotine in the blood of smokers or vapers.

Results and Discussion

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 interact with the first-shell residues [1, 6]. Because (a) mutations at greater distances from the binding site sometimes affect ligand binding and/or enzymatic activity [8, 9], and (b) mutations computationally designed for enhanced stability often show enhanced function [10], we applied computational design to identify possible mutation hotspots outside the first and second shell.

We first used AutoDock Vina to dock nicotine into iNicSnFR3a [11]. 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 2.S1**), from which we selected 34 positions (**Table 2.S2**) 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].

Table 2.S1: Top 100 mutation outputs from single mutant stability analysis on iNicSnFR3a ranked by energy score.

Ranking	Mutation	Energy Score
1	R467W	-7.11
2	D72Y	-6.77
3	D72F	-6.28
4	N39W	-5.97
5	P475W	-5.95
6	P79W	-5.79
7	R36F	-5.51
8	E24W	-5.28
9	P328W	-5.28
10	R36Y	-5.18
11	P79H	-5.14
12	N20W	-4.82
13	P79M	-4.78
14	N46W	-4.69
15	Q431Y	-4.60
16	K499W	-4.54
17	P79F	-4.47

18	N39F	-4.42
19	R36W	-4.42
20	K342W	-4.41
21	D72H	-4.35
22	E24Y	-4.32
23	P400W	-4.19
24	E24F	-4.13
25	P328Y	-4.13
26	N39Y	-4.13
27	E513W	-4.11
28	E24L	-4.11
29	S325W	-4.03
30	P79L	-3.99
31	P475F	-3.98
32	P507A	-3.96
33	P475Y	-3.96
34	D453W	-3.88
35	N20Y	-3.85
36	K454Y	-3.84
37	R36I	-3.82
38	R52W	-3.75
39	P400Y	-3.75
40	E64L	-3.73
41	P328F	-3.73
42	R467Y	-3.73
43	S325F	-3.72
44	E476W	-3.65
45	Q431F	-3.62
46	N497Y	-3.61
47	N497W	-3.58
48	R467L	-3.55
49	N497F	-3.52
50	T413W	-3.52
51	R36V	-3.48
52	P79Y	-3.48
53	E78W	-3.47
54	S325Y	-3.44
55	R36L	-3.43
56	S9Y	-3.42
57	K336W	-3.41
58	P77W	-3.39
59	K454W	-3.38
60	P323W	-3.37

61	T413Y	-3.35
62	D439M	-3.34
63	E27L	-3.34
64	M418Y	-3.33
65	P79I	-3.32
66	E78Y	-3.31
67	P323F	-3.31
68	K454F	-3.27
69	P475L	-3.27
70	P464I	-3.26
71	L384W	-3.24
72	E24H	-3.24
73	P400F	-3.23
74	E517W	-3.21
75	P475H	-3.21
76	D72W	-3.20
77	Q431W	-3.19
78	H68W	-3.19
79	N46L	-3.19
80	P79T	-3.16
81	P350I	-3.15
82	K342F	-3.15
83	K51W	-3.14
84	D434V	-3.13
85	P464V	-3.13
86	E78F	-3.12
87	K362W	-3.11
88	D501I	-3.11
89	R341W	-3.07
90	P79Q	-3.06
91	D452H	-3.05
92	E429W	-3.04
93	P350W	-3.04
94	R467F	-3.04
95	P507T	-3.02
96	N497V	-3.02
97	P400L	-3.01
98	N20F	-2.97
99	D453H	-2.97
100	K414W	-2.97

Table 2.S2: Mutation outputs from sequence design on iNicSnFR3a on 34 previously identified positions ranked by energy gain.

Ranking	Position	Mutation	Energy Gain
1	K342	F	-4.51
2	R52	K	-3.98
3	Q431	Y	-3.96
4	K51	I	-3.85
5	K454	R	-3.84
6	N20	D	-3.56
7	R467	K	-3.56
8	E476	S	-2.88
9	N46	L	-2.74
10	E27	L	-2.04
11	D452	P	-2.04
12	T413	I	-1.95
13	D453	S	-1.85
14	E64	D	-1.73
15	N497	A	-1.73
16	K414	I	-1.61
17	M418	D	-1.59
18	S325	F	-1.43
19	R36	I	-1.42
20	S9	T	-1.11
21	N39	H	-0.67
22	E517	R	-0.58
23	E78	Y	-0.24
24	E24	-	0.00
25	K499	-	0.00
26	E513	-	0.00
27	K336	-	0.00
28	D439	-	0.00
29	L384	-	0.00
30	D434	-	0.00
31	K362	-	0.00
32	D501	-	0.00
33	R341	-	0.00
34	E429	-	0.00

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 2.S1**). 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 2.2A**. 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) displayed 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.

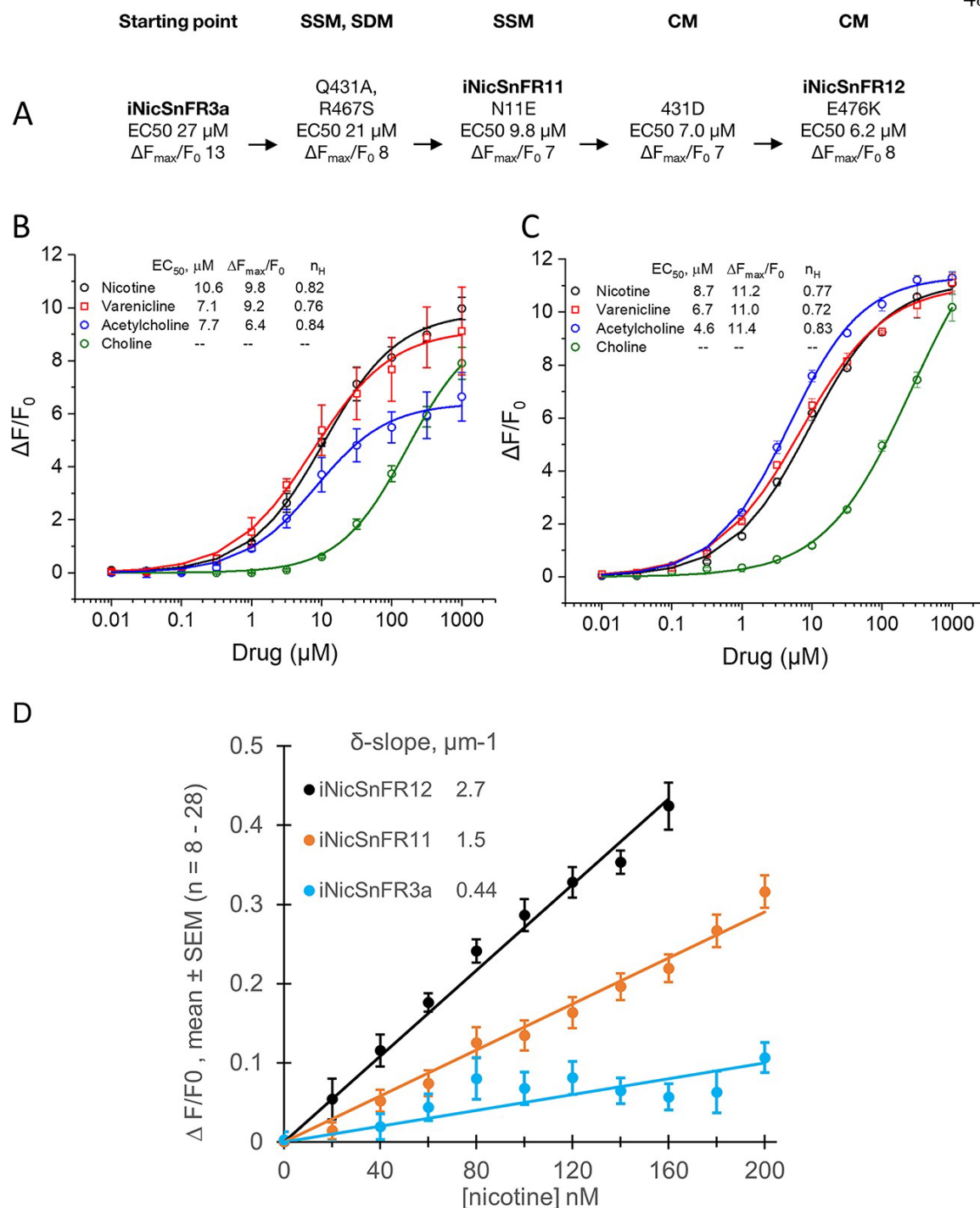


Figure 2.2 Directed Evolution and Fluorescence Characteristics of iNicSnFR11 and iNicSnFR12. (A) The path of protein engineering of iNicSnFR constructs. Data for iNicSnFR3a are the average of two measurements reported previously for the purified

protein [1]. Other measurements reported in this figure were made on bacterial lysates, which often have different $\Delta F_{\max}/F_0$ and EC_{50} values than the purified proteins (compare with panels B and C). SSM, site saturation mutagenesis; SDM, site-directed mutagenesis; CM, combinatorial mutagenesis. **(B)** Dose-response relations for purified iNicSnFR11. Dose-response relations for several nicotinic drugs. Purified iNicSnFR11 was measured at 100 nM in 3× PBS, pH 7.40. Data are mean $\pm$ SEM (n = 3). **(C)** Dose-response relations for purified iNicSnFR12. Dose-response relations for several nicotinic drugs. Purified iNicSnFR12 was measured at 100 nM in 3× PBS, pH 7.40. Data are mean $\pm$ SEM (n = 3). **(D)** Dose-response relations for iNicSnFR variants at sub- μ M [nicotine]. Purified iNicSnFR3a, iNicSnFR11, and iNicSnFR12 were measured at 100 nM in 3× PBS, pH 7.40. The legend shows values for the fitted δ -slopes. Data points are mean $\pm$ SEM (n = 8 to 28).

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 [12]. 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.

Fluorescence Characteristics 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 an EC_{50} of 10.6 μM (**Figure 2.2B**), or a 2.5-fold improvement over iNicSnFR3a [1]. The iNicSnFR11 data show roughly equal EC_{50} values for varenicline, ACh, and nicotine (**Figure 2.2B**), 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, $\Delta F/F_0$ vs [nicotine], was linear in this range. We term the proportionality constant the δ -slope to emphasize that it is measured at low [nicotine] ($<0.1 \times EC_{50}$). We consider the measured δ -slope more relevant than the metric called the S-slope = $(\Delta F_{\text{max}}/F_0) / EC_{50}$. The $\Delta F_{\text{max}}/F_0$ and EC_{50} values, which are fitted to the Hill equation from measurements over the entire dose-response relation, could be influenced by complexities at high [nicotine]. For iNicSnFR11, the δ -slope was 1.5 μM^{-1} (**Figure 2.2D**), representing a 3.4-fold improvement over the value measured for iNicSnFR3a, 0.44 μM^{-1} .

For iNicSnFR12, the EC_{50} for nicotine was 8.7 μM , the lowest value we have measured in the iNicSnFR series (**Figure 2.2C**). Varenicline and ACh had EC_{50} values similar (within two-fold) to the value for nicotine (**Figure 2.2C**); this similarity resembles the pattern for iNicSnFR3a and iNicSnFR11. Again, 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_{50}).

As expected, in the sub- μM range, nicotine displays a linear dose-response relation at iNicSnFR12 (**Figure 2.2D**). The δ -slope for nicotine is $2.7 \mu\text{M}^{-1}$. Therefore, iNicSnFR12 responds to nicotine with sensitivity representing a 6.1-fold improvement over iNicSnFR3a.

Absorbance Characteristics of iNicSnFR12

Figure 2.3A presents optical absorbance data for iNicSnFR12 in the presence of various nominal [nicotine] between zero and the nearly saturating value of $120 \mu\text{M}$ at pH 7.0. Typical fluorescence experiments are conducted at excitation wavelengths between 470 and 500 nm. At the absorption peak (496 nm), the fitted molar absorptivity of the apo and nicotine-bound iNicSnFR12 are $\epsilon_{\text{apo}} = 2900 \text{ M}^{-1} \text{ cm}^{-1}$ and $\epsilon_{\text{bound}} = 48\,800 \text{ M}^{-1} \text{ cm}^{-1}$, respectively (**Figure 2.3B**). Because of uncertainties in concentrations, dilutions, and blanks, we consider that the ratio $\epsilon_{\text{bound}}/\epsilon_{\text{apo}}$ is in the range of 14 to 21. The fitted K_D value is $14.5 \mu\text{M}$ (**Figure 2.3B**). In nicotine-induced fluorescence experiments on the same sample of purified iNicSnFR12 (diluted to 70–200 nM), we measured $\Delta F_{\text{max}}/F_0 = 10.2 \pm 0.8$ and $EC_{50} = 10.6 \pm 1.2 \mu\text{M}$ (mean $\pm$ SEM, $n = 3$). The modestly greater (~ 1.5 -fold) EC_{50} value at pH 7.0 vs 7.4 is consistent with previous observations on increases in EC_{50} with lower pH for other iDrugSnFRs [1, 5] and may be explained by the role of protonation in the ‘candle snuffer’ mechanism [6].

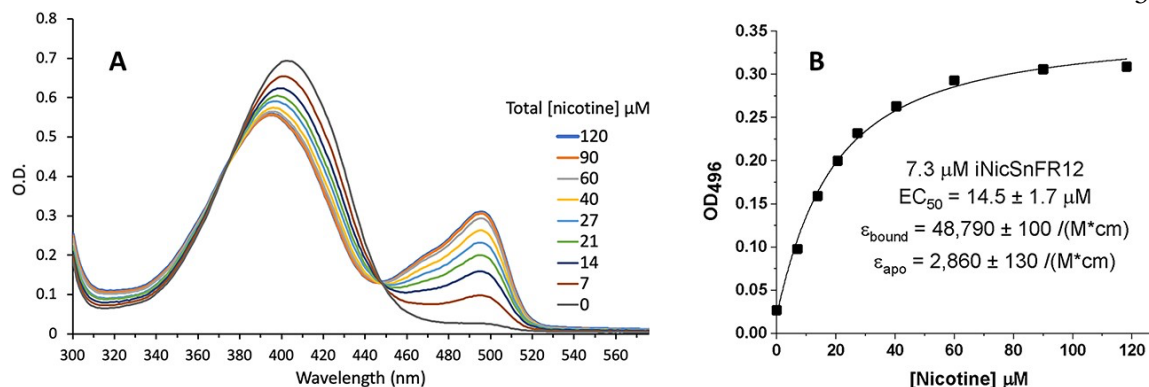


Figure 2.3 iNicSnFR12 shows increased absorption at 450–520 nm upon nicotine binding. (A) Nicotine tartrate solution (1 mM) was added to the iNicSnFR solution (7.3 μM) to produce the indicated nominal [nicotine]. (B) Fitting the OD at 496 nm (the local peak absorption) to a single binding model yielded the parameters indicated.

Fluorescence Quantum Yield of iNicSnFR3a and iNicSnFR12

The Φ for iNicSnFR3a (corrected as described in Methods) is 0.733; for iNicSnFR12, 0.783. We consider that these estimates are reliable within $\pm 15\%$, so that Φ for iNicSnFR3a and iNicSnFR12 are between 0.623 and 0.843 and between 0.666 and 0.901, respectively. Therefore, the Φ of the liganded iNicSnFR variants resembles that of EGFP itself (0.6) and may be slightly higher.

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 2.4A). 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.

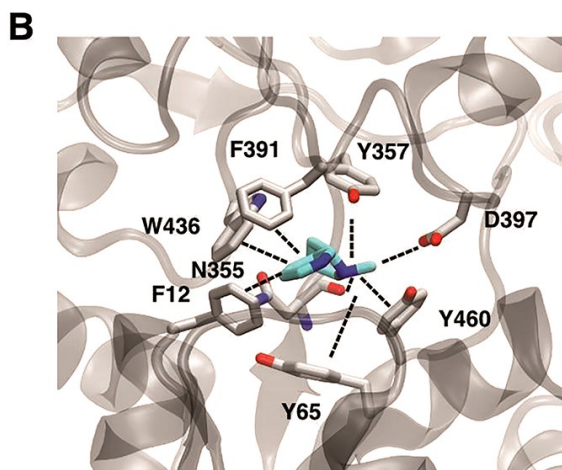
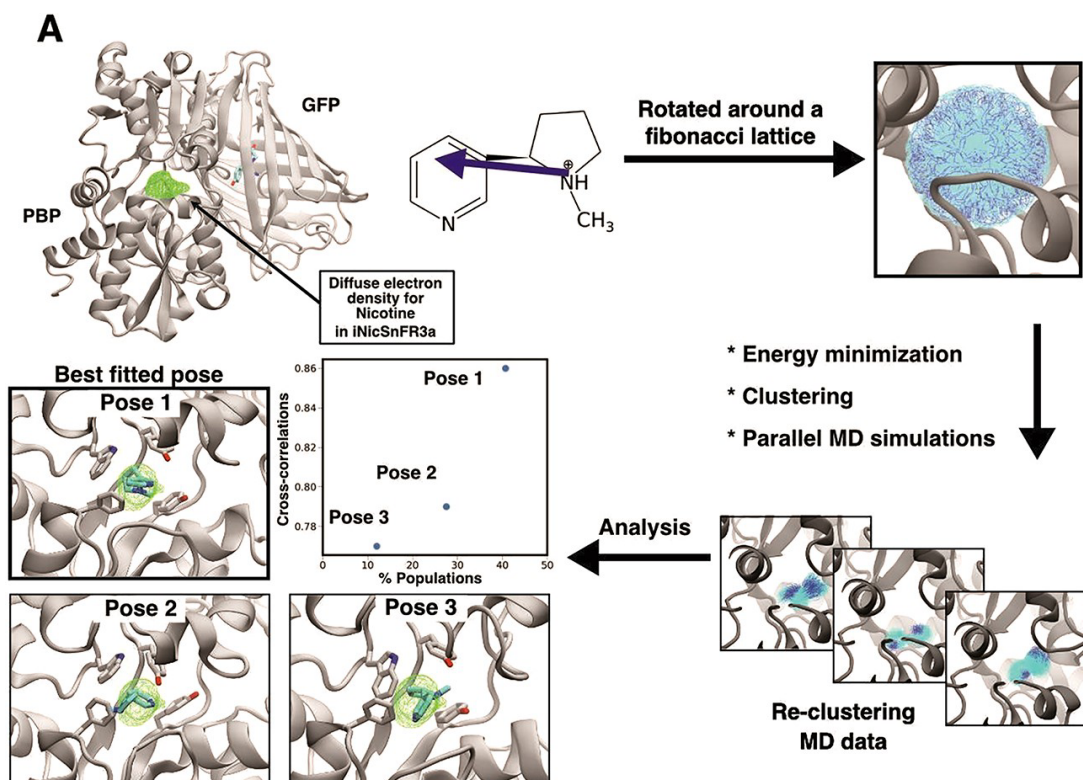


Figure 2.4 An MD simulation-based workflow to assign a nicotine-bound pose in the X-ray map of nicotine-bound iNicSnFR3a (PDB: 7S7U). (A) The PBP and the GFP (chromophore shown as sticks). The diffuse X-ray density of nicotine is shown in green. To search for possible nicotine poses in the binding pocket, 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 cross-correlation with the model. (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.

Next, we analyzed detailed interactions of nicotine within the binding pocket using all the members of cluster 1 (**Figure 2.4B**). 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.

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 μ s, totaling 12 μ s 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-Kobayashi *et al.* [13, 14] and connects the PBP to GFP. Interestingly, in NicSnFR12, helix 4 was shifted ~ 4 Å away from the centroid of the PBP (**Figure 2.5A**),

compared to NicSnFR3a. One measure of this shift is the probability distribution of the distance between the C α atom of a terminal residue D329 and the C α atom Y461, the terminal of a β -sheet that locates nearest to helix 4, residue Y461 (**Figure 2.5A**). 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.

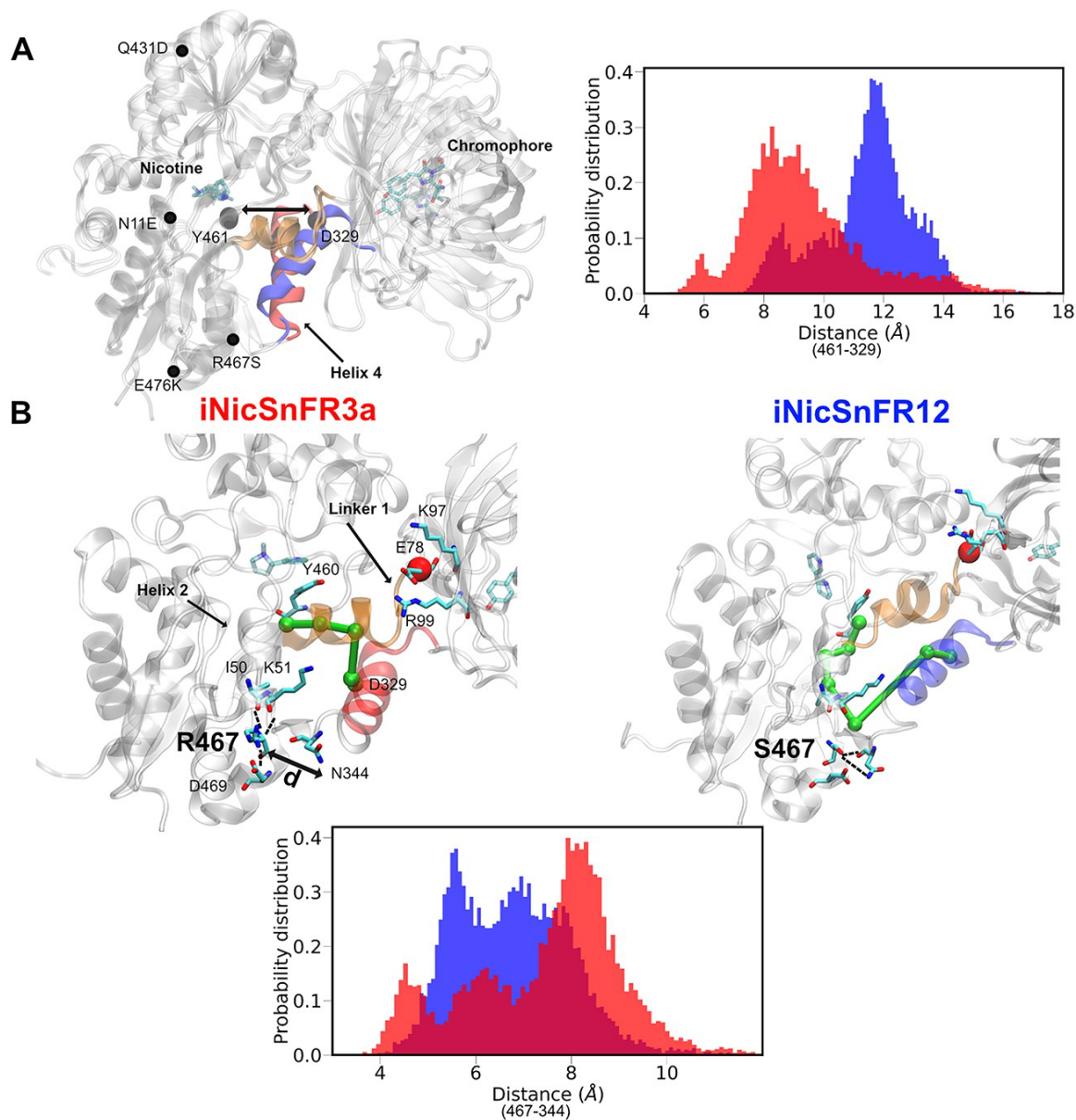


Figure 2.5 Tilting of linker helix 4 in iNicSnFR12 might modulate ligand binding. (A)

All the mutations present in iNicSnFR12 from iNicSnFR3a are labeled. In iNicSnFR12, helix 4 (colored in blue) shifts toward the GFP. To measure this process, we analyzed the probability distribution of the distance between the C α atom of a terminal residue D329 and the C α 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.

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 2.5B**). 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, which affects the thermal stability of the variant as described below. 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 2.5B**). 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 toward the centroid of the PBP (**Figure 2.S1**); this action could help to force the other end of the helix to tilt away from the PBP as a balancing effect.

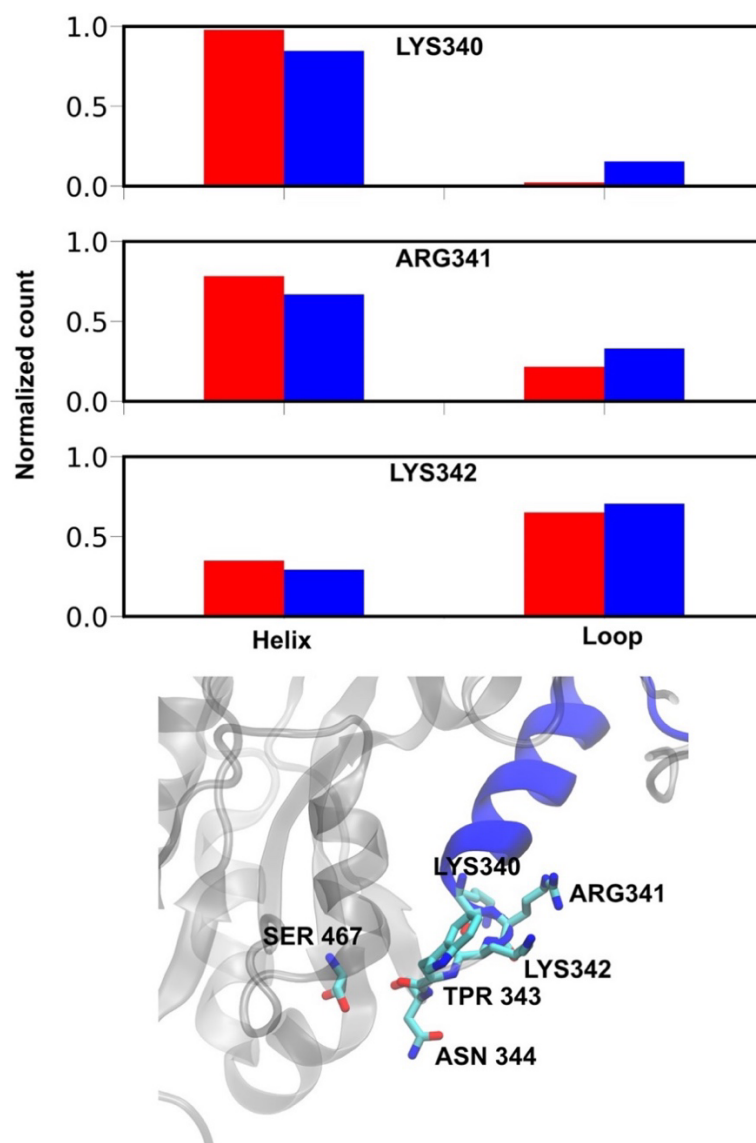


Figure 2.S1 Partial unwinding of Helix 4 in iNicSnFR12. Due to the interactions of SER467 with ASN344, located at the C-terminus of helix 4, this part of helix 4 spends a slightly lower fraction of the simulation in well-defined helical secondary structure in iNicSnFR12 (blue bars) in comparison to iNicSnFR3a (red bars), indicative of stretching/partial unwinding of the helix.

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 [15]. 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) [15-17]. The source was selected as Y460, a key aromatic residue coordinating the cationic amine of nicotine as described above (**Figure 2.4B**). 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.

Thermal Stability Characterization of iNicSnFR3a, iNicSnFR12, and iNicSnFR12-S467R

To explore the stability implications of the R467S mutation in iNicSnFR12 that putatively deletes the helix capping interaction between the sidechain of R467 and the C-terminal residues of helix 2, differential scanning fluorimetry (DSF) was performed on samples of iNicSnFR3a, iNicSnFR12, and iNicSnFR12-S467R at pH 7.0 in the absence of ligand (**Figure 2.6A**). iNicSnFR12-S467R differs from iNicSnFR12 by a single S467R point mutation, which reintroduces the putative helix capping interaction. After subtraction of intrinsic fluorescence, the three variants exhibited two distinct melting transitions (**Figures 2.S2 and 2.S3**). The unfolding of the PBP domain occurred at the lower apparent melting temperature, $T_m(\text{low})$, while the unfolding of the GFP moiety occurred at the higher apparent

melting temperature, $T_m(\text{high})$ (**Figures 2.S2 and 2.S3**). While the $T_m(\text{high})$ values for iNicSnFR3a, iNicSnFR12, and iNicSnFR12-S467R were indistinguishable, the $T_m(\text{low})$ value associated with PBP unfolding was lower for iNicSnFR12 compared to iNicSnFR3a and iNicSnFR12-S467R (**Figure 2.6A**). This observation supports the hypothesis that R467 plays an important role in stability of the PBP. To determine the effect of nicotine on the stability of these variants, DSF was performed on each variant with a concentration of nicotine equal to twice the respective EC_{50} values (**Figure 2.6B**). In nicotine-induced fluorescence experiments on the sample of purified iNicSnFR12-S467R, we measured an EC_{50} of $25.6 \pm 1.8 \mu\text{M}$ (mean $\pm$ SEM, $n = 3$). The presence of nicotine at twice the EC_{50} concentrations increased values for $T_m(\text{low})$ ranging from 1.7°C to 3.3°C and decreased values for $T_m(\text{high})$ by less than 1°C . Nicotine, therefore, has an overall stabilizing effect on the PBP domains of the tested iNicSnFR variants.

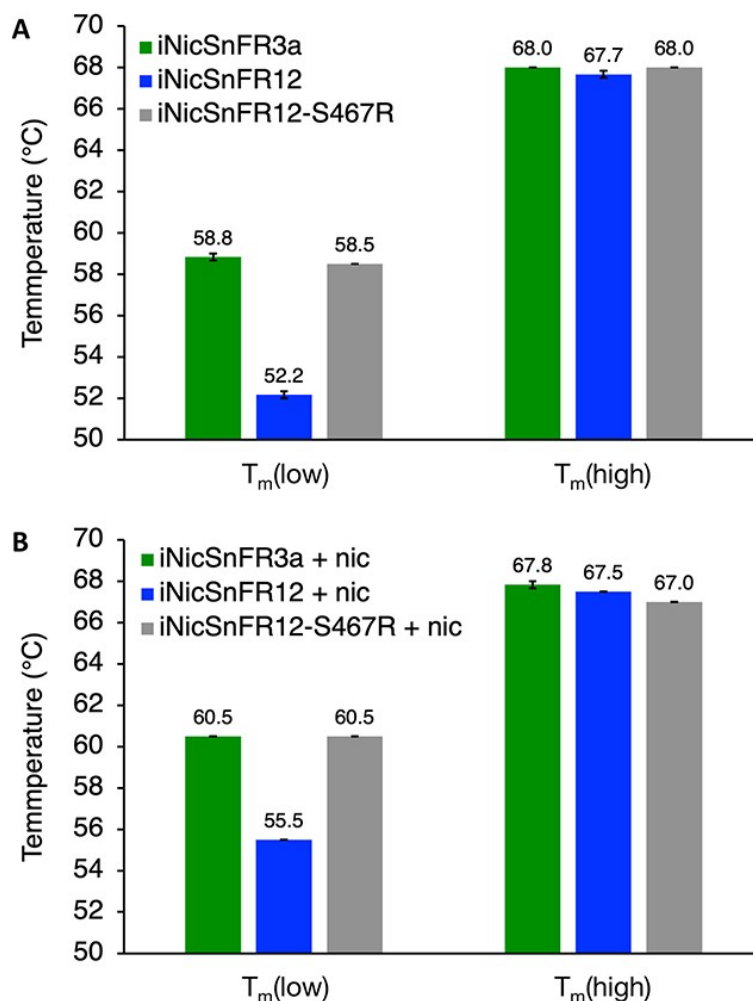


Figure 2.6 S467R mutation to iNicSnFR12 restores enhanced stability inherent to iNicSnFR3a. (A) DSF for iNicSnFR3a, iNicSnFR12, and iNicSnFR12-S467R in $3 \times$ PBS, pH 7.0 in the absence of ligand. $T_m(\text{low})$ values were determined to be 58.8 ± 0.2 °C, 52.2 ± 0.2 °C, and 58.5 ± 0.0 °C for iNicSnFR3a, iNicSnFR12, and iNicSnFR12-S467R respectively. $T_m(\text{high})$ values were determined to be 68.0 ± 0.0 °C, 67.7 ± 0.2 °C, and 68.0 ± 0.0 °C for iNicSnFR3a, iNicSnFR12, and iNicSnFR12-S467R respectively. (B) DSF for iNicSnFR3a, iNicSnFR12, and iNicSnFR12-S467R at pH 7.0 with $[\text{nicotine}] = 2 \times EC_{50}$. $T_m(\text{low})$ values were determined to be 60.5 ± 0.0 °C, 55.5 ± 0.0 °C, and 60.5 ± 0.0 °C for iNicSnFR3a, iNicSnFR12, and iNicSnFR12-S467R respectively. $T_m(\text{high})$ values were determined to be 67.8 ± 0.2 °C, 67.5 ± 0.0 °C, and 67.0 ± 0.0 °C for iNicSnFR3a, iNicSnFR12, and iNicSnFR12-S467R respectively. Data are mean $\pm$ SEM, $n = 3$.

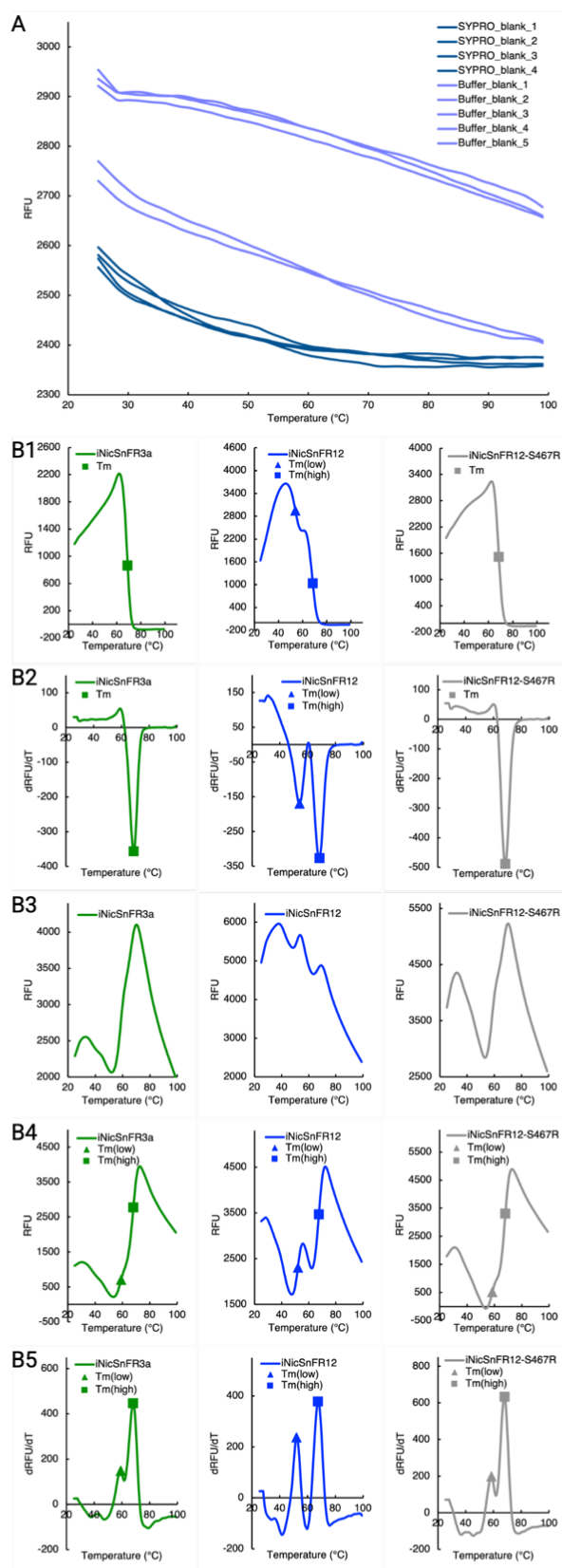


Figure 2.S2 Thermal melt data in the absence of nicotine. (A) Fluorescence of blank samples. Samples with the “SYPRO” and “Buffer” prefixes are with or without SYPRO® Orange, respectively. (B1) Intrinsic fluorescence. Fluorescence values are averages of technical triplicates. Values for T_m , $T_m(\text{low})$, and $T_m(\text{high})$ are associated with decreases in intrinsic fluorescence due to GFP unfolding and were determined from the local minima of $dRFU/dT$ curves for each replicate, the averages of which are presented in **Figure 2.S2B2**. T_m for iNicSnFR3a was 69.0 ± 0.0 °C. $T_m(\text{low})$ and $T_m(\text{high})$ for iNicSnFR12 were 54.0 ± 0.0 °C and 68.3 ± 0.2 °C, respectively. T_m for iNicSnFR12-S467R was 68.5 ± 0.0 °C. Data are mean $\pm$ SEM, $n = 3$. (B2) GFP apparent melting temperature determination. Values of $dRFU/DT$ are averages of numerical derivatives taken on technical replicates. (B3) Background-subtracted DSF. (B4) DSF minus intrinsic fluorescence. Average intrinsic fluorescence was subtracted from the fluorescence of triplicate DSF samples. Averages are shown. Values for $T_m(\text{low})$, and $T_m(\text{high})$ were determined from the peaks of $dRFU/dT$ curves for each replicate, the averages of which are presented in **Figure 2.S2B5**. Values for $T_m(\text{low})$ and $T_m(\text{high})$ are summarized in **Figure 2.6A**. (B5) Apparent melting temperature determination. Values of $dRFU/DT$ presented are averages of numerical derivatives taken on technical triplicates.

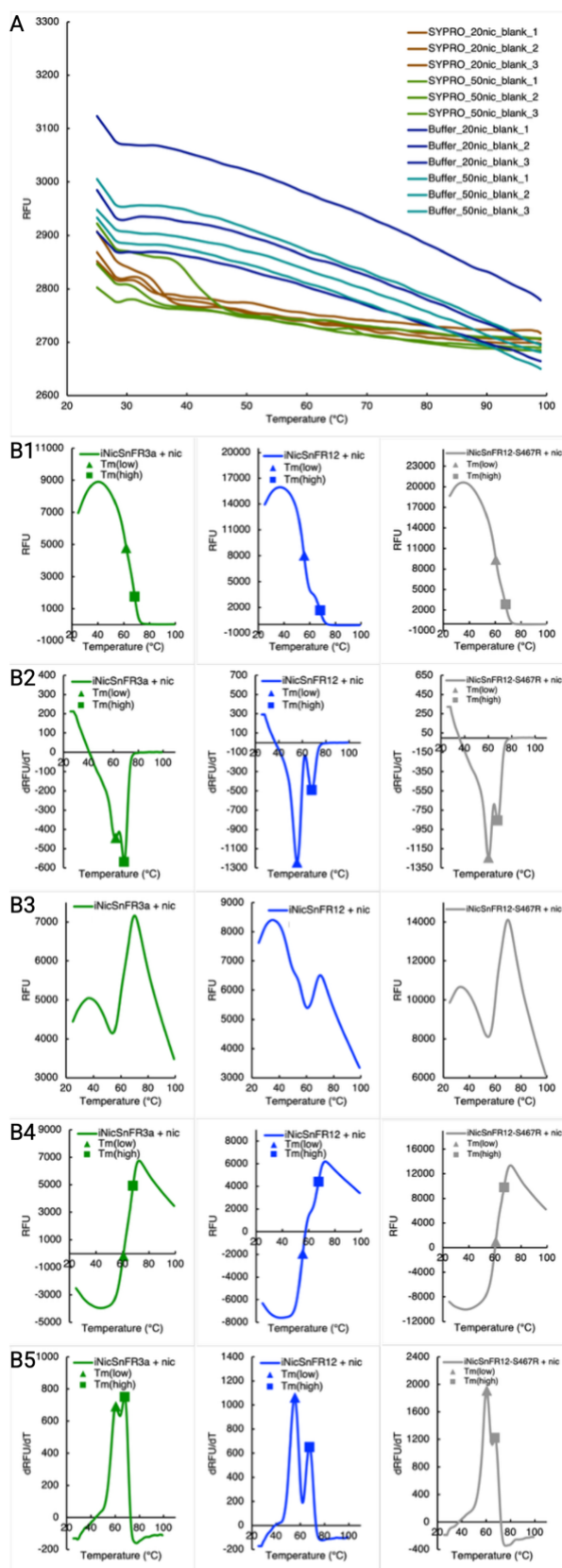


Figure 2.S3 Thermal melt data in the presence of nicotine. (A) Fluorescence of blank samples with nicotine. Samples with the “SYPRO” and “Buffer” prefixes are with or without SYPRO® Orange, respectively. Concentrations of nicotine are in μM (e.g., “20nic” refers to 20 μM nicotine). (B1) Intrinsic fluorescence with nicotine. Fluorescence values are averages of technical triplicates. Values for T_m , $T_m(\text{low})$, and $T_m(\text{high})$ are associated with decreases in intrinsic fluorescence due to GFP unfolding and were determined from the local minima of dRFU/dT curves for each replicate, the averages of which are presented in **Figure 2.S3B2**. $T_m(\text{low})$ and $T_m(\text{high})$ for iNicSnFR3a were 61.8 ± 0.2 °C and 68.5 ± 0.0 °C, respectively. $T_m(\text{low})$ and $T_m(\text{high})$ for iNicSnFR12 were 55.5 ± 0.0 °C and 68.0 ± 0.0 °C, respectively. $T_m(\text{low})$ and $T_m(\text{high})$ for iNicSnFR12-S467R were 60.3 ± 0.2 °C and 68.0 ± 0.3 °C, respectively. Data are mean $\pm$ SEM, $n = 3$. (B2) GFP apparent melting temperature determination with nicotine. Values of dRFU/DT are averages of numerical derivatives taken on technical replicates. (B3) Background-subtracted DSF with nicotine. (B4) DSF minus intrinsic fluorescence, each with nicotine. Average intrinsic fluorescence was subtracted from the fluorescence of triplicate DSF samples. Averages are shown. Values for $T_m(\text{low})$, and $T_m(\text{high})$ were determined from the peaks of dRFU/dT curves for each replicate, the averages of which are presented in **Figure 2.S3B5**. Values for $T_m(\text{low})$ and $T_m(\text{high})$ are summarized in **Figure 2.6B**. (B5) Apparent melting temperature determination with nicotine. Values of dRFU/DT presented are averages of numerical derivatives taken on technical triplicates.

iNicSnFR12 Detects Nicotine at 100 nM Concentrations in Rat and Human Serum

Figure 2.2C shows that solutions containing iNicSnFR12 detect nicotine at <100 nM. We tested whether this capability also applies to solutions containing rat or human biofluids. Firstly, in PBS alone, fluorescence increased as the percentage of rat serum increased, possibly due to the presence of bilirubin. Secondly, we define F_0' as the fluorescence in PBS plus rat serum plus iNicSnFR12. F_0' was greater than the arithmetic sum of F_0 , plus the fluorescence of 10% rat serum alone, by a factor 3.52 ± 0.1 . This suggested that rat serum contains endogenous molecule(s) that activate iNicSnFR12 in the absence of nicotine. With

heat-treated serum, the ratio was only slightly less (3.22-fold). This suggested that the elevated F_0' arises from a small molecule(s) in serum. We infer that the endogenous non-protein ligand(s) in rat serum activate iNicSnFR12 to 31% (3.52/11.2; see **Figure 2.3C**) of the maximum for nicotine.

Thirdly, despite the elevated F_0' , in PBS containing 10% rat serum, iNicSnFR12 displayed fluorescence responses at nicotine concentrations as low as 50 nM nicotine (**Figure 2.7A**). The small variability and linearity suggest that the limit of detection is somewhat lower, perhaps 20 nM. Under these conditions, the δ -slope had a decreased value, $0.37 \pm 0.02 \mu\text{M}^{-1}$; we term this parameter the δ' -slope, and it is several fold lower than the zero-serum δ -slope value of $2.7 \mu\text{M}^{-1}$ (**Figure 2.2D**).

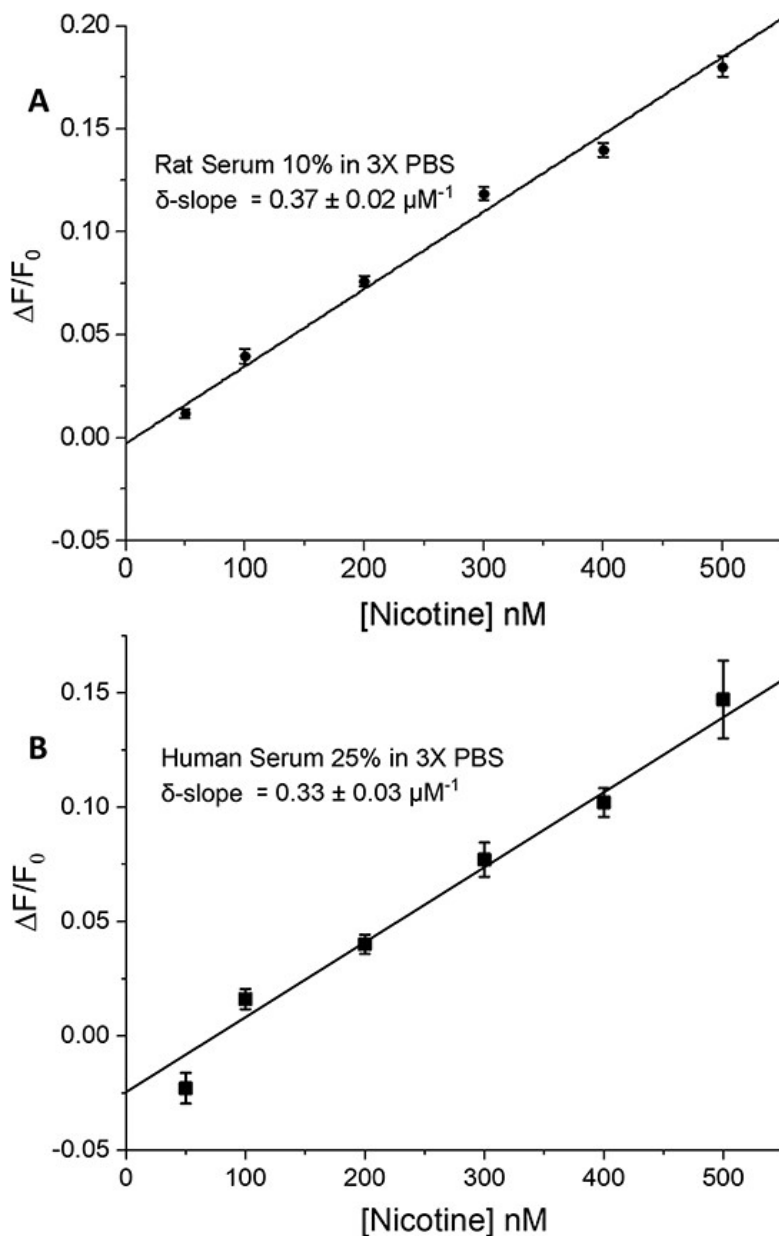


Figure 2.7 Dose-response data for nicotine at iNicSnFR12 in biofluids. (A) Rat serum, 10% in 3× PBS. iNicSnFR concentration was 100 nM. Data were measured 5–30 min after addition of nicotine. The fitted line corresponds to δ' -slope = $0.52 \pm 0.01 \mu\text{M}^{-1}$. Data are mean $\pm$ SEM, N = 9 samples per concentration. (B) Human serum, 25% in 3 × PBS. iNicSnFR concentration was 100 nM. Data were measured 15–30 min after addition of nicotine. The fitted line corresponds to δ' -slope = $0.33. \pm 0.03 \mu\text{M}^{-1}$. Data are mean $\pm$ SEM, N = 9 samples per concentration.

We simulated the fluorescence responses in the presence of an endogenous ligand. The relationship between δ -slope and decreased δ' -slope was consistent with simulations of partial activation by endogenous ligand(s) (**Figure 2.S4**).

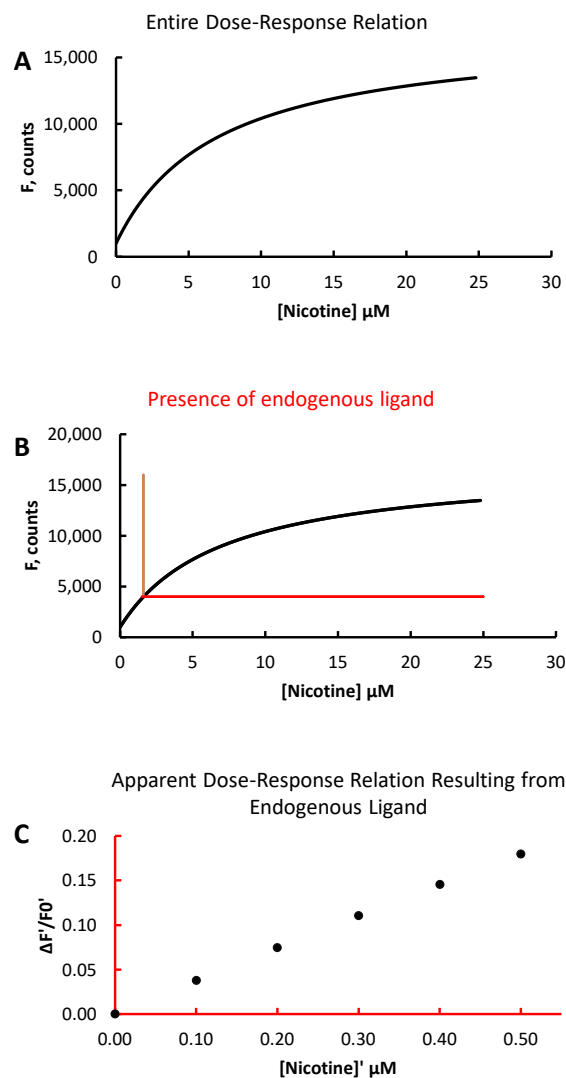


Figure 2.S4 Simulations of decreased S-slope caused by partial activation. We assume a Michaelis-Menton relation between [nicotine] and ΔF , with a Hill coefficient of 1. Other assumptions: F_0 : 1000 counts; EC_{50} : 7, μM ; ΔF_{max} : 16000 counts. Concentration step: 0.1 μM . (A) Unmodified dose-response relation on linear axes. Conditions as in the experiment

of **Figure 2.7A**. The relation intersects the Y-axis at F_0 . **(B)** as in **(A)**, overlaid with red axes denoting $F'_0 = 4F_0$, corresponding to 1.6 μM nicotine (or an unidentified endogenous ligand). **(C)** Simulation of the experiment of **Figure 2.7A**. The plot shows the dose-response relation near the origin of the red overlay. The X-axis, $[\text{nicotine}]'$, represents incremental amounts of [nicotine]. The Y-axis, $\Delta F'/F'_0$, represents incremental fluorescence.

PBS containing 25% human serum from nonsmokers had an intrinsic fluorescence 2.0 ± 0.1 -fold larger than PBS alone. In 25% human serum plus iNicSnFR12, F'_0 was $1.9 \pm$ fold higher than the arithmetic sum of F_0 in PBS and the fluorescence of 25% human serum. This led us to define a new parameter, F'_0 , suggesting that human serum, like rat serum, has endogenous ligand(s) that activate iNicSnFR12. The calculated activation is 17% ($1.9/11$, see **Figure 2.3C**) of the maximum for nicotine.

We tested responses to nicotine at iNicSnFR12 in PBS containing 25% serum (**Figure 2.7B**). The data had greater variance than in 10% rat serum, possibly because of the elevated fluorescence of serum alone. The measurements at 50 nM nicotine had a negative $\Delta F/F_0$, distorting the extrapolated zero-concentration value. Therefore, in 25% human serum, the elevated F'_0 and resulting variability lead to a limit of detection limit of ~ 100 nM. The measured δ' -slope was $0.33 \mu\text{M}^{-1}$, decreased from the zero-serum value of $2.7 \mu\text{M}^{-1}$ (**Figure 2.2D**).

These preliminary data show that iNicSnFR12 responds to nicotine in two biofluids, with a reduced sensitivity (termed S' -slope), as expected from the elevated F_0 . See Discussion for the possible nature of the endogenous ligand(s).

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 [18, 19]. 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 2.5**).

R467S is one of the mutations that distinguish iNicSnFR12 from previous iNicSnFR variants (**Figure 2.2A**). Comparative thermal stability data for Arg and Ser at position 467 support the idea that the side chain at position 467 affects the stability of Apo-iNicSnFR12 *via* the presence (Arg) or absence (Ser) of a helix capping interaction (**Figure 2.6**). Also, at least part of the increased nicotine sensitivity of iNicSnFR12 correlates with the increased thermal stability conferred by nicotine binding to the Ser-467 variant (**Figure 2.6**).

Our measurements represent the first data comparing the absorption characteristics of an iDrugSnFR between the apo and drug-bound state (**Figure 2.3**). The ϵ_{apo} and ϵ_{bound} values resemble those for various GCaMP constructs, which also have cpGFP moieties [20, 21]. For iNicSnFR12, the absorption ratio $\epsilon_{\text{bound}}/\epsilon_{\text{apo}}$ is in the range of 14 to 21, whereas the fluorescence data yield an F_{max}/F_0 ratio of 10.2. It should be noted that the absorption data were taken at iNicSnFR12 concentrations 37- to 90-fold higher than the fluorescence experiments, so that the ϵ_{apo} and F_0 values may be separately distorted by systematic errors that affect these ratios. The good agreement between the EC_{50} from fluorescence measurements and the K_D from absorption experiments supports the idea that the same

binding event and conformational change (movement of the candle snuffer [6] governs both the fluorescence measurements and absorption changes, by perturbing the hydrogen bonds around the chromophore [21, 22]. In the 400–450 nm range, iNicSnFR12 shows only minor ligand-induced changes in absorption, like other cpGFP-based sensors [21].

We also present the first measurements of fluorescence quantum yield, Φ , on an iDrugSnFR. We find that Φ for nicotine-bound iNicSnFR3a and iNicSnFR12 is at least as great as that for EGFP itself and exceeds 0.6, implying that further attempts to improve the quantum yield would yield <1.7-fold enhancements in signal. The rather low ϵ_{apo} values for the iNicSnFR family vitiate quantum yield measurements for the unbound state. For the present data, we cannot determine whether the absorption and fluorescence of the unbound state arise (a) completely from a small fraction of the protein molecules in the active conformation, *versus* (b) a small amount of absorption and fluorescence by the protein in the ‘snuffed’ conformation. We do know that ΔF_0 becomes larger at higher pH, presumably because deprotonation of Glu78 drives the apoprotein toward the active conformation [6], suggesting that mechanism (a) dominates.

Contemplated Uses for iNicSnFR12

Nicotine sensors are selected/selected to bind nicotine; we have not selected against sensitivity to ACh or varenicline. The constructs do retain sensitivity to varenicline and acetylcholine. Neither of these molecules are expected to interfere with contemplated continuous measurements in human biofluids such as sweat and interstitial fluid.

Selection against choline sensitivity has been a secondary goal. The concentration of choline in healthy adult human and rat plasma is 10 μM [23–27]. The EC_{50} for choline at iNicSnFR12 is at least 100 μM (**Figure 2.2C**); therefore, [choline] in 10% rat serum and 25% human serum is 1 μM and 2.5 μM , respectively. These [choline] are expected to produce at most 1% and 2.5% activation at iNicSnFR12, respectively. The measured activation (F_0'/F_0) produced by the serum samples is > an order of magnitude higher (31% and 17% for 10% rat and 25% human serum, respectively).

These data suggest that another endogenous compound(s) activates iNicSnFR12 in rat or human serum. These compounds could be biogenic amines other than choline. In humans, the elevated F_0 could also arise from the presence of varenicline. It remains to be tested whether the human biofluids of most interest, sweat and interstitial fluid, have similar concentrations of endogenous activated ligand(s). If so, identifying the endogenous ligand(s), and evolving iNicSnFR variants with decreased sensitivity to these ligands, remains a goal of further improvements in the iNicSnFR series.

A related goal is to selectively improve the sensitivity to nicotine so that ~ 10 nM gives a larger signal. This would allow an iNicSnFR to become the sensor protein in minimally invasive or non-invasive devices that continuously monitor nicotine concentration in biofluids. The encouraging data with rat serum (**Figure 2.7A**) suggest that it would be worthwhile to measure [nicotine] continuously in rats performing intravenous nicotine self-administration [28].

In human pharmacokinetic data, smokers and vapers have as much as ten-fold variability in their nicotine half-life, in part because of polymorphisms in cytochrome P450 2A6 [29]. This variability also may inform the tactic that an individual uses to quit smoking, *e.g.* nicotine patch or varenicline [30]. A continuous nicotine monitor with sensitivity down to 10 nM would enable one to test the personal pharmacokinetics of nicotine during smoking or vaping sessions, as well as during the decay phase that continues for several hours after a session.

Methods

Mutagenesis and Library Screening

Site-saturated mutagenesis (SSM) libraries were generated employing the so-called ‘22c trick’ [31] for PCR primers. For site-directed mutagenesis (SDM), individual primers 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 [32] supplemented with 100 µg/ml ampicillin for 30 h at 30 °C, 250 rpm. Then the cells were harvested ($3000 \times g$, 10 min, 4 °C), washed with PBS, pH 7.40, and resuspended in $3 \times$ PBS, pH 7.40. Resuspended cells were lysed by freezing and thawing using liquid N₂ and a room-temperature water bath. Intact cells and cell debris were removed by centrifugation at $3500 \times g$ 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 near the EC_{50} of the parent construct for the library being screened. A Spark[®] 20M 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 relations against nicotine (described below). The optimally responding construct in each round of mutagenesis was used as a template for the next round of SSM, SDM, or combinatorial mutagenesis (CM).

Purification of iNicSnFRs

Proteins were expressed in *E. coli* BL21-Gold (DE3) cells using 50 ml ZYM5052 autoinduction media [32]. 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 $15\,000 \times g$ for 30 min at 4 °C. The supernatant was collected and loaded onto a prewashed Ni-NTA column with wash buffer. 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 $3 \times$ PBS, pH 7.40. The dialyzed protein was then subjected 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). Solutions were in $3 \times$ PBS, pH 7.0 or 7.4. A Tecan Spark® 20M 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 Origin 9.2 software (OriginLabs) to the Hill equation [1, 6].

Absorbance Measurements

Absorption spectra were measured in a Tecan Spark® 20M instrument. Chromophore concentration was measured by denaturation in 1 M NaOH using the absorption coefficient of $44\,000\text{ cm}^{-1}\text{ M}^{-1}$ at 447 nm [20]. To a solution of iNicSnFR12 ($7.3\text{ }\mu\text{M}$ in $3 \times$ PBS, pH 7.0), we added increments of the same solution with added 1 mM nicotine tartrate (Thermo Fisher, Waltham, MA). The relationship between absorption at 496 nm and [nicotine] was fitted to the molar absorptivity of the apo and bound sensors (ϵ_{apo} and ϵ_{bound} , respectively) and the K_m for the nicotine-iNicSnFR12 complex (K) assuming a single binding site, corrected for depletion of the free [nicotine] by binding to iNicSnFR12:

$$A_{496} = \epsilon_{\text{apo}}S + (\epsilon_{\text{bound}} - \epsilon_{\text{apo}}) \left(K + S + x - \sqrt{(S + x + K)^2 - 4Sx} \right) / 2$$

where S is the concentration of iNicSnFR12 and x is the total nicotine concentration.

Quantum Yield Measurements

Quantum yields (Φ) were measured by comparison with the value for EGFP. As above, chromophore concentration was measured by denaturation in 0.1 M NaOH using the absorption coefficient of $44\,000\text{ cm}^{-1}\text{ M}^{-1}$ at 447 nm [20]. The uncorrected relative fluorescence quantum yield for the sample is given by: [33]

$$\Phi_{f,x} = \Phi_{f,st} \cdot \frac{F_x}{F_{st}} \cdot \frac{A_{st}}{A_x}.$$

$\Phi_{\text{f, st}}$ is the fluorescence quantum yield of the reference standard, which was purified EGFP ($\Phi_{\text{f, st}} = 0.60$ at $\lambda_{\text{ex}} = 489$ nm) [34]. F is the integrated fluorescence emission for 489 nm excitation. A is the absorbance at 489 nm. The iNicSnFR variants were measured in $3 \times$ PBS, pH 7.4 with $[\text{nicotine}] = 10 \times EC_{50}$ for each variant. EGFP was measured in $3 \times$ PBS, pH 7.4. Background-subtracted absorption spectra for samples at protein concentrations between 3–4 μM were measured in a 10-mm pathlength cuvette in the Tecan Spark[®] 20M instrument. Samples were then diluted 20-fold in $3 \times$ PBS, pH 7.4 with the same $[\text{nicotine}]$ in a clear-bottom 96-well plate, and fluorescence measurements were taken by exciting at 489 nm (20 nm nominal bandwidth) with the monochromator and detecting emissions (1 nm step size; 20 nm bandwidth). Fluorescence assay volumes were 337 μL . Background-subtracted emission was integrated between 528 (to avoid overlap with the excitation) and 650 nm.

The Φ values for nicotine-saturated iNicSnFR fluorescence were corrected by three factors. Factor 1 accounts for the extrapolation of iNicSnFR values taken at $10 \times EC_{50}$ to predicted values at saturated $[\text{nicotine}]$; this factor equals 1.1. Factor 2 accounts for relative integrated absorbed photons by the iNicSnFR3a, iNicSnFR11, or iNicSnFR12 vs GFP at the peak excitation wavelength (489 nm) and effective excitation bandwidth of 26 nm, equal to 0.957. Factor 3 accounts for the shift in peak emission of iNicSnFR3a, iNicSnFR11, or iNicSnFR12 vs that of EGFP, resulting in a different fraction of total fluorescence when we integrated from 528 to 700 nm. The emission peak for iNicSnR3a is shifted by 5 nm vs that for EGFP [1], and we noted that the emission decreases at a rate of e-fold per 31 nm. Factor 3 is therefore $e^{-5/31} = 0.851$. We corrected the observed quantum yield by the product of these factors, 0.895.

Computational Search for Docking Poses of Nicotine

A workflow resembling a previous study [35] 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. Both ligands have a protonated amine group. Varenicline is known to form multiple cation- π interactions within the binding

site of iNicSnFR3a [1, 6]; and nicotine is known to form at least one these cation- π interactions, with Tyr-357 [1]. Assuming that other cation- π interactions known for varenicline also occur for nicotine, we first aligned the protonated amine of nicotine to the secondary amine of varenicline. To optimize possible resolution and placement, we placed the pyrrolidine amine using the coordinates of the highest-resolution liganded structure in the iNicSnFR series, the complex between varenicline and the PBP from iCytSnFR (PDB: 7S7X). We defined a vector, from the pyrrolidine amine in nicotine to the C3 carbon atom of the pyridine ring. We mapped this vector onto a Fibonacci spherical lattice (FSL) [36] of 25 points, generating multiple orientations. During rotation, the amine 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 around the vector, with an interval of $\Psi = 45^\circ$, creating a total of 200 poses. To account for the internal degrees of freedom, multiple ligand conformations were generated using the Open Babel toolkit [37]. 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 [35] and removed, resulting in a total of ~600 distinct poses.

Each pose was then subjected to 1000 steps of energy minimization using the generalized Born implicit solvent (GBIS) module implemented in NAMD2 [38, 39]. 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 liganded iNicSnFR3a (PDB 7S7U) 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 $10.9 \times 10.9 \times 10.9$ nm. Each system was energy-minimized using the steepest-descent algorithm for 5000 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 2.S3**). Then, unrestrained simulations were performed for all 10 simulations, each for 200 ns, gathering a total of 2 μ s of simulations.

Table 2.S3: Protocol for gradually releasing the restraints during MD simulations of the 10 nicotine-iDrugSnFR3a complexes, during the search for nicotine docking poses.

Step	Backbone restraint force-constant (kJ mol ⁻¹ nm ⁻²)	Side-chain restraint force-constant (kJ mol ⁻¹ nm ⁻²)	Nicotine heavy atom restraints (kJ mol ⁻¹ nm ⁻²)	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

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 2.S5**) 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* [40] 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.

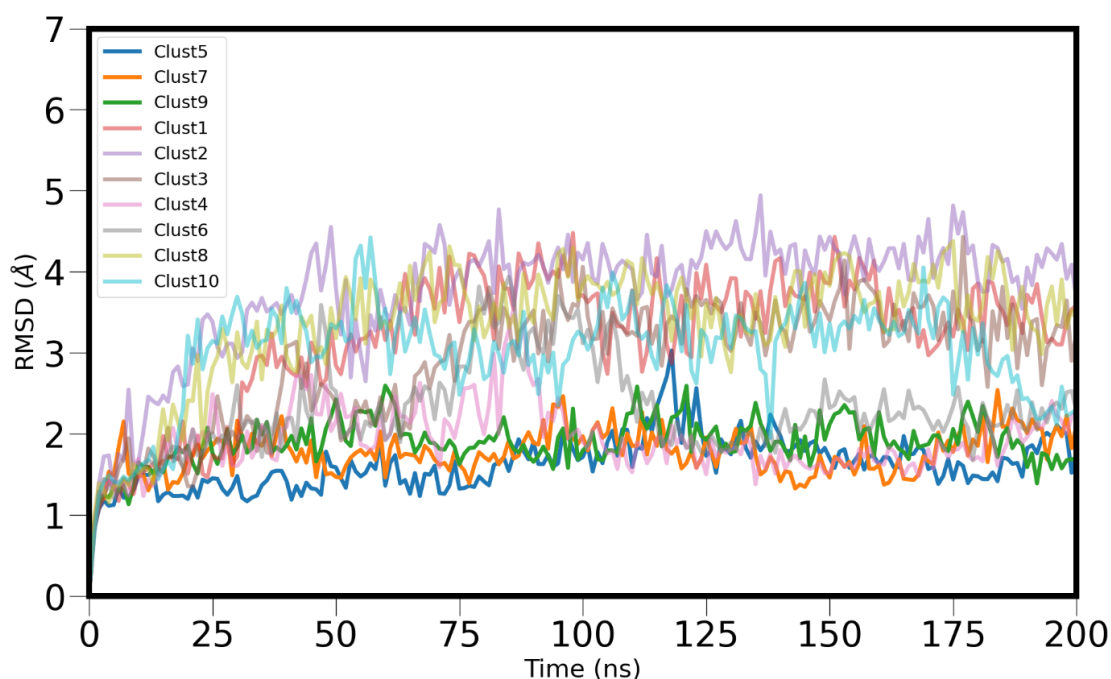


Figure 2.S5 Stability of PBP complexes in simulations. Root mean squared deviation (RMSD) of the PBP (after aligning the PBP to the initial frame) of the 10 nicotine-iDrugSnFR3a complexes, during the search for nicotine docking poses. Darker colors are used for the three trajectories that consistently stayed close to the holo conformation, while brighter curves illustrate simulations where the PBP binding cleft is (partially) opening towards the apo conformation.

Biofluids

Human serum was obtained from Sigma-Aldrich. Rat serum was obtained from Sigma-Aldrich. Heat treatment was for 65 °C, for 10 min, followed by passage through a 0.2 µm Millipore filter.

MD Simulations

MD simulations were performed using GROMACS2022 [41] utilizing AMBER ff19SB [42] and GAFF2 [43] 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 [44] 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 [45]. Temperature coupling was maintained at 300 K with the V-rescale algorithm [46]. 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 [15-17]. For each system, a network map or graph was created where nodes are the C α 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 i^{th} 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.

Differential Scanning Fluorimetry and Intrinsic Fluorescence Melts

Differential scanning fluorimetry (DSF) [47] was used to evaluate the thermal stability of iNicSnFR variants. In the DSF assay, samples containing 45 μ L of 5 μ M purified protein ($3 \times$ PBS, pH 7.0) with and without nicotine at twice the respective variant's EC_{50} value were mixed with 5 μ L of a stock solution of 200 $\times$ SYPRO[®] Orange dye (Invitrogen, Eugene, OR) in triplicate. Blank samples with the same concentrations of nicotine and SYPRO[®] Orange dye were also prepared in triplicate for blank subtraction. All samples were assayed on a CFX96 Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA). Samples underwent a thermal melt from 25 $^{\circ}$ C to 99 $^{\circ}$ C with temperature holds every 30 s at 0.5 $^{\circ}$ C increments. Fluorescence values were measured in relative fluorescence units (RFU) using the FRET scan mode. Backgrounds were subtracted from all samples.

Since the iNicSnFR variants have intrinsic fluorescence arising from the cpGFP moiety which overlaps with that of SYPRO[®] Orange dye, intrinsic fluorescence melts were performed in triplicate and the average fluorescence values were subtracted from the DSF samples. For the intrinsic fluorescence samples, 5 μ L of buffer ($3 \times$ PBS, pH 7.0) was added

to the protein samples in place of SYPRO[®] Orange dye. The intrinsic fluorescence samples underwent the same thermal melts as the DSF samples, and blank subtraction was performed. Apparent melting temperatures were determined from peaks in the first derivative plots of the melting curves by subtracting average intrinsic fluorescence.

Acknowledgments

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

Bibliography

1. Shivange AV, Borden PM, Muthusamy AK, Nichols AL, Bera K, Bao H, Bishara I, Jeon J, Mulcahy MJ, Cohen B, O'Riordan SL, Kim C, Dougherty DA, Chapman ER, Marvin JS, Looger LL, Lester HA. Determining the pharmacokinetics of nicotinic drugs in the endoplasmic reticulum using biosensors. *J Gen Physiol* 2019;151(4):738-57. 10.1085/jgp.201812201
2. Henderson BJ, Lester HA. Inside-out neuropharmacology of nicotinic drugs. *Neuropharmacol.* 2015;96(Pt B):178-93. 10.1016/j.neuropharm.2015.01.022
3. Benowitz NL, Hukkanen J, Jacob P, 3rd. Nicotine chemistry, metabolism, kinetics and biomarkers. *Handb Exp Pharmacol.* 2009(192):29-60. 10.1007/978-3-540-69248-5_2
4. Unger EK, Keller JP, Altermatt M, Liang R, Matsui A, Dong C, Hon OJ, Yao Z, Sun J, Banala S, Flanigan ME, Jaffe DA, Hartanto S, Carlen J, Mizuno GO, Borden PM, Shivange AV, Cameron LP, Sinning S, Underhill SM, Olson DE, Amara SG, Temple Lang D, Rudnick G, Marvin JS, Lavis LD, Lester HA, Alvarez VA, Fisher AJ, Prescher JA, Kash TL, Yarov-Yarovoy V, Gradinaru V, Looger LL, Tian L. Directed Evolution of a Selective and Sensitive Serotonin Sensor via Machine Learning. *Cell.* 2020;183(7):1986-2002.e26. 10.1016/j.cell.2020.11.040
5. Bera K, Kamajaya A, Shivange AV, Muthusamy AK, Nichols AL, Borden PM, Grant S, Jeon J, Lin E, Bishara I, Chin TM, Cohen BN, Kim CH, Unger EK, Tian L, Marvin JS, Looger LL, Lester HA. Biosensors Show the Pharmacokinetics of S-Ketamine in the Endoplasmic Reticulum. *Frontiers in cellular neuroscience.* 2019;13:499-. 10.3389/fncel.2019.00499
6. Nichols AL, Blumenfeld Z, Fan C, Luebbert L, Blom AEM, Cohen BN, Marvin JS, Borden PM, Kim CH, Muthusamy AK, Shivange AV, Knox HJ, Campello HR, Wang JH, Dougherty DA, Looger LL, Gallagher T, Rees DC, Lester HA. Fluorescence activation mechanism and imaging of drug permeation with new sensors for smoking-cessation ligands. *eLife.* 2022;11:e74648. 10.7554/eLife.74648
7. Nichols AL, Blumenfeld Z, Luebbert L, Knox HJ, Muthusamy AK, Marvin JS, Kim CH, Grant SN, Walton DP, Cohen BN, Hammar R, Looger L, Artursson P, Dougherty DA, Lester HA. Selective Serotonin Reuptake Inhibitors within Cells: Temporal Resolution in Cytoplasm, Endoplasmic Reticulum, and Membrane. *J Neurosci.* 2023;43:2222-41. 10.1523/jneurosci.1519-22.2022
8. Moore JC, Jin HM, Kuchner O, Arnold FH. Strategies for the in vitro evolution of protein function: enzyme evolution by random recombination of improved sequences. *J Mol Biol.* 1997;272(3):336-47. 10.1006/jmbi.1997.1252
9. Spiller B, Gershenson A, Arnold FH, Stevens RC. A structural view of evolutionary divergence. *Proc Natl Acad Sci U S A.* 1999;96(22):12305-10. 10.1073/pnas.96.22.12305
10. Treynor TP, Vizcarra CL, Nedelcu D, Mayo SL. Computationally designed libraries of fluorescent proteins evaluated by preservation and diversity of function. *Proc Natl Acad Sci U S A.* 2007;104(1):48-53. 10.1073/pnas.0609647103

11. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*. 2010;31(2):455-61. 10.1002/jcc.21334
12. Packer MS, Liu DR. Methods for the directed evolution of proteins. *Nat Rev Genet*. 2015;16(7):379-94. 10.1038/nrg3927
13. Fukami-Kobayashi K, Tateno Y, Nishikawa K. Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *J Mol Biol*. 1999;286(1):279-90. 10.1006/jmbi.1998.2454
14. Berntsson RP, Smits SH, Schmitt L, Slotboom DJ, Poolman B. A structural classification of substrate-binding proteins. *FEBS Lett*. 2010;584(12):2606-17. 10.1016/j.febslet.2010.04.043
15. Sethi A, Eargle J, Black AA, Luthey-Schulten Z. Dynamical networks in tRNA:protein complexes. *Proc Natl Acad Sci U S A*. 2009;106(16):6620-5. 10.1073/pnas.0810961106
16. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996;14(1):33-8, 27-8. 10.1016/0263-7855(96)00018-5
17. Eargle J, Luthey-Schulten Z. NetworkView: 3D display and analysis of protein-RNA interaction networks. *Bioinformatics*. 2012;28(22):3000-1. 10.1093/bioinformatics/bts546
18. Khan HM, MacKerell AD, Jr., Reuter N. Cation- π Interactions between Methylated Ammonium Groups and Tryptophan in the CHARMM36 Additive Force Field. *J Chem Theory Comput*. 2019;15(1):7-12. 10.1021/acs.jctc.8b00839
19. Liu H, Fu H, Shao X, Cai W, Chipot C. Accurate Description of Cation- π Interactions in Proteins with a Nonpolarizable Force Field at No Additional Cost. *Journal of Chemical Theory and Computation*. 2020;16(10):6397-407. 10.1021/acs.jctc.0c00637
20. Chen TW, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA, Orger MB, Jayaraman V, Looger LL, Svoboda K, Kim DS. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*. 2013;499(7458):295-300. 10.1038/nature12354
21. Barnett LM, Hughes TE, Drobizhev M. Deciphering the molecular mechanism responsible for GCaMP6m's Ca^{2+} -dependent change in fluorescence. *PLoS One*. 2017;12(2):e0170934. 10.1371/journal.pone.0170934
22. Nasu Y, Shen Y, Kramer L, Campbell RE. Structure- and mechanism-guided design of single fluorescent protein-based biosensors. *Nat Chem Biol*. 2021;17(5):509-18. 10.1038/s41589-020-00718-x
23. Bligh J. The level of free choline in plasma. *J Physiol*. 1952;117(2):234-40. 10.1113/jphysiol.1952.sp004743
24. Zeisel SH, Epstein MF, Wurtman RJ. Elevated choline concentration in neonatal plasma. *Life Sci*. 1980;26(21):1827-31.
25. Klein J, Koppen A, Löffelholz K, Schmitthenner J. Uptake and Metabolism of Choline By Rat-Brain After Acute Choline Administration. *Journal of Neurochemistry*. 1992;58(3):870-6.

26. Savendahl L, Mar MH, Underwood LE, Zeisel SH. Prolonged fasting in humans results in diminished plasma choline concentrations but does not cause liver dysfunction. *Am J Clin Nutr.* 1997;66(3):622-5. 10.1093/ajcn/66.3.622
27. Hoffman MC, Hunter SJ, D'Alessandro A, Christians U, Law AJ, Freedman R. Maternal Plasma Choline during Gestation and Small for Gestational Age Infants. *Am J Perinatol.* 2022. 10.1055/s-0042-1759775
28. Walker NB, Yan Y, Tapia MA, Tucker BR, Thomas LN, George BE, West AM, Marotta CB, Lester HA, Dougherty DA, Holleran KM, Jones SR, Drenan RM. $\beta 2$ nAChR Activation on VTA DA Neurons Is Sufficient for Nicotine Reinforcement in Rats. *eNeuro.* 2023;10(5):ENEURO.0449-22.2023. 10.1523/eneuro.0449-22.2023
29. Dempsey D, Tutka P, Jacob P, 3rd, Allen F, Schoedel K, Tyndale RF, Benowitz NL. Nicotine metabolite ratio as an index of cytochrome P450 2A6 metabolic activity. *Clin Pharmacol Ther.* 2004;76(1):64-72. 10.1016/j.clpt.2004.02.011
30. Liakoni E, Tyndale RF, Jacob P, Dempsey DA, Addo N, Benowitz NL. Effect of race and glucuronidation rates on the relationship between nicotine metabolite ratio and nicotine clearance. *Pharmacogenetics and genomics.* 2021;31(5):97-107. 10.1097/fpc.0000000000000427
31. Kille S, Acevedo-Rocha CG, Parra LP, Zhang ZG, Opperman DJ, Reetz MT, Acevedo JP. Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS synthetic biology.* 2013;2(2):83-92. 10.1021/sb300037w
32. Studier FW. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification.* 2005;41(1):207-34. 10.1016/j.pep.2005.01.016
33. Wurth C, Grabolle M, Pauli J, Spieles M, Resch-Genger U. Relative and absolute determination of fluorescence quantum yields of transparent samples. *Nat Protoc.* 2013;8(8):1535-50. 10.1038/nprot.2013.087
34. Patterson G, Day RN, Piston D. Fluorescent protein spectra. *J Cell Sci.* 2001;114(Pt 5):837-8. 10.1242/jcs.114.5.837
35. Haloi N, Vasani AK, Geddes EJ, Prasanna A, Wen P-C, Metcalf WW, Hergenrother PJ, Tajkhorshid E. Rationalizing the generation of broad spectrum antibiotics with the addition of a positive charge. *Chemical Science.* 2021;12(45):15028-44. 10.1039/D1SC04445A
36. Marques R, Bouville C, Ribardi re M, Santos LP, Bouatouch K. Spherical Fibonacci Point Sets for Illumination Integrals. *Computer Graphics Forum.* 2013;32(8):134-43. 10.1111/cgf.12190
37. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. *J Cheminform.* 2011;3:33. 10.1186/1758-2946-3-33
38. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kal  L, Schulten K. Scalable molecular dynamics with NAMD. *Journal of computational chemistry.* 2005;26(16):1781-802. 10.1002/jcc.20289
39. Phillips JC, Hardy DJ, Maia JDC, Stone JE, Ribeiro JV, Bernardi RC, Buch R, Fiorin G, H nin J, Jiang W, McGreevy R, Melo MCR, Radak BK, Skeel RD, Singharoy A, Wang Y, Roux B, Aksimentiev A, Luthey-Schulten Z, Kal  LV, Schulten K, Chipot

- C, Tajkhorshid E. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys.* 2020;153(4):044130. 10.1063/5.0014475
40. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica D Biological Crystallography.* 2010;66(Pt 2):213-21. 10.1107/S0907444909052925
 41. Páll S, Zhmurov A, Bauer P, Abraham M, Lundborg M, Gray A, Hess B, Lindahl E. Heterogeneous parallelization and acceleration of molecular dynamics simulations in GROMACS. *J Chem Phys.* 2020;153(13):134110. 10.1063/5.0018516
 42. Tian C, Kasavajhala K, Belfon KAA, Raguette L, Huang H, Migués AN, Bickel J, Wang Y, Pincay J, Wu Q, Simmerling C. ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J Chem Theory Comput.* 2020;16(1):528-52. 10.1021/acs.jctc.9b00591
 43. He X, Man VH, Yang W, Lee TS, Wang J. A fast and high-quality charge model for the next generation general AMBER force field. *J Chem Phys.* 2020;153(11):114502. 10.1063/5.0019056
 44. Darden T, York D, Pedersen L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics.* 1993;98(12):10089-92. 10.1063/1.464397
 45. Parrinello M, Rahman A. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Physical Review Letters.* 1980;45(14):1196-9. 10.1103/PhysRevLett.45.1196
 46. Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys.* 2007;126(1):014101. 10.1063/1.2408420
 47. Niesen FH, Berglund H, Vedadi M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols.* 2007;2(9):2212-21. 10.1038/nprot.2007.321

COMPUTATIONAL DESIGN OF PERIPLASMIC BINDING PROTEIN-BASED BIOSENSOR ENABLES MEASUREMENT OF FULL RANGE OF PHARMACOLOGICALLY RELEVANT NICOTINE CONCENTRATIONS

Abstract

To understand differences in the time course of [nicotine] between individuals and nicotine delivery systems, sensors capable of robustly measuring 10-200 nM nicotine are needed. By employing computational protein design strategies aimed at improving the thermostability of the previously reported iNicSnFR12, we discovered four mutants with improved sensitivity for nicotine, including iNicSnFR12-D72A. This improved fluorescent protein sensor of nicotine has an EC_{50} of 440 nM and an $11.6 \mu\text{M}^{-1}$ δ -slope, $(\Delta F/F_0)/([\text{nicotine}])$ at the beginning of the dose-response relation, at intradermal conditions. These sensitivity enhancements represent 4.4-fold and 3.3-fold improvements in EC_{50} and δ -slope, respectively, over iNicSnFR12. In iNicSnFR3a, a previous variant, published structures indicate that a hydrogen bond between the sidechains of H68 and D72 present in the apo state is broken upon binding of nicotine or varenicline, resulting in H68 adopting a sidechain rotamer oriented towards the binding pocket. We hypothesize that in iNicSnFR12-D72A, the H68 sidechain is free to interact with the ligand since no hydrogen bond would exist with the sidechain of A72, favoring the ligand-bound state of the sensor. With its improved sensitivity and suitable thermostability characteristics, iNicSnFR12-D72A represents the most promising biosensor for measuring temporal changes in nicotine concentration, [nicotine], in a continuous monitor worn by nicotine users.

Introduction

Globally, approximately 1.2 billion people consume tobacco and tobacco-derived products [1], resulting in over 7 million annual deaths from tobacco-related diseases [2]. Understanding the pharmacological effects of addiction to nicotine, the addictive component of tobacco, has therefore been a critical area of investigation. The mechanisms of addiction and dependence to nicotine are thought to involve the drug acting as a chaperone of nicotinic acetylcholine receptors (nAChR), resulting in their upregulation in neurons [3]. The full time course of [nicotine]_t, denoted as [nicotine]_t, is relevant to nicotine addiction [4]. This time course exhibits nicotine concentrations as low as 10 nM between smoking sessions and as high as 200 nM following nicotine ingestion [5], with waveforms varying widely between individuals and methods of nicotine delivery [4].

Due to widespread usage of tobacco, especially in low- and middle-income countries [1], the need for monitoring [nicotine]_t in a cheaper fashion than the standard method involving blood draws by intravenous catheters and analyzing samples by liquid chromatography-mass spectrometry (LC-MS) assays. The series of intensity-based nicotine-sensing fluorescent reporters (iNicSnFRs) have come closest to realizing this goal [4, 6-8], with initial progress having been made at encapsulation of biosensors in hydrogels [9], which would be amenable to continuous monitoring of [nicotine]_t in interstitial fluid (ISF) via an implantable format [4].

Members of the iNicSnFR series are comprised of engineered variants of the periplasmic binding protein (PBP) OpuBC from *Thermoanaerobacter* sp. X513 with Superfolder GFP [10] circularly permuted (cpsfGFP). As determined from analyses of unbound (apo) and bound structures of iNicSnFR3a [7], fluorescence activation occurs via a “candle-snuffer” mechanism. In the unbound (apo) state, the lobes of the PBP are in an open conformation, and in the ligand-bound state, the lobes of the PBP close. In the apo state, the side chain of the “candle-snuffer” residue E78, which is on a linker between the PBP and cpsfGFP domains, interacts with the chromophore of cpsfGFP, diminishing its fluorescence. The apo state fluorescence is denoted as F_0 . In the bound state, E78 pulls away from the chromophore,

resulting in increased fluorescence, denoted as F . F and fluorescence fold change, $\Delta F/F_0$, vary with [drug] in a dose-dependent manner [6-8].

To enhance the sensitivity of iNicSnFR12 [8], protein design techniques that have been demonstrated to improve thermostability were employed on the PBP moiety, denoted as PBP12. This approach was informed by the observation that enhancing a protein's thermostability often improves protein function [11].

Results and Discussion

Potentially stabilizing mutants of iNicSnFR12 were found using homology-based consensus design [12] of PBP12 and protein design calculations aimed at stabilizing the α -helices [13] of PBP12. Consensus design yielded 22 single mutants (see **Table A.1**), and protein design calculations of α -helices yielded 66 single mutants (see **Table A.2**). All single mutants were constructed and assayed for nicotine-induced fluorescence. The fourteen most promising variants from the lysate screen were purified, and full nicotine dose-response relations were determined (see **Figure A.1**). The four purified variants with significantly better nicotine sensitivity than iNicSnFR12 as determined by EC_{50} and S-slope, calculated as $\Delta F_{\max}/F_0$, are presented in **Figure 3.1** with nicotine sensitivity metrics shown in **Table 3.1**.

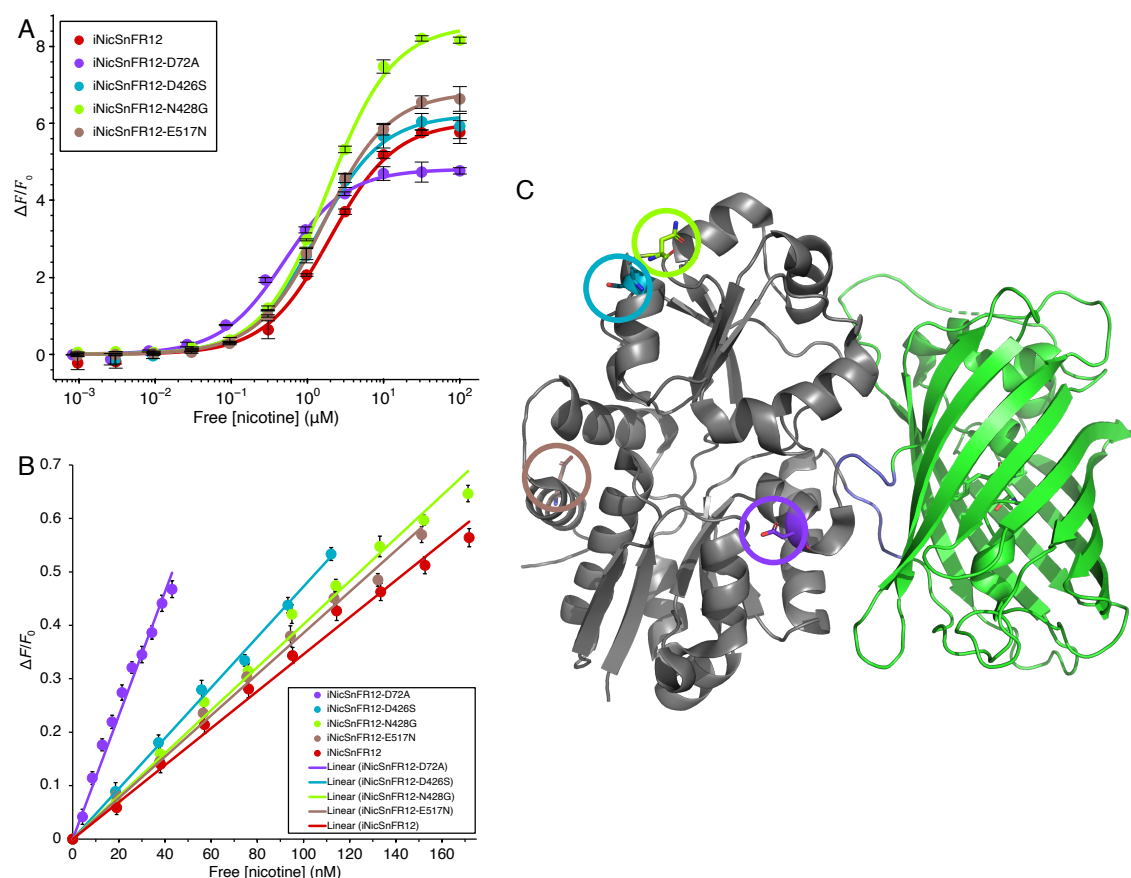


Figure 3.1 Single mutants of iNicSnFR12 with enhanced sensitivity for nicotine. (A) Full dose-response relations for purified iNicSnFR12 [8] and purified single mutants of iNicSnFR12 that exhibited significantly better nicotine sensitivity (see **Table 3.1**) compared to the parent iNicSnFR12. Data are mean $\pm$ SEM (n=3). (B) Dose-response relations in linear regime for purified variants as in **Figure 3.1A**. Data are mean $\pm$ SEM (n=8). (C) Positions of sensitivity-enhancing mutations as shown in the most recent nicotine-bound iNicSnFR experimental structure, that of iNicSnFR3a [6, 7] (PDB: 7S7U).

Table 3.1: Nicotine sensitivity metrics from fitted dose-response data. P-values for comparing metrics between single mutants and iNicSnFR12 are denoted as follows— $p < 0.05$: *, $p < 0.01$: **, $p < 0.001$: ***.

Variant	EC_{50} (μM)	S-slope (μM^{-1})	δ -slope (μM^{-1})
iNicSnFR12	1.92 ± 0.15	3.1 ± 0.2	3.5
D72A	0.44 ± 0.03 (**)	11 ± 0.8 (***)	11.6
D426S	1.28 ± 0.11	4.9 ± 0.4 (*)	4.7
N428G	1.79 ± 0.06	4.8 ± 0.2 (*)	4.0
E517N	1.57 ± 0.09	4.3 ± 0.3 (*)	3.9

iNicSnFR12-N428G was found via consensus design, while iNicSnFR12-D72A, iNicSnFR12-D426S, and iNicSnFR12-E517N came from the protein design calculations on α -helices. While the D426S, N428G, and E517N variants of iNicSnFR12 exhibited significantly better S-slope values as derived from fitted full dose-response data, the only variant with a significantly improved EC_{50} value for nicotine is iNicSnFR12-D72A, which is 4.4 times lower than parent iNicSnFR12 (see **Figure 3.1A** and **Table 3.1**). Additionally, iNicSnFR12-D72A exhibits a 3.3-fold improvement in δ -slope (see **Table 3.1**), the slope of the linear regime of the dose-response relation (see **Figure 3.1B**). Interestingly, these four sensitivity-enhancing mutations do not occur in the binding pocket between the upper and lower lobes of the PBP. These four mutations were combined in multi-mutants of iNicSnFR12 denoted as iNicSnFR12-M1(D72A, N428G) and iNicSnFR12-M2 (D72A, D426S, N428G, and E517N), but neither multi-mutant was as sensitive to nicotine as iNicSnFR12-D72A (see **Figure A.2**).

A structural analysis of helix 3 (according to conventional helix numbering [14]) was performed to rationalize the sensitivity-enhancing effect of the D72A mutation on iNicSnFR12 and is presented in **Figure 3.2**.

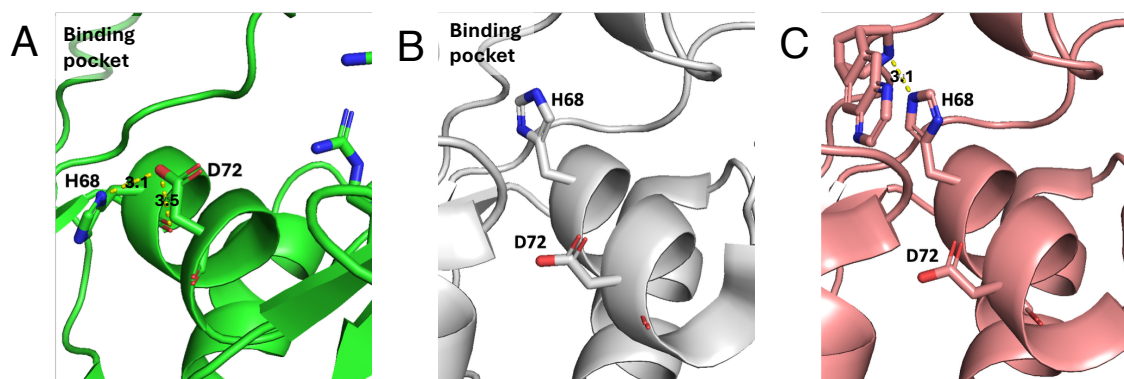


Figure 3.2 Structural rationale for improved nicotine sensitivity conferred by D72A. Interatom distances measured using PyMOL. Snapshots exhibiting helix 3 in crystal structures for iNicSnFR3a, the most recent iNicSnFR variant with experimental structures are shown [6, 7]. Residues in helix 3 and its environs were not mutated in the development of iNicSnFR12 [8]. (A) Apo structure (PDB: 7S7V). (B) Nicotine-bound structure (PDB: 7S7U). (C) Varenicline-bound structure (PDB: 7S7T).

In the apo structure shown in **Figure 3.2A**, a hydrogen bond is formed between the sidechains of H68 and D72. In the nicotine-bound structure (**Figure 3.2B**) and varenicline-bound structure (**Figure 3.2C**), the hydrogen bond between the sidechains of H68 and D72 is abrogated, enabling the sidechain of H68 to orient towards the binding pocket. Since nicotine has a rotatable bond, its atomic coordinates could not be resolved in the nicotine-bound structure; however, it has been shown to occupy the same binding pocket as varenicline, a smoking cessation ligand [7]. The D72A mutation would abrogate side-chain hydrogen bonding with H68, allowing its sidechain to adopt a rotamer oriented towards the binding packet, even in the apo state. This presumed reorientation of the sidechain of H68, biasing the ligand-bound conformation, could contribute to the enhanced sensitivity of iNicSnFR12-D72A over iNicSnFR12.

For the proposed use of iNicSnFR to measure $[\text{nicotine}]_t$ in biofluids, exogenous ligand varenicline and the endogenous ligand choline are relevant for biosensor activation. Comparisons of fluorescence dose-response relations for nicotine, varenicline, and choline are presented in **Figure 3.3**.

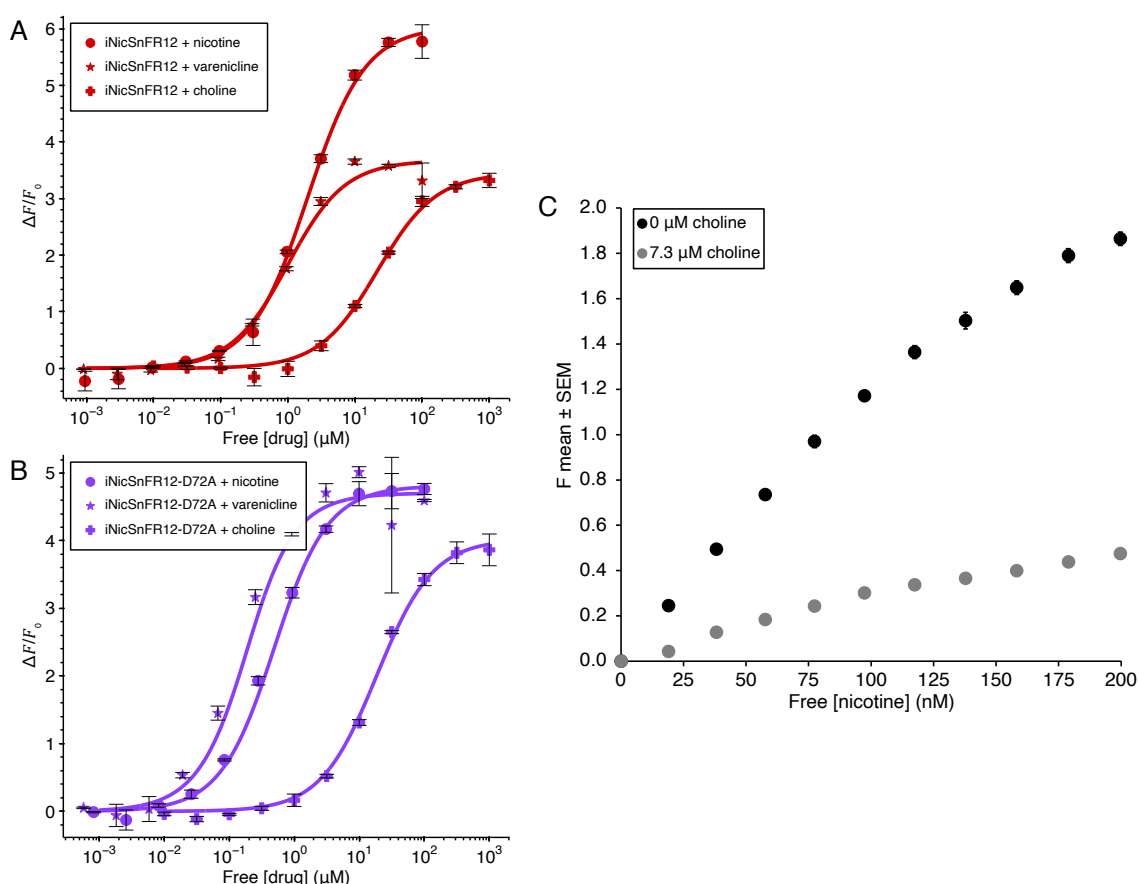


Figure 3.3 Dose-response relations for nicotine, varenicline, and choline. Measurements were made with purified biosensors. (A) iNicSnFR12 dose-response relations with each drug. Data are mean $\pm$ SEM ($n=3$). The fitted EC_{50} values are $1.92 \pm 0.15 \mu\text{M}$, $0.95 \pm 0.11 \mu\text{M}$, and $21.54 \pm 1.82 \mu\text{M}$ for nicotine, varenicline, and choline, respectively. (B) iNicSnFR12-D72A dose-response relations with each drug. Data are mean $\pm$ SEM ($n=3$). The fitted EC_{50} values are $0.44 \pm 0.03 \mu\text{M}$, $0.13 \pm 0.03 \mu\text{M}$, and $18.55 \pm 1.39 \mu\text{M}$ for nicotine, varenicline, and choline, respectively. (C) Dose-response relation of iNicSnFR12-D72A at

pharmacologically relevant nicotine concentrations with and without 7.3 μM choline, the mean [choline] in adult plasma [15]. Data are mean $\pm$ SEM (n=8).

As seen in **Figures 3.3A** and **3.3B**, the sensitivity-enhancing effect conferred by the D72A mutation applies to varenicline as well as nicotine, lending support to the hypothesis that H68 is important to the binding of nicotine. The current sensor could be useful for measuring personal [varenicline]_i, with the relevant concentration range below the maximum measured plasma varenicline concentration of 134 nM [16] falling within the range of sensor-activating varenicline concentrations. Although iNicSnFR12-D72A is 42-fold more sensitive for nicotine than choline, sensor activation by choline poses a challenge for determining [nicotine], as the mean [choline] for adults is 7.3 μM [15] while [nicotine]_i is less than 200 nM [5], representing a 37-fold difference in ligand abundance. **Figure 3.3C** illustrates the dampening effect that the physiological [choline] has on the dose-response relation of iNicSnFR12-D72A for the pharmacologically relevant [nicotine]. Fluctuations in [choline] could foreseeably be background-subtracted from measured [nicotine] with the use of an orthogonal iDrugSnFR with activation for choline but not nicotine. However, the low fluorescence response of iNicSnFR12-D72A in the presence of choline (see **Figure 3.3C**), makes measuring the lower range of pharmacologically relevant [nicotine] challenging. Alternatively, encapsulating hydrogels bearing iNicSnFR in a choline-impermeable but nicotine-permeable material could allow the measurement of [nicotine] without interference from choline, while introducing an additional transport limitation.

The thermostability of iNicSnFR12 and its four improved variants are shown in **Figure 3.4**.

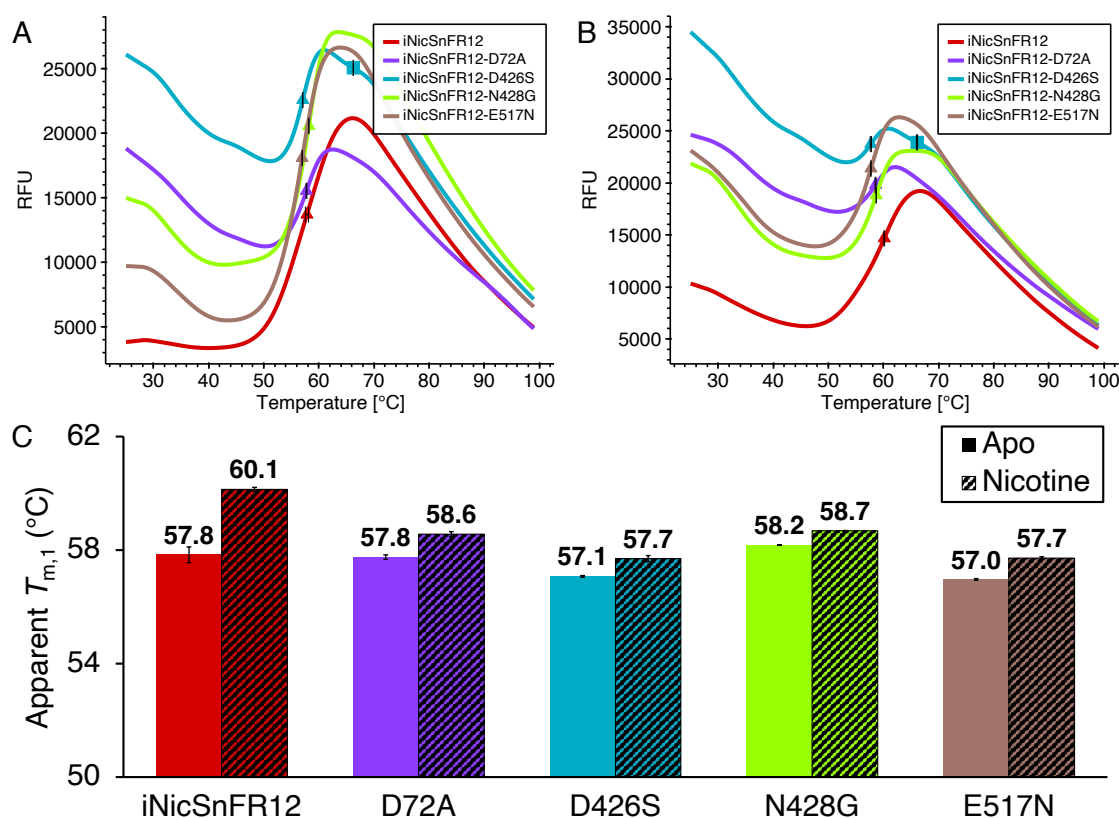


Figure 3.4 Thermostability assessment of iNicSnFR variants via DSF. Line plots represent mean background subtracted fluorescence for each purified biosensor (n=3). Markers indicating the mean $\pm$ SEM (n=3) first and second apparent melting temperatures are shown as triangles and squares, respectively. Purified biosensors were assayed at 5 μ M except for iNicSnFR12-D72A, which was assayed at 3.5 μ M. **(A)** DSF melt curve without nicotine. Values for the first apparent melting temperature $T_{m,1}$ are presented in **Figure 3.4C**. The only variant exhibiting a second melting temperature was iNicSnFR12-D426S, with $T_{m,2}$ occurring at 66.2 ± 0.0 °C. **(B)** DSF melt curve with nicotine. Values for the first apparent melting temperature $T_{m,1}$ are presented in **Figure 3.4C**. The only variant exhibiting a second melting temperature was iNicSnFR12-D426S, with $T_{m,2}$ occurring at 66.1 ± 0.1 °C. **(C)** Summary of $T_{m,1}$ values from **Figure 3.4A-B**.

Although the mutations were expected to thermostabilize the protein, only iNicSnFR12-N428G had a modestly higher $T_{m,1}$ value than iNicSnFR12 in apo conditions (**Figure 3.4A**

and **3.4C**), and none of the variants had higher $T_{m,1}$ values than iNicSnFR12 in the presence of nicotine (**Figure 3.4B-C**). To ensure stability at physiological temperature for the duration of a weeklong continuous nicotine monitoring trial, thermal hold experiments were performed with iNicSnFR12 and the top mutant iNicSnFR12-D72A. **Figure 3.5** shows functional and stability results over the course of the 7-day, 37 °C hold.

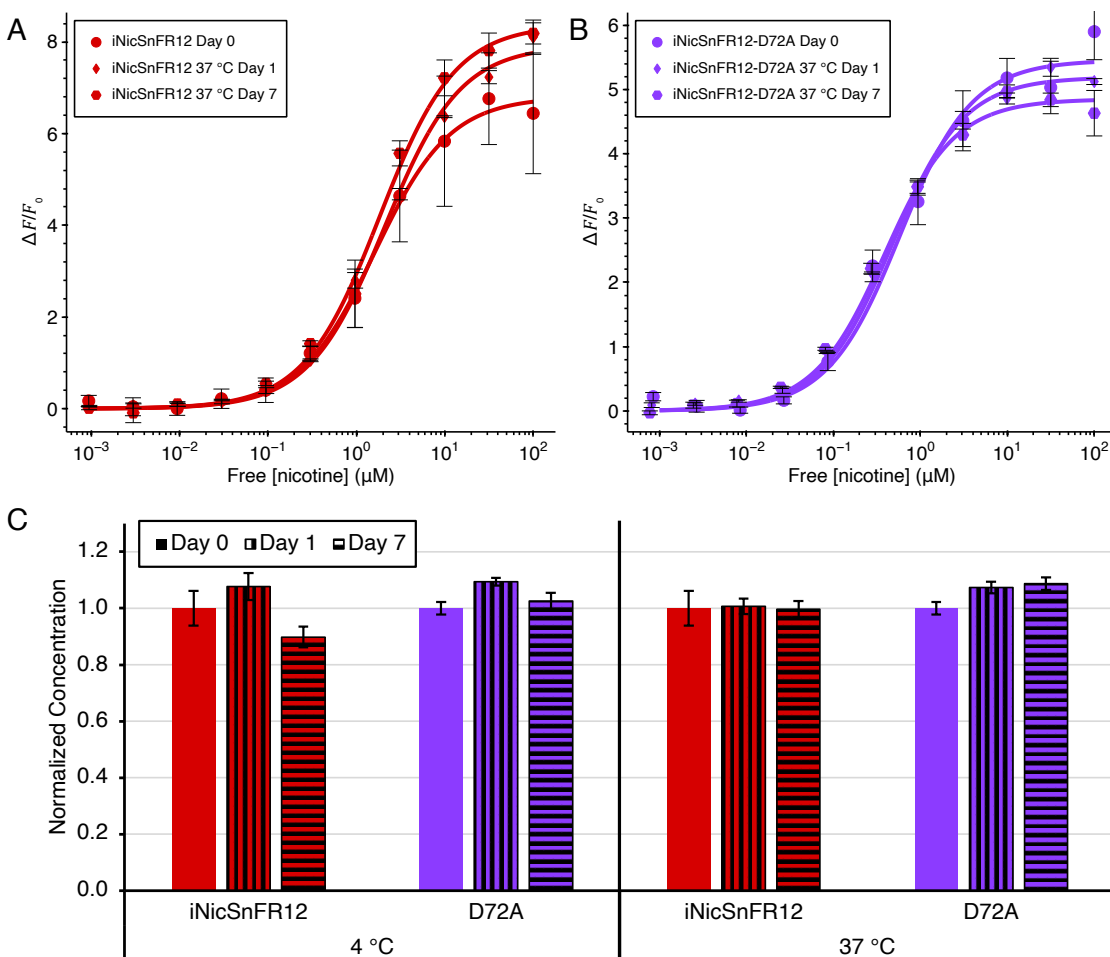


Figure 3.5 Assessments of iNicSnFR12 and iNicSnFR12-D72A resilience to thermal stress. (A) Full dose-response relations for purified iNicSnFR12 on days 0, 1, and 7 of 37 °C incubation. Fitted parameter values for the Hill function were not statistically different. (B) Full dose-response relations for purified iNicSnFR12-D72A on days zero, one, and seven of 37 °C incubation. Fitted parameter values for the Hill function were not statistically

different. (C) Normalized concentrations of iNicSnFR12 and iNicSnFR12-D72A on days 0, 1, and 7 of 37 °C incubation.

Figures 3.5A-B indicate that nicotine-sensitivity was unaffected by the 7-day, 37 °C hold, and **Figure 3.5C** demonstrates that both iNicSnFR12 and iNicSnFR12-D72A remain stable in solution under this thermal stress.

The fluorescence quantum yield Φ was measured for iNicSnFR12 and iNicSnFR12-D72A as 0.669 and 0.733, respectively, as measured in PBS, pH 7.4. These values resemble Φ values for iNicSnFR3a and iNicSnFR12 measured in $3 \times$ PBS, pH 7.4, which were reported previously [8].

iNicSnFR12-D72A is the most sensitive nicotine biosensor to date, with an EC_{50} of 440 nM and a δ -slope of $11.6 \mu\text{M}^{-1}$. Further development of iNicSnFR12-D72A as a continuous nicotine monitor should involve the incorporation of the biosensor in a hydrogel, as has been demonstrated previously using iNicSnFR12 [9], with a method for distinguishing [nicotine] from [choline].

Methods

Mutant Library Generation via Physics-Based Design of α -Helices

Fixed-backbone sequence design was used to find putatively stabilizing mutations to the inherent dipoles in α -helices [13] of the PBP in iNicSnFR. Helix dipoles result from three unsatisfied hydrogen bonds on the helix backbone at each of the termini [17-20]. iNicSnFR3a is the closest variant to iNicSnFR12 (N11E, Q431D, R467S, E476K) [8] with experimental structures [7]. Therefore, the structure for iNicSnFR3a with nicotine (PDB: 7S7U) [7] was used as the target structure using the Triad Sequence-Design module (Protabit LLC). For each α -helix, the eight positions of interest were the position immediately preceding the helix (“N-capping” position), the three most N-terminal helical positions, the three most C-

terminal helical positions, and the position immediately following the helix. Additionally, wild-type Gly residues within helices were mutable since Gly has low helix propensity [21, 22]. Residues with pre-existing interactions and positions with wild-type residues that did not violate helix dipole “rules” (positively charged residues at the four most N-terminal positions of interest and negatively charged residues at the four most C-terminal positions of interest) were not allowed to mutate.

Two different schemes were used to determine which residues were allowed at mutable positions. Under the regular scheme, the wild-type amino acid and residues that would not violate helix dipole “rules” (positively charged residues at the four most N-terminal positions of interest and negatively charged residues at the four most C-terminal positions of interest) were allowed. Under the strict scheme, only wild-type and charge-stabilizing residues (negatively charged residues at four most N-terminal positions of interest and positively charged residues at the four most C-terminal positions of interest) were allowed. Rotamers for amino acids within 5 Å of each mutable position were allowed to repack. Design calculations were run under the regular and strict schemes with the Rosetta [23] and Phoenix [24] force yields for a total of four calculations per helix. Out of designs with calculated force-field energies equivalent to ($\pm 0.1\%$) or more stable than the wildtype sequence, mutations in the top five designs were included in the mutagenic library, excluding His and Cys.

Mutant Library Generation via Consensus Design

Protein basic local alignment search tool (BLAST) searches were performed using the Bio.Blast package in Biopython [25] to access the National Center for Biotechnology Information (NCBI) BLAST [26] web service to construct a multiple sequence alignment (MSA) with sequences having 30-90% sequence identity with the PBP of iNicSnFR12. Consensus mutations were determined for more conserved positions in the MSA because these mutations are more likely to stabilize a protein [12]. Positions were deemed more conserved if the relative entropy (RE) values were greater than the average RE for all positions in the PBP. RE for each position is calculated using **Equation 3.1**:

$$RE = \sum_x p_x \ln \frac{p_x}{f_x} \quad (3.1)$$

with p_x denoting the frequency of amino acid x in the MSA and f_x denoting the frequency of amino acid x in proteins of *Bacillus subtilis*, which makes a homologous OpuBC to that of *Thermoanaerobacter sp. X513*, which was the source of the PBP engineered for the first iNicSnFR [6]. Amino acid frequencies for *T. sp. X513* are unknown; therefore, amino acid frequencies of *B. subtilis* were used in calculations and are as follows—A: 7.6%, C: 0.8%, D: 5.2%, E: 7.1%, F: 4.4%, G: 7.0%, H: 2.3%, I: 7.3%, K: 6.9%, L: 9.6%, M: 2.6%, N: 4.1%, P: 3.8%, Q: 3.9%, R: 4.1%, S: 6.4%, T: 5.4%, V: 6.6%, W: 1.1%, Y: 3.6% [27, 28]. At each position in the PBP of iNicSnFR12 with above average RE , if the consensus (most prevalent) amino acid was different than the wildtype amino acid, the consensus amino acid was included in the mutagenic library. The degree of correlation in pairwise interactions between positions in the MSA was determined using mutual information (MI), which was calculated using **Equation 3.2**:

$$MI(i, j) = \sum_i \sum_j p_{x,y} \ln \frac{p_{x,y}}{p_x p_y} \quad (3.2)$$

with p_x and p_y denoting the frequencies of amino acids x and y , respectively, in the MSA and $p_{x,y}$ denoting the frequency of co-occurrence of amino acid x at position i and amino acid y at position j in the MSA. Since the RE screen yielded a manageable number of mutations to test experimentally, MI was not used to eliminate positions from consideration.

Mutant Construction and Screening in Cell Lysate

Mutants of the parent molecule iNicSnFR12 [8] were generated via site-directed mutagenesis (SDM) or Gibson assembly [29] using commercially available reagents (New England Biolabs). Following sequence confirmation, *E. coli* BL21-Gold (DE3) chemically competent cells (Agilent) were transformed. Cultures for iNicSnFR variants were expressed in 1-mL cultures in 96-well deep well plates (Thermo Scientific) and screened in 96-well plate format. Cultures were grown in ZYM5052 autoinduction media [30] with 100 µg/mL ampicillin at 30 °C for 30 h shaking at 250 rpm. Cells were pelleted at 4,000 rpm for 12 min at 4 °C, media

was decanted off and cells were resuspended in 1 mL of $3 \times$ PBS pH 7.4. Resuspended cells were lysed via two freeze-thaw cycles using liquid N_2 and an ambient-temperature water bath. The plate was then centrifuged at 4,000 rpm for 12 min at 4 °C to pellet intact cells and cell debris. Supernatant was transferred to black polystyrene 96-well plates (Corning) and analyzed in triplicate. Baseline fluorescence (F_0) and nicotine-induced fluorescence (F) were measured using a Spark® 20M 96-well fluorescence plate reader (Tecan). Nicotine concentrations within an order of magnitude of the EC_{50} of the purified parent iNicSnFR12 [8] were used. Variants with the top five highest $\Delta F/F_0$ values for 300 nM and 1 μ M nicotine as well as variants with positive correlations between $\Delta F/F_0$ and temperature were further characterized.

Expression and Purification of iNicSnFRs

Protein expression was performed in *E. coli* BL21-Gold (DE3) chemically competent cells (Agilent) in 200-300 mL of ZYM5052 autoinduction media [30]. Cells were pelleted at $5,000 \times g$ for 10 min and stored at -20 °C until used. Thawed cell pellets were resuspended in 20-30 mL PBS, pH 7.4 (5 mL PBS per 50 mL of media), and lysed by sonication in an ice-water bath to mitigate sample warming. Remaining intact cells and cell debris were removed by two rounds of centrifugation at 5,000 rpm for 12 min with supernatant transferred to a clean 50-mL conical tube between spins. Resulting supernatant was then filtered using a 25 mm diameter 0.22 μ M sterile filter (VWR) and loaded onto a Ni-NTA column equilibrated with wash buffer. A 20 mM imidazole in PBS, pH 7.4 wash buffer was used. Fractions were collected from eluting with a 20-200 mM imidazole gradient. Fractions containing proteins of the target mass (determined by SDS-PAGE) were pooled, concentrated and buffer exchanged into PBS, pH 7.40 by centrifugation with Amicon Ultra 15 filter units with 30-kD molecular weight cutoff (Millipore).

Concentration Determination of iNicSnFRs

Biosensor concentration was determined by denaturing an aliquot in 1 M NaOH and measuring 447-nm absorption on a NanoDrop spectrophotometer using the known denatured chromophore extinction coefficient of 44,000 M⁻¹cm⁻¹ [31].

Measurement of Fluorescence Dose-Response Relation

Purified variants of iNicSnFR12 were mixed with drug solutions to generate samples containing 100 nM biosensor and drug at the appropriate concentration for dose-response characterization. Solutions were prepared in 1% BSA, PBS pH 7.40 to mitigate biosensor adsorption to wells. Sample replicates were prepared and analyzed in three to eight wells (see individual figures) of black polystyrene flat bottom 96-well plates (Corning) using an epMotion[®] 5075 automated liquid handler (Eppendorf). Biosensor fluorescence (F_0 and ΔF) was measured with 485 nm excitation and 535 nm emission wavelengths after plates were incubated at 32 °C for 20 min using a Spark[®] 20M plate reader (Tecan). The analysis temperature was chosen to reflect data for skin surface temperatures [32] since an intended use of iNicSnFRs is wearable sensing of nicotine for personal pharmacokinetics [4]. Dose-response data in the linear regime, where [nicotine] < 0.1 × EC_{50} for a given variant, were fitted using Excel (Microsoft). Sigmoidal dose-response data were fitted to the Hill equation [33, 34] using SciPy [35] with the free ligand concentration [L] calculated using **Equation 3.3** from the ligand depletion model assuming a single binding site:

$$[L] = [L_{\text{tot}}] - \frac{[L_{\text{tot}}] + [S_{\text{tot}}] + K_D - \sqrt{([L_{\text{tot}}] + [S_{\text{tot}}] + K_D)^2 - 4[L_{\text{tot}}][S_{\text{tot}}]}}{2} \quad (3.3)$$

[L_{tot}] is the total ligand concentration. [S_{tot}] is the total biosensor concentration. K_D is the equilibrium dissociation constant, which is assumed equal to EC_{50} , the half-maximal effective concentration of ligand. Method development efforts for these experiments are exhibited in **Figures A.3-5**.

Differential Scanning Fluorimetry

Differential scanning fluorimetry (DSF) [36] corrected for intrinsic biosensor fluorescence was used to determine apparent melting temperatures of iNicSnFR variants. Samples with 50- μ L assay volumes were prepared in triplicate containing 3.5 or 5 μ M (see legend of **Figure 3.4**) purified biosensor (PBS, pH 7.40) with and without nicotine at free ligand concentrations (see **Equation 3.3**) representing twice the respective variant's EC_{50} value. Average fluorescence values for replicates lacking SYPRO[®] Orange dye (Thermo Fisher) were subtracted from replicates of the same variant with $20 \times$ SYPRO[®] Orange dye. Fluorescence values were measured in relative fluorescence units (RFU) at 0.5 $^{\circ}$ C increments from 25 $^{\circ}$ C to 99 $^{\circ}$ C with 30 s holds using a CFX96 Real-Time PCR Detection System (Bio-Rad Laboratories) using the FRET scan mode. Background fluorescence was subtracted from all samples. Apparent melting temperatures were determined from inflection points in melt curves.

Thermal Hold

Biosensor solutions in 1% BSA, PBS, pH 7.40 with protein concentrations of approximately 10 μ M, the concentration used as reported in experiments with iNicSnFR embedded in hydrogels [9], were prepared and assayed for stability under thermal stress. Nicotine dose-response relations and biosensor concentrations were measured (see above) for samples stored at 4 $^{\circ}$ C and 37 $^{\circ}$ C on days zero, one and seven.

Fluorescence Quantum Yield Measurements

Fluorescent quantum yields (Φ) for iNicSnFRs in the liganded, un-snuffed state [7] were determined by comparison with the known value for EGFP. Uncorrected relative fluorescence quantum yield is calculated using **Equation 3.4**, with 'x' denoting the unknown species and 'st' denoting the standard: [37]

$$\Phi_x = \Phi_{st} \cdot \frac{F_x}{F_{st}} \cdot \frac{A_{st}}{A_x} \quad (3.4)$$

The fluorescence quantum yield of the reference standard (Φ_{st}), which was purified EGFP, is 0.60 at $\lambda_{ex} = 489$ nm [38]. F is the integrated fluorescence emission for $\lambda_{ex} = 489$ nm. A is the absorbance at 489 nm. iNicSnFR variants were measured in PBS, pH 7.40 with the free nicotine concentration (**Equation 3.3**) equal to $10 \times EC_{50}$ for each variant. EGFP was measured in PBS, pH 7.4. Background-subtracted absorbance at 489 nm was measured in a 10-mm pathlength cuvette (VWR International) with a Spark[®] 20M instrument (Tecan). Samples were then diluted 20-fold in PBS, pH 7.4 with the appropriate [nicotine] in a clear-bottom 96-well plate (Corning), and fluorescence emission spectra were taken by exciting at 489 nm (20 nm nominal bandwidth) and detecting emissions (1 nm step size; 20 nm bandwidth). Assay volumes for absorbance and fluorescence measurements were 300 μ L and 337 μ L, respectively. Background-subtracted emission spectra were integrated between 528 and 650 nm, with shorter wavelengths omitted to avoid overlap with the excitation.

The Φ values for nicotine-activated iNicSnFR fluorescence required correction by three factors. Factor 1 accounts for extrapolating the integrated fluorescence spectra measured with free nicotine concentrations of $10 \times EC_{50}$ to saturating levels. Factor 2 accounts for the difference in relative integrated absorbance by the iNicSnFRs compared to EGFP using a bandwidth of 20 nm around 489 nm, the peak excitation wavelength for the standard. Factor 3 accounts for the shift in peak emission for the iNicSnFRs compared to EGFP, which results in different fractions of emitted photons falling within the 528 to 700 nm interval. The emission peaks for iNicSnFR12 and iNicSnFR12-D72A were shifted 2 and 3 nm, respectively, vs. that of EGFP. We observed that emission decreased at a rate of e-fold per 31 nm. Therefore, Factor 3 was determined by $e^{-x/31}$, where x is the increase in emission peak wavelength compared to the standard in nm. Quantum yield values as determined by **Equation 3.4** were corrected by the product of these factors as shown in **Table 3.2**.

Table 3.2: Summary of correction factors of fluorescence quantum yield measurements.

Correction Factors	Variant	
	iNicSnFR12	iNicSnFR12-D72A
1	1.082	1.046
2	0.949	0.954
3	0.938	0.908
1×2×3	0.963	0.906

Bibliography

1. WHO global report on trends in prevalence of tobacco use 2000-2024 and projections 2025-2030. Geneva2025.
2. Evaluation IfHMa. Global Burden of Disease Compare. Seattle2023.
3. Henderson BJ, Lester HA. Inside-out neuropharmacology of nicotinic drugs. *Neuropharmacol.* 2015;96(Pt B):178-93. 10.1016/j.neuropharm.2015.01.022
4. Nichols AL, Marotta CB, Lukas H, Friesenhahn NJ, Wagenaar DA, Mayo SL, Dougherty DA, Benowitz NL, Gao W, Muthusamy AK, Lester HA. Continuous Nicotine Monitors for Personal Nicotine Pharmacokinetics: A Receptor-Aware Research Agenda. *Nicotine & Tobacco Research.* 2025;27(9):1499-512. 10.1093/ntr/ntaf062
5. Benowitz NL, Hukkanen J, Jacob P, 3rd. Nicotine chemistry, metabolism, kinetics and biomarkers. *Handb Exp Pharmacol.* 2009(192):29-60. 10.1007/978-3-540-69248-5_2
6. Shivange AV, Borden PM, Muthusamy AK, Nichols AL, Bera K, Bao H, Bishara I, Jeon J, Mulcahy MJ, Cohen B, O'Riordan SL, Kim C, Dougherty DA, Chapman ER, Marvin JS, Looger LL, Lester HA. Determining the pharmacokinetics of nicotinic drugs in the endoplasmic reticulum using biosensors. *J Gen Physiol* 2019;151(4):738-57. 10.1085/jgp.201812201
7. Nichols AL, Blumenfeld Z, Fan C, Luebbert L, Blom AEM, Cohen BN, Marvin JS, Borden PM, Kim CH, Muthusamy AK, Shivange AV, Knox HJ, Campello HR, Wang JH, Dougherty DA, Looger LL, Gallagher T, Rees DC, Lester HA. Fluorescence activation mechanism and imaging of drug permeation with new sensors for smoking-cessation ligands. *eLife.* 2022;11:e74648. 10.7554/eLife.74648
8. Haloi N, Huang S, Nichols AL, Fine EJ, Friesenhahn NJ, Marotta CB, Dougherty DA, Lindahl E, Howard RJ, Mayo SL, Lester HA. Interactive computational and experimental approaches improve the sensitivity of periplasmic binding protein-based nicotine biosensors for measurements in biofluids. *Protein Engineering, Design and Selection.* 2024;37:gzae003. 10.1093/protein/gzae003
9. Nichols AL, Marotta CB, Friesenhahn NJ, Dalfonzo SR, Muthusamy A, Wagenaar DA, Mayo SL, Dougherty DA, Lester HA. Hydrogel Encapsulation of a Designed Fluorescent Protein Biosensor for Continuous Measurements of Sub-100 nM Nicotine. *ACS Applied Bio Materials.* 2026;9(2):765-74. 10.1021/acsabm.5c01418
10. Pédélec J-D, Cabantous S, Tran T, Terwilliger TC, Waldo GS. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology.* 2006;24(1):79-88. 10.1038/nbt1172
11. Treynor TP, Vizcarra CL, Nedelcu D, Mayo SL. Computationally designed libraries of fluorescent proteins evaluated by preservation and diversity of function. *Proc Natl Acad Sci U S A.* 2007;104(1):48-53. 10.1073/pnas.0609647103
12. Sullivan BJ, Nguyen T, Durani V, Mathur D, Rojas S, Thomas M, Syu T, Magliery TJ. Stabilizing Proteins from Sequence Statistics: The Interplay of Conservation and Correlation in Triosephosphate Isomerase Stability. *Journal of Molecular Biology.* 2012;420(4):384-99. 10.1016/j.jmb.2012.04.025

13. Marshall SA, Morgan CS, Mayo SL. Electrostatics significantly affect the stability of designed homeodomain variants¹¹ Edited by J. Thornton. *Journal of Molecular Biology*. 2002;316(1):189-99. 10.1006/jmbi.2001.5326
14. Fukami-Kobayashi K, Tateno Y, Nishikawa K. Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *J Mol Biol*. 1999;286(1):279-90. 10.1006/jmbi.1998.2454
15. Garcia E, Shalaurova I, Matyus SP, Wolak-Dinsmore J, Oskardmay DN, Connelly MA. Quantification of choline in serum and plasma using a clinical nuclear magnetic resonance analyzer. *Clinica Chimica Acta*. 2022;524:106-12. 10.1016/j.cca.2021.11.031
16. Ravva P, Gastonguay MR, Tensfeldt TG, Faessel HM. Population pharmacokinetic analysis of varenicline in adult smokers. *British Journal of Clinical Pharmacology*. 2009;68(5):669-81. 10.1111/j.1365-2125.2009.03520.x
17. Walter S, Hubner B, Hahn U, Schmid FX. Destabilization of a Protein Helix by Electrostatic Interactions. *Journal of Molecular Biology*. 1995;252(1):133-43. 10.1006/jmbi.1995.0480
18. Hol WGJ, van Duijnen PT, Berendsen HJC. The α -helix dipole and the properties of proteins. *Nature*. 1978;273(5662):443-6. 10.1038/273443a0
19. Lockhart DJ, Kim PS. Internal Stark Effect Measurement of the Electric Field at the Amino Terminus of an α Helix. *Science*. 1992;257(5072):947-51. 10.1126/science.1502559
20. Lockhart DJ, Kim PS. Electrostatic Screening of Charge and Dipole Interactions with the Helix Backbone. *Science*. 1993;260(5105):198-202. 10.1126/science.8469972
21. Lyu PC, Liff MI, Marky LA, Kallenbach NR. Side Chain Contributions to the Stability of Alpha-Helical Structure in Peptides. *Science*. 1990;250(4981):669-73. 10.1126/science.2237416
22. Muñoz V, Serrano L. Intrinsic secondary structure propensities of the amino acids, using statistical ϕ - ψ matrices: Comparison with experimental scales. *Proteins: Structure, Function, and Bioinformatics*. 1994;20(4):301-11. 10.1002/prot.340200403
23. Alford RF, Leaver-Fay A, Jeliaskov JR, O'Meara MJ, DiMaio FP, Park H, Shapovalov MV, Renfrew PD, Mulligan VK, Kappel K, Labonte JW, Pacella MS, Bonneau R, Bradley P, Dunbrack RL, Jr., Das R, Baker D, Kuhlman B, Kortemme T, Gray JJ. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of Chemical Theory and Computation*. 2017;13(6):3031-48. 10.1021/acs.jctc.7b00125
24. Mou Y, Huang P-S, Hsu F-C, Huang S-J, Mayo SL. Computational design and experimental verification of a symmetric protein homodimer. *Proceedings of the National Academy of Sciences*. 2015;112(34):10714-9. 10.1073/pnas.1505072112
25. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. 2009;25(11):1422-3. 10.1093/bioinformatics/btp163

26. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *Journal of Molecular Biology*. 1990;215(3):403-10. 10.1016/S0022-2836(05)80360-2
27. Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from the international DNA sequence databases. *Nucleic Acids Research*. 1997;25(1):244-5. 10.1093/nar/25.1.244
28. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Research*. 2013;41(D1):D36-D42. 10.1093/nar/gks1195
29. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*. 2009;6(5):343-5. 10.1038/nmeth.1318
30. Studier FW. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*. 2005;41(1):207-34. 10.1016/j.pep.2005.01.016
31. Chen TW, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA, Orger MB, Jayaraman V, Looger LL, Svoboda K, Kim DS. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*. 2013;499(7458):295-300. 10.1038/nature12354
32. Lee CM, Jin S-P, Doh EJ, Lee DH, Chung JH. Regional Variation of Human Skin Surface Temperature. *Ann Dermatol*. 2019;31(3):349-52.
33. Hill AV. The possible effects of the aggregation of the molecules of hæmoglobin on its dissociation curves. *Journal of Physiology*. 1910;40:iv-vii.
34. Weiss JN. The Hill equation revisited: uses and misuses. *The FASEB Journal*. 1997;11(11):835-41. 10.1096/fasebj.11.11.9285481
35. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Burovski E, Peterson P, Weckesser W, Bright J, van der Walt SJ, Brett M, Wilson J, Millman KJ, Mayorov N, Nelson ARJ, Jones E, Kern R, Larson E, Carey CJ, Polat İ, Feng Y, Moore EW, VanderPlas J, Laxalde D, Perktold J, Cimrman R, Henriksen I, Quintero EA, Harris CR, Archibald AM, Ribeiro AH, Pedregosa F, van Mulbregt P, Vijaykumar A, Bardelli AP, Rothberg A, Hilboll A, Kloeckner A, Scopatz A, Lee A, Rokem A, Woods CN, Fulton C, Masson C, Häggström C, Fitzgerald C, Nicholson DA, Hagen DR, Pasechnik DV, Olivetti E, Martin E, Wieser E, Silva F, Lenders F, Wilhelm F, Young G, Price GA, Ingold G-L, Allen GE, Lee GR, Audren H, Probst I, Dietrich JP, Silterra J, Webber JT, Slavič J, Nothman J, Buchner J, Kulick J, Schönberger JL, de Miranda Cardoso JV, Reimer J, Harrington J, Rodríguez JLC, Nunez-Iglesias J, Kuczynski J, Tritz K, Thoma M, Newville M, Kümmerer M, Bolingbroke M, Tartre M, Pak M, Smith NJ, Nowaczyk N, Shebanov N, Pavlyk O, Brodtkorb PA, Lee P, McGibbon RT, Feldbauer R, Lewis S, Tygier S, Sievert S, Vigna S, Peterson S, More S, Pudlik T, Oshima T, Pingel TJ, Robitaille TP, Spura T, Jones TR, Cera T, Leslie T, Zito T, Krauss T, Upadhyay U, Halchenko YO, Vázquez-Baeza Y, SciPy C. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*. 2020;17(3):261-72. 10.1038/s41592-019-0686-2
36. Niesen FH, Berglund H, Vedadi M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols*. 2007;2(9):2212-21. 10.1038/nprot.2007.321

37. Wurth C, Grabolle M, Pauli J, Spieles M, Resch-Genger U. Relative and absolute determination of fluorescence quantum yields of transparent samples. *Nat Protoc.* 2013;8(8):1535-50. 10.1038/nprot.2013.087
38. Patterson G, Day RN, Piston D. Fluorescent protein spectra. *J Cell Sci.* 2001;114(Pt 5):837-8. 10.1242/jcs.114.5.837

Chapter IV

CONCLUSION

This thesis provides evidence for the utility of developing an iNicSnFR-based continuous nicotine monitor (Chapter 1) as well as research narratives of the development from iNicSnFR3a to iNicSnFR12 (Chapter 2) and from iNicSnFR12 to iNicSnFR12-D72A (Chapter 3). Supplemental material on mutants found by consensus design and by the computational design of α -helices are in Appendix A along with summaries of method development for dose-response experiments. Appendix B includes the development of nanobody-drug conjugates with computationally designed target-binding synergy (CDRxAbs), which I supported experimentally. Computational protein design represents a useful, rational approach to protein engineering, which has enabled the development of CDRxAbs and the optimization of sensitivity for nicotine in the iNicSnFR series.

In future work involving iNicSnFRs, activation by endogenous amines such as choline, which results in an elevated fluorescence baseline, will need to be minimized through selectively permeable materials or additional protein engineering aimed at selectivity for nicotine over other sensor-activating ligands found in biofluids. We look forward to the insights made possible by iNicSnFR12-D72A, from cellular- and organelle-level mechanisms of nicotine addiction to affordable, wearable measuring of personal nicotine pharmacokinetics leading to more effective smoking cessation interventions.

Appendix A

SUPPLEMENTAL INFORMATION ON THE DESIGN AND CHARACTERIZATION OF NICOTINE BIOSENSORS

The work presented herein complements Chapter 3.

Table A.1: Consensus design mutations for positions in iNicSnFR12 with above-average *RE*.

Amino Acid Number	PBP12 amino acid	Consensus amino acid	<i>RE</i>
5	V	I	1.409
15	G	Q	1.401
18	V	L	1.587
43	E	T	1.583
50	I	L	1.398
63	V	P	1.470
68	H	T	2.601
343	W	Y	1.783
391	F	E	1.433
395	G	R	2.577
407	N	G	1.591
420	I	L	1.751
428	N	G	1.410
436	W	Y	2.083
437	A	S	1.655
442	L	I	1.670
445	H	Y	1.476
467	S	R	1.579
474	H	Y	2.114
476	K	E	1.697
486	N	G	1.597
489	S	T	1.547

Table A.2: Single mutants of iNicSnFR12 from α -helix stabilization calculations.

Position and PBP12 amino acid	Amino acids substitutions presumed to stabilize the biosensor
G15	S
E27	K, R, Q, A, I, L, M
G67	D, E, S, N, A
H68	D, E, N, Q, L, W, G
G69	D, S, Q, N, A
D72	K, R, S, N, A, M, F
E394	K, R, N, Q, L, M
D426	K, R, S, T, Q, A, L
K478	E, A, G
D501	K, R, N, I, M
K509	D, E, S, N, Q, A, L
E517	K, R, N, Q, A, L

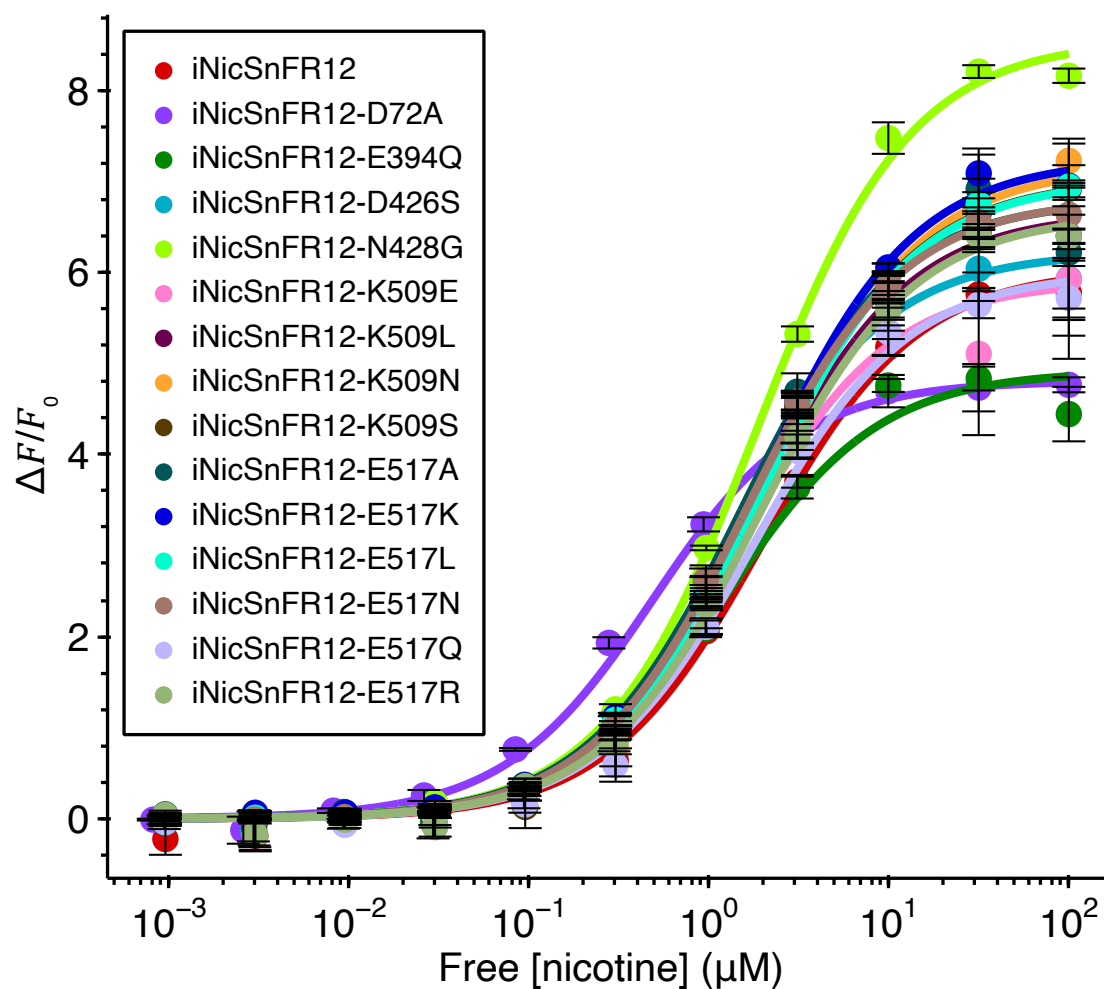


Figure A.1 Dose-response relations of iNicSnFR12 and the most sensitive single mutants from lysate screen. Data are mean $\pm$ SEM (n=3).

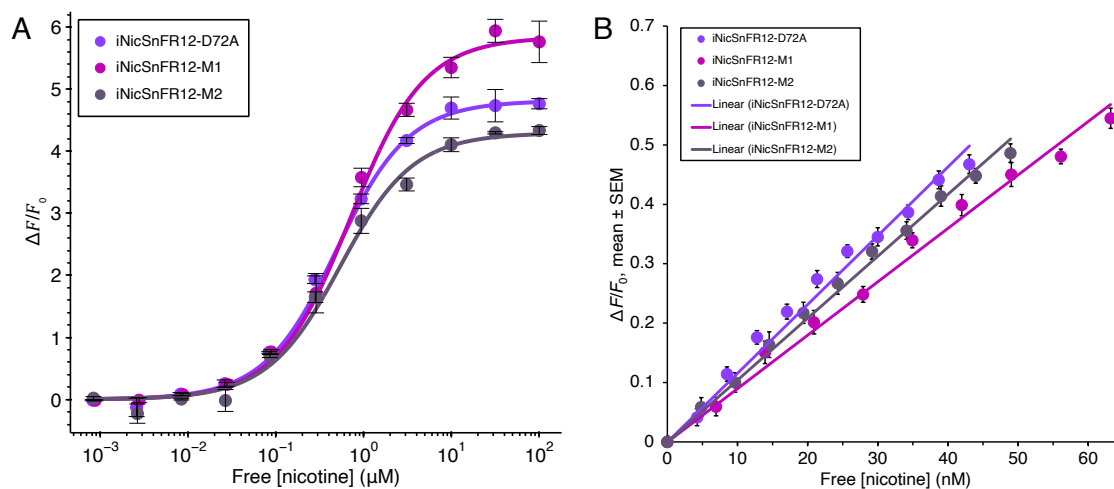


Figure A.2 Dose-response relations of multi-mutants of iNicSnFR12 compared to iNicSnFR12-D72A. (A) Full dose-response relations. The EC_{50} values for iNicSnFR12-D72A, iNicSnFR12-M1, and iNicSnFR12-M2 are $0.44 \pm 0.03 \mu\text{M}$, $0.65 \pm 0.06 \mu\text{M}$, and $0.49 \pm 0.05 \mu\text{M}$, respectively. Data are mean $\pm$ SEM ($n=3$). (B) Dose-response relations in linear regime. The δ -slope values for iNicSnFR12-D72A, iNicSnFR12-M1, and iNicSnFR12-M2 are $11.6 \mu\text{M}^{-1}$, $9.0 \mu\text{M}^{-1}$, and $10.4 \mu\text{M}^{-1}$. Data are mean $\pm$ SEM ($n=8$).

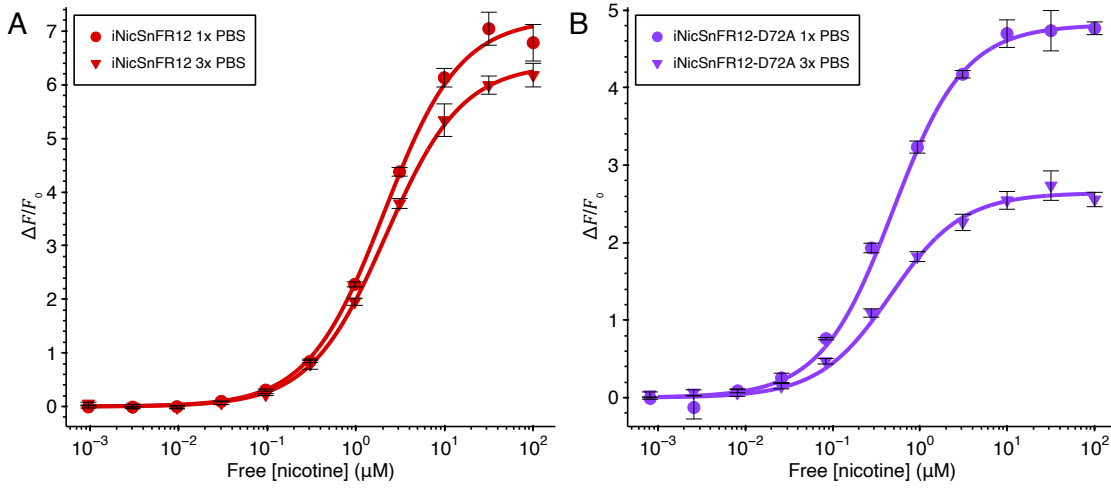


Figure A.3 Comparison of dose-response relations of sensors in different equivalents of PBS buffer. (A) Full dose-response relations of iNicSnFR12 with $1 \times \text{PBS}$ and $3 \times \text{PBS}$. There was no statistical difference in fitted parameters. (B) Full dose-response relations of iNicSnFR12-D72A with $1 \times \text{PBS}$ and $3 \times \text{PBS}$. The fitted $\Delta F_{\text{max}}/F_0$ was statistically different with $p < 0.001$, and the S-slope was statistically different with $p < 0.01$.

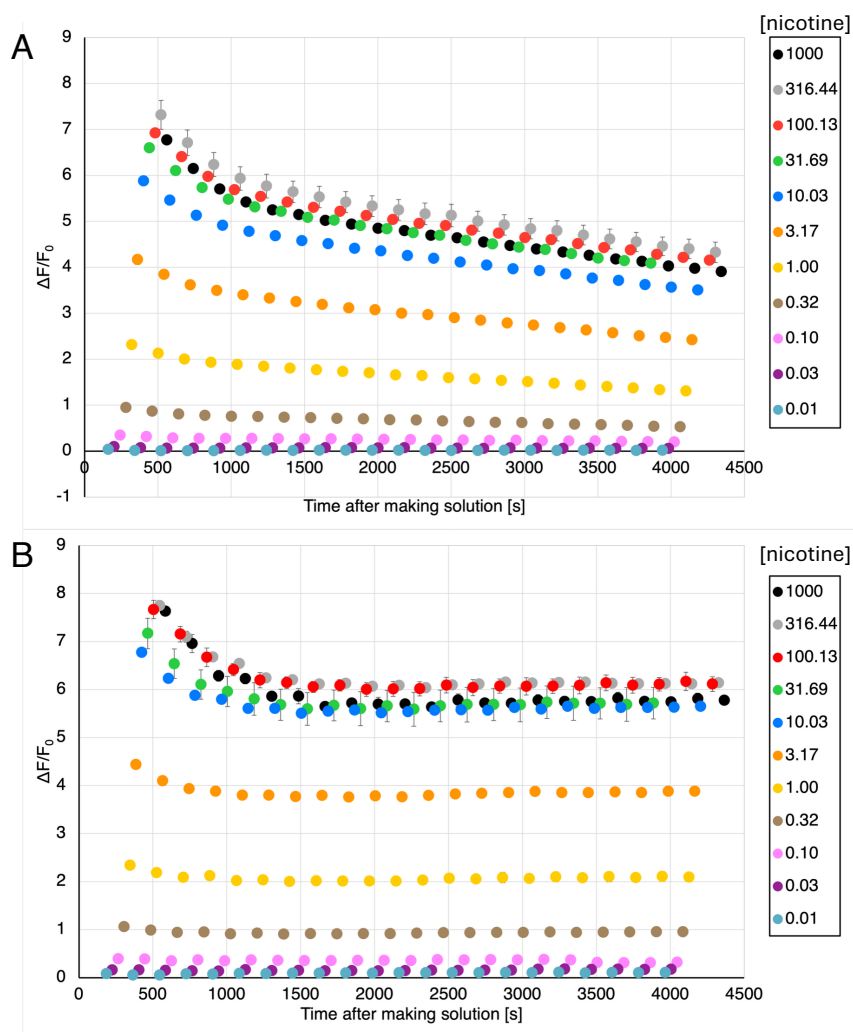


Figure A.4 Dose-response time-course experiments to determine appropriate incubation time and BSA equivalents. Fluorescence scans were taken every 3 min. The plate reader was maintained at 32 °C (A) iNicSnFR12 fluorescence time-course without BSA. (B) iNicSnFR12-D72A fluorescence time-course with BSA.

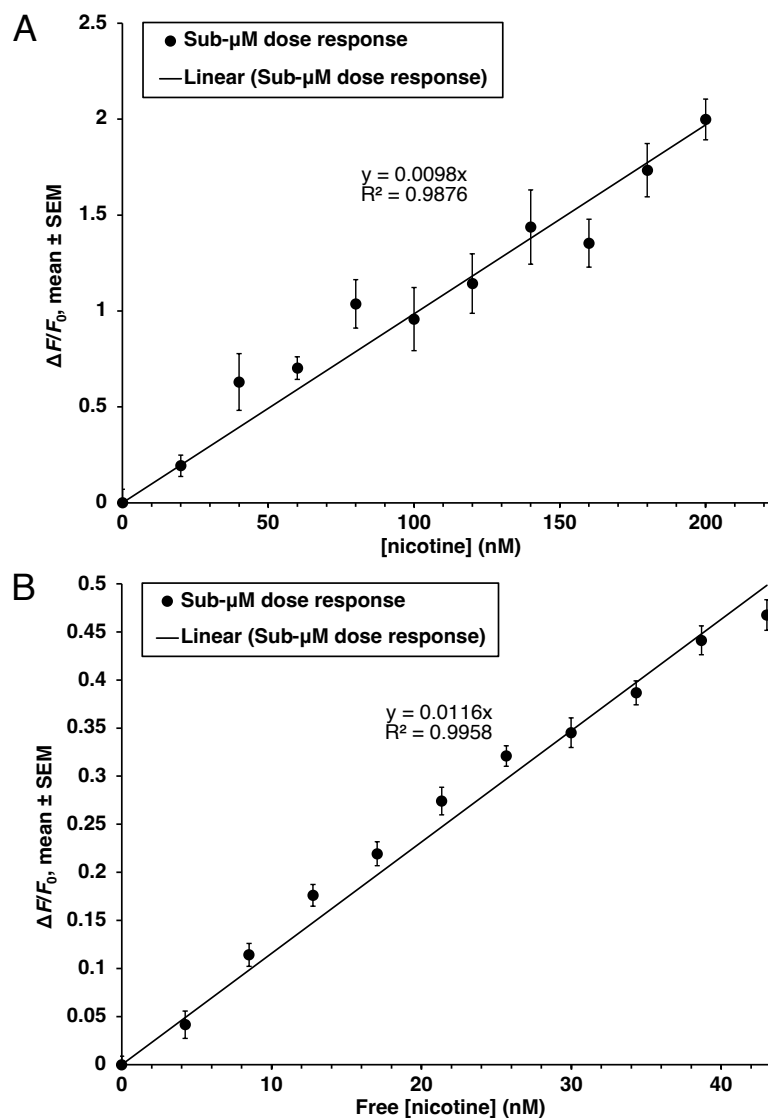


Figure A.5 Reduction in intra-replicate variability in δ -slope experiment by inclusion of BSA in samples and optimization of pipetting by liquid-handling robot. Demonstrated with iNicsnFR12-D72A (A) iNicSnFR12-D72A with solution mixing performed by 50- μ L multichannel pipette directly in microtiter plate. (B) iNicSnFR12-D72A with mixing performed by 300- μ L multichannel pipette in deep-well plate followed by manual transfer by multichannel pipette to microtiter plate.

CDRXABS: ANTIBODY SMALL-MOLECULE CONJUGATES WITH COMPUTATIONALLY DESIGNED TARGET-BINDING SYNERGY

Adapted from:

Wang J, Aceves AJ, Friesenhahn NJ, Mayo SL. CDRxAbs: antibody small-molecule conjugates with computationally designed target-binding synergy. *Protein Engineering, Design and Selection*. 2025;38:gzaf004. 10.1093/protein/gzaf004

Reproduced with permission from Oxford University Press.

Abstract

Bioconjugates as therapeutic modalities combine the advantages and offset the disadvantages of their constituent parts to achieve a refined spectrum of action. We combine the concept of bioconjugation with the full atomic simulation capability of computational protein design to define a new class of molecular recognition agents: CDR-extended antibodies, abbreviated as CDRxAbs. A CDRxAb incorporates a covalently attached small molecule into an antibody/target binding interface using computational protein design to create an antibody small-molecule conjugate that binds tighter to the target of the small molecule than the small molecule would alone. CDRxAbs are also expected to increase the target binding specificity of their associated small molecules. In a proof-of-concept study using monomeric streptavidin/biotin pairs at either a nanomolar or micromolar-level initial affinity, we designed nanobody-biotin conjugates that exhibited >20-fold affinity improvement against their protein targets with stepwise optimization of binding kinetics and overall protein stability. The workflow explored through this process promises a novel approach to optimize small-molecule based therapeutics and to explore new chemical and target space for molecular-recognition agents in general.

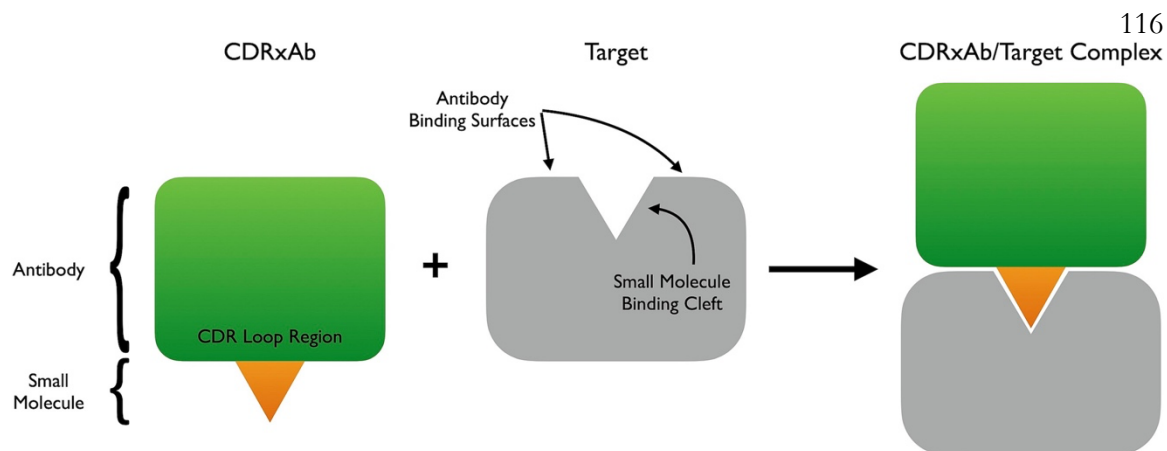


Figure B.1 Graphical Abstract.

Introduction

The mechanisms of action of most pharmaceuticals involve drug-target interactions that are mediated by synthetic small molecules or monoclonal antibodies [1, 2]. Despite impressive successes, many biological pathways are still considered to be ‘undruggable’ because the difficulty of engineering the desired interactions or the pharmacological trade-offs for establishing the interactions outweigh the potential benefits [3-6]. This observation drives creation of new approaches and modalities for addressing challenging targets, which can also expand the targetable molecular space itself [7]. To create new drug modalities, a common approach is to combine existing modalities in order to capitalize on individual advantages and offset individual flaws [7, 8]. Antibody-drug conjugates (ADCs), for example, take advantage of the high specificity and biological compatibility of monoclonal antibodies to improve therapeutic indices of small-molecule drugs [8, 9]. The antibody and small-molecule drug components of ADCs are typically developed separately and bind to different targets during their course of action [8, 9]. Generally, ADCs improve the specificity of their conjugated drugs by delivering the small molecule payloads into target cells through specific antibody-induced receptor endocytosis [9, 10]. Some ADCs and peptide-drug conjugates have also been reported to improve the metabolic stability, circulation half-life, and solubility of linked small molecules through antibody-associated pharmacokinetics, chemical

environment around the conjugation sites, and linker design, indicating that bioconjugation could modulate a wide range of small-molecule properties [10-14].

Successfully engineered hybrid molecules have been reported for various applications. Incorporating reactive moieties in cyclic peptide libraries allowed direct isolation of specific covalent binders [15]. Conjugating kinase inhibitors to a separately developed monoclonal antibody showed improvements in both potency and selectivity [16]. Recently, Cheng *et al.* developed ADCs whose antibody and drug components bind to an identical protein target to achieve synergistic binding/inhibition effects [17]. In their study, the co-crystal structure of a small molecule drug, sitagliptin, a separately developed antibody, 11A19, and the protein target, DPP-IV, was solved first [18]. Based on the structure, optimal conjugation sites and linker sequences were then searched to create ADCs that exhibited 13 to 32-fold IC₅₀ improvement relative to sitagliptin alone against the target [17]. Cheng *et al.*'s work demonstrated that small molecule binding could be improved through conjugating to antibodies, suggesting the potential to expand the chemical space and therefore target space of molecular recognition agents that involve small molecules. However, the approach requires the independent development and characterization of the individual antibody and small molecule components and doesn't incorporate small-molecule binding directly into the antibody/antigen interface, which could be advantageous for certain applications. In order to realize the full potential of this approach, a workflow that can efficiently identify a compatible antibody sequence and conjugation strategy for a desired small molecule binding event would be ideal.

Recently, Lewis *et al.* explored a similar concept of fibronectin small-molecule conjugates, where the fibronectin component was engineered by yeast display in presence of the bioconjugation [19]. The engineered conjugates displayed 9 and 80-fold improvement respectively in potency and selectivity against enzyme targets, further strengthening the hypothesis that a small molecule binding behavior could be directly tuned by a conjugated protein sequence.

In this study, we explored the feasibility of computationally designing the antibody and linker components of synergistically binding ADCs. We introduced the concept of CDR-extended

antibodies (CDRxAbs), which are antibodies whose complementary-determining regions (CDRs) contain a covalently attached small molecule ligand that binds to a desired target with surrounding CDR amino acid sequences tailored to strengthen the target-binding interactions (**Figure B.2A**). For this proof-of-concept study, we focused our design on nanobodies, which are llama-derived single-domain antibody fragments [20-22]. Using a modified streptavidin-biotin interaction pair as model system, we demonstrated that with only the structural knowledge of small-molecule/target interactions, nanobody small-molecule conjugates can be computationally designed to bind a target tighter than the small molecule itself. Through subsequent computationally directed sequence design, the affinity, binding kinetics, and overall stability of the conjugates can be improved in a stepwise manner. Greater than 20-fold affinity improvements together with targeted kinetic-tuning were achieved when the starting small-molecule/target affinity is as weak as 1 μ M or as strong as 7 nM. Exploration of various computational methods revealed key design principles from which we proposed a general design strategy for this novel molecular modality.

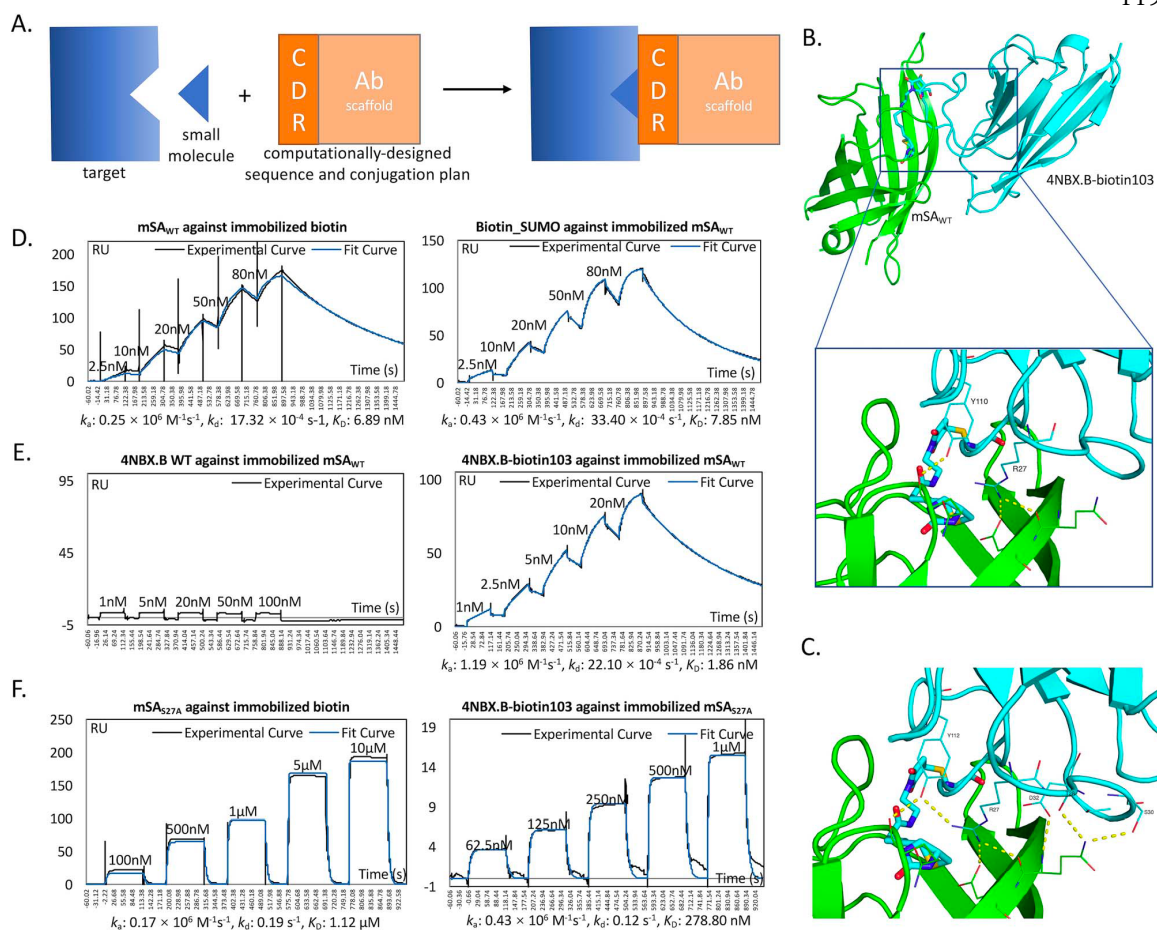


Figure B.2 Computationally designed nanobody-biotin conjugates. (A) Schematic representation of workflow. (B) Final model of 4NBX.B-biotin103 in complex with mSA_{WT} streptavidin. mSA_{WT} shown in green (left side of figures); nanobody shown in cyan (right side of figures); Biotin103 side chain shown as stick. (C) Potential hydrogen bond interactions between nanobody and mSA_{WT}. (D) SPR measurement of mSA_{WT}/biotin binding with immobilized biotin (left) and immobilized mSA_{WT}. (E) SPR measurement of 4NBX.B-WT/mSA_{WT} binding (left) and 4NBX.B-biotin103/mSA_{WT} binding (right). (F) SPR measurement of mSA_{S72A}/biotin (left) and 4NBX.B-biotin103/mSA_{S72A} binding (right).

Results and Discussion

Computationally Designed Nanobody Small-Molecule Conjugation Creates Tighter Binders Against the Small-Molecule Target Protein

We first asked whether computationally determined nanobody sequences and their designed conjugation to a small molecule can exhibit an enhanced binding affinity to the small-molecule target. Designing the antibody components of synergistically-binding ADCs involves creating a new antibody/target interface, which is challenging because of the difficulty in predicting the conformation of antibody CDR loops when bound to a target protein and the general challenge of accurately modeling long structured loops [23-26]. To avoid the challenge of accurately modeling CDR conformations, we adopted an approach similar to anchored-design methods [27]. Anchored-design creates new protein–protein interfaces by first identifying hotspot residues that favorably interact with the target, then designing protein scaffolds to stabilize the anchoring hotspots [27-29]. For synergistically binding ADCs, the conjugated small molecule is defined as a hotspot ‘residue’ that interacts with the target protein. Therefore, to create co-targeting ADCs, the small molecule can be designed as an anchoring noncanonical CDR residue with binding interactions strengthened by additional CDR/target interactions, integrating the drug/target interaction into the antibody/target binding event and biasing the CDRs to adopt the designed conformation.

We divided the design strategy into the following two steps: predicting the optimal CDR binding poses against the target in the small molecule binding region and searching for a conjugation strategy that accommodates both the optimized CDR pose and the target small molecule interaction. As a proof-of-concept we chose monomeric streptavidin as a target and biotin as the small molecule. Streptavidin/biotin interactions have been structurally well characterized. Moreover, monomeric streptavidin variants have been reported with biotin binding affinities reduced by more than 10^5 -fold relative to the tetrameric streptavidin parent [29, 30] which allows for a range of initial starting affinities.

To search optimal CDR binding conformations, we first docked various nanobody scaffolds onto a monomeric streptavidin structure with a computationally modeled triple mutant, S45A/T90A/D180A, that was reported to monomerize streptavidin and reduce the biotin-binding affinity to 1.7 μ M [30]. We then performed loop-modeling for polyalanine CDRs 1, 2 and 3 on 12 docked poses using TRIAD’s loop-modeling module in an attempt to optimize

the CDR conformations against the target surface. Optimized CDR3s were assessed for structural integrity by alignment with CDR3s of the same length from 40 nanobodies with high-resolution crystal structures. Most of the remodeled CDR3 loops showed greater than 6 Å CA RMSD to the reference structures (**Figure B.S8**) which gave us low confidence in the accuracy of the remodeled loops.

In order to constrain the search to sample viable CDR structures, we restricted the search to previously observed nanobody CDR binding conformations [31]. We curated a library of 154 nanobody structures with diverse target-binding CDR conformations from the Protein Data Bank (PDB) and individually docked them onto the target streptavidin surface (**Figure B.S1A**). 2310 docked poses were generated and filtered to identify the most realizable binding conformations, resulting in 7 final binding poses (**Figure B.S1C**). Optimal conjugation strategies were then searched on the finalized poses. We chose to conjugate biotin to nanobody CDRs *via* cysteine-maleimide chemistry, a commonly used conjugation method in ADCs (**Figure B.S2A**) [32]. We wanted to select the shortest, readily available linker possible for this system in order to minimize the number of side chain conformations (*i.e.* rotamers) that needed to be evaluated in the computational modeling and to minimize side chain entropy loss upon binding the target. Biotin C2 maleimide was selected as the conjugation reagent. Optimal nanobody scaffolds and conjugation sites were determined by computationally screening a rotamer library of the cysteine-conjugated side chain at various CDR positions on the nanobody/streptavidin poses. The top-ranked conjugation plan was amino acid position 103 of the 4NBX.B nanobody scaffold (chain B of PDB structure 4NBX), which natively binds to the bacterial toxin TcdA, a target unrelated to streptavidin [33]; the resulting molecule is named 4NBX.B-biotin103. The computationally relaxed complex shows hydrogen bonds between the sidechains of residues Y112 and R27 on 4NBX.B-biotin103 and the backbone of streptavidin. In the native 4NBX.B complex with TcdA, these two residues also participate in intermolecular sidechain/backbone hydrogen bonds, suggesting that the designed pose is closely related to the natural binding mode of 4NBX.B and that these interactions could potentially stabilize the biotin-modified nanobody's interactions with streptavidin. (**Figure B.S2B and C**).

We then synthesized 4NBX.B with site 103 mutated to cysteine and coupled it with biotin C2 maleimide. We attempted to purify and refold the S45A/T90A/D128A mutant of streptavidin to perform binding measurements, but the molecule was insoluble in our hands (data not shown). Therefore, we structurally aligned our triple-mutant streptavidin model and a different previously reported monomeric streptavidin variant, mSA (57% sequence pairwise identity) (Lim *et al.*, 2013). The C-alpha atom RMSD for the alignment was 0.5 Å. We then relaxed the resulting complex of 4NBX.B-biotin103 and mSA (**Figure B.S2D**). The 4NBX.B-biotin103/mSA model preserved the rotamer configuration of the conjugated biotin from the triple-mutant streptavidin model as well as the intermolecular Y112 and R27 hydrogen bonds (**Figure B.S2D**).

Surface plasmon resonance (SPR) binding experiments conducted at 25 °C confirmed that 4NBX.B-biotin103 binds to immobilized mSA with a K_D of 1.8 ± 0.1 nM compared to mSA binding to immobilized biotin with a K_D of 7.0 ± 0.1 nM, indicating a moderate 4-fold affinity improvement that is largely due to a higher k_a (**Figures B.2D–E and B.6A**). Wildtype 4NBX.B did not show binding to immobilized mSA at concentrations up to 100 nM, indicating that the biotin in the 4NBX.B-biotin103 conjugate is required for binding to mSA (**Figure B.2E left panel**). The SPR-measured biotin/mSA affinity is similar to previously published fluorescence polarization spectroscopy data: 2.8 ± 0.5 nM at 4 °C and 5.5 ± 0.2 nM at 37 °C [34]. However, given the lower data-fitting quality of the mSA/biotin binding curves relative to the 4NBX.B-biotin103 binding curves, we performed an alternative determination of the mSA/biotin affinity by binding Smt3 SUMO protein biotinylated at the N-terminus with biotin C2 maleimide to immobilized mSA. Smt3 SUMO protein has an unstructured N-terminus that we hypothesized would minimize the interaction between the protein components [35]. The data quality was improved and the resulting K_D , 9.6 ± 1.2 nM, is similar to what we determined above (**Figure B.2D right panel**).

To determine whether computationally designed nanobody conjugation shows improved affinity when starting with weakly-binding small molecules, we created a single mutation of mSA, mSA_{S27A}, whose counterpart S45A in wild type streptavidin reduces biotin-binding

strength and was predicted by molecular dynamics (MD) simulation to minimally affect the overall structure [36]. Size-exclusion chromatography (SEC) shows that mSA_{S27A} elutes at the same time as mSA_{WT} (**Figure B.S3A**). SPR measurements show that mSA_{S27A} binds to biotin with a K_D of $1.14 \pm 0.02 \mu\text{M}$, while 4NBX.B-biotin103 binds to mSA_{S27A} with a K_D of $245 \pm 41 \text{ nM}$, indicating a similarly moderate 5-fold improvement in binding affinity (**Figures B.2F and 6B**). Together, the above results demonstrate that a purely computational method based solely on structural information of the individual components can be used to design CDRxAbs that exhibit affinity-enhancing synergistic binding effects.

Sequence Design Improves the Binding Affinity and Kinetics for Computationally Designed Conjugates

Next, we performed computational sequence design on the CDR loops of 4NBX.B-biotin103 with the goal of improving its binding affinity to mSA. We performed *in silico* site-saturation mutagenesis on each CDR position in the complex to identify positions tolerant to mutation, resulting in a set of 9 positions (SI Appendix). Computational sequence designs using combinations of four identified CDR positions each were tested in parallel. The amino acids allowed for each position in these calculations were set based on the known sequence diversity of nanobody CDR loops [37]. Analysis of the design output revealed that designs with positions 31, 32, 104, and 105 most frequently returned sequences that were predicted to form additional intermolecular hydrogen bonds and that were also energetically stable. The top-ranked variant by energy, v119, has CDR1 mutations M31H and D32A, and CDR3 mutations N104S and W105H. New hydrogen bonds were predicted between H31 and Q108 of mSA and between H105 and E105 of mSA (**Table B.S2, Figure B.3A**). The D32A mutation eliminates a buried, unpaired charged residue. SPR measurements show that the v119 variant binds mSA_{WT} with a K_D of $0.9 \pm 0.2 \text{ nM}$, indicating an ~ 2 -fold improvement from 4NBX.B-biotin103_{WT} (**Figures B.3A and B.6A**). The K_D improvement was primarily due to an increase in k_a (**Figures B.3A and B.6A**). With the goal of reducing k_d , we selected variant v149, which has the largest number of predicted hydrogen bonds, 15, of the top 20 ranked sequences. The v149 variant has mutations M31R, D32S, N104A, and W105R that

were predicted to form extensive interactions with Y96, E105, and Q108 of mSA, including a potential salt bridge between nanobody R105 and mSA E105 (**Figure B.3B**). Interestingly, R-E interactions appear to be frequently used by nanobodies [38]. Compared to v119, SPR measurements for v149 show an ~ 2 -fold lower k_d and an ~ 4 -fold higher k_a that together contribute to a K_D of 0.12 ± 0.01 nM, a > 20 -fold K_D improvement compared to biotin/mSA_{WT} affinity (**Figure B.3B and B.5A**). However, SEC analysis of v149 shows signs of aggregation (**Figure B.4A**).

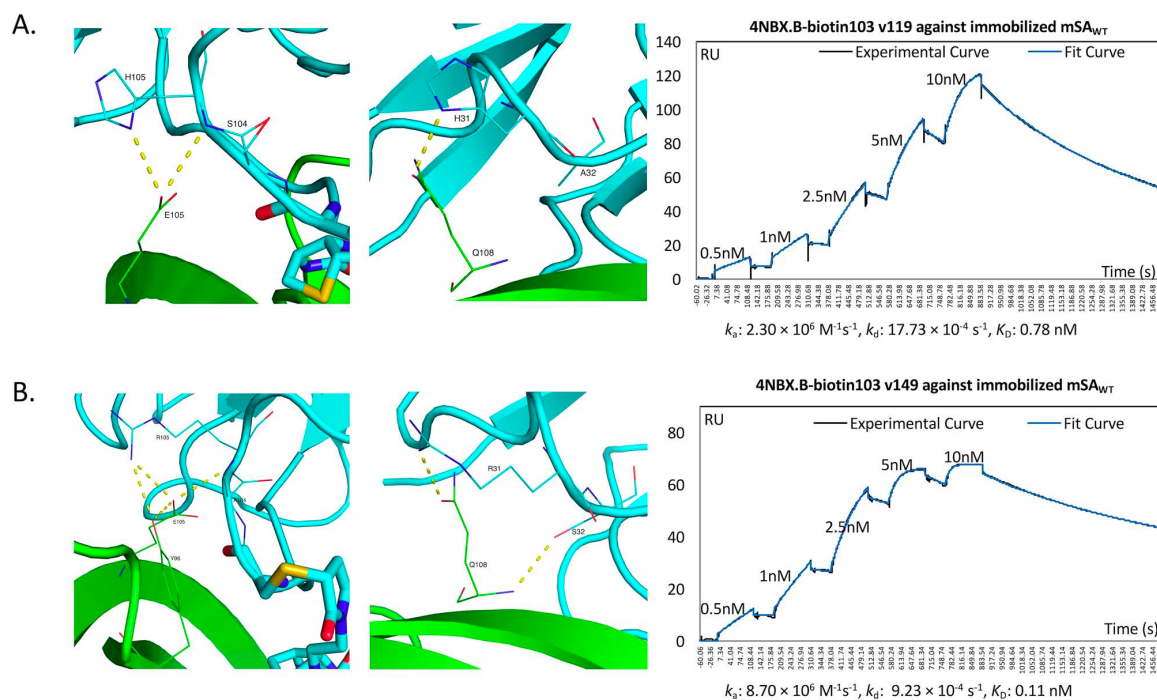


Figure B.3 Design models and SPR binding for CDR sequence designs v119 and v149. (A) v119 design showing potential inter-protein hydrogen bonds (left and middle -- mSA_{WT}, green (bottom of figures); v119, cyan (top of figures); Biotin103 thick sticks) and SPR measurement of v119/mSA_{WT} binding (right). (B) v149 design showing potential inter-protein hydrogen bonds (left and middle) and SPR measurement of v149/mSA_{WT} binding (right).

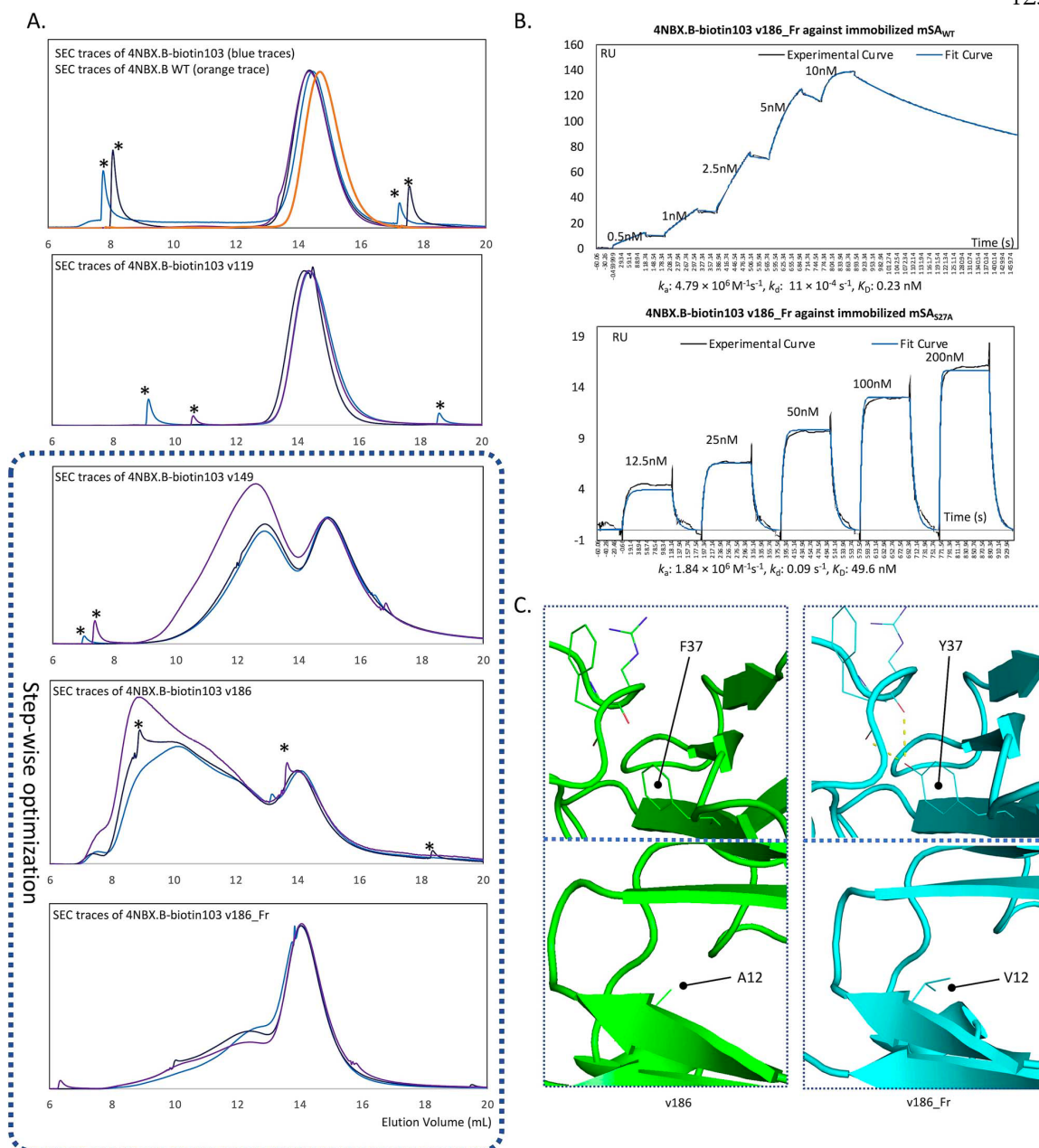


Figure B.4 Framework design for reduced aggregation. (A) SEC traces normalized by monomer peak height of 4NBX.B-biotin103 (biological triplicates – blue traces) and 4NBX.B WT (orange trace) (top panel), v119 (second panel), v149 (third panel), v186 (fourth panel) and v186_Fr (bottom panel). ‘*’ indicates instrument artifact. See **Figure B.S3** for additional SEC details. (B) SPR measurement of v186_Fr/mSA_{WT} binding (top panel) and v186_Fr/mSA_{S27A} binding (bottom panel). (C) Structural models showing sequence

changes between v186 (green, left side of figure) and v186_Fr (cyan, right side of figure). Top panels show F37Y mutation in v186_Fr and potential hydrogen bonds for the introduced tyrosine. Bottom panels show A12V mutation.

Sequence Design Reduces Aggregation While Preserving Binding Affinity for the Designed Conjugates

4NBX.B-biotin 103 WT and v119 are well behaved by SEC with single peaks eluting at roughly the same time as the wild type 4NBX nanobody (**Figure B.4A**). The apparent reduced stability of v149 is consistent with its predicted higher energy score compared to v119 (**Table B.S2**). To improve the stability of v149, we explored further CDR optimizations that would better accommodate the biotin103 side chain in order to stabilize the loop and overall structure. We performed two additional rounds of CDR residue mutability analysis followed by in-parallel combinatorial designs on v149 until no further CDR mutations were predicted to be energetically favorable. Mutations accumulated in previous rounds of design were kept in subsequent rounds. In the top 20 sequences ranked by the energy score of both additional rounds of design, no additional hydrogen bonds were predicted to form with mSA, so the design with the best energy improvement, v186, was selected. The v186 variant has 6 additional CDR mutations on top of those in v149: Y101L, R107F, R56T, Y106K, D108A, and Y110S. The hydrogen bonds predicted in the v149 variant are preserved in v186 and v186 appears to bind to mSA_{WT} with a similar K_D as v149 (**Figure B.S5**). However, SEC analysis of v186 shows even more aggregation than v149 (**Figure B.4A**).

Molecular dynamics (MD) simulations have been successfully applied to reveal the source of unexpected or undesired properties in designed proteins [39]. In order to explore the flaws of the current designs and inform a design strategy going forward, we perform MD simulation of 4NBX.B-biotin103 v186 in complex with mSA_{WT}. The simulations showed that the CDR3 loop separates from the β -barrel framework region (Supplementary Movie in original publication) suggesting that optimizing the interactions of CDR3 with the framework region could lead to improved designs. We performed framework sequence design on v186. The top-ranked variant, v186_Fr, was predicted to form additional hydrogen

bonds with CDR3 residues through the F37Y mutation (**Figure B.4C, Top**). In addition, the A12V mutation shows increased hydrophobic shielding of the β -barrel core and improved β -strand propensity relative to alanine (**Figure B.4C, Bottom**). Interestingly, when the same framework sequence design was performed on v149, the parent of v186, the A12V and F37Y mutations were predicted to destabilize v149, suggesting that the v186 mutations are required to realize the benefit of the A12V and F37Y mutations (**Tables B.S2–S4**). **Table S5** summarizes the energy scores for proteins experimentally evaluated in this study.

4NBX.B-biotin103 v186_Fr showed significantly reduced aggregation by SEC. Collected main peak fractions did not show signs of re-aggregation when reanalyzed by SEC (**Figures B.4A and B.S3B**). SPR measurements showed the K_D of v186_Fr to be 0.20 ± 0.03 nM, preserving the >20-fold K_D improvement observed with v149 over biotin/mSA_{WT} (**Figure B.4B top panel, B.6A**). The kinetics profile of v186_Fr binding to mSA_{WT} was also similar to v149 (**Figure B.4B top panel, B.6A**). When binding to mSA_{S27A}, v186_Fr showed a K_D of 54 ± 3 nM, indicating an ~20-fold improvement resulting from both improved association and dissociate rates (**Figure B.4B bottom panel, B.6B**).

To further investigate the structural features of v186 and v186_Fr, we performed additional 100 ns MD simulations of v186 and v186_Fr bound to mSA_{WT}. During the simulations both the overall geometry of the complexes and the conformation of the biotin103 side chain remained constant with CA backbone RMSD of less than 1.5 Å (**Figure B.S4 and B.5B first panel**). The 4NBX.B nanobody scaffold has two notable solvent-inaccessible clusters of hydrophobic residues one located in the β -barrel core and the other shielded from solvent by the CDR3 loop (**Figure B.5A**). Buried hydrophobic surface area is generally correlated with protein folding stability, while hydrophobic surface patches and the exposure of buried hydrophobic residues are correlated with loss of stability and aggregation [9, 40]. The MD simulations show reduced solvent-accessible surface area exposure for the two clusters of hydrophobic residues for v186_Fr relative to v186 supporting the idea that the F37Y and A12V mutations in v186_Fr reduce aggregation by stabilizing the framework region and reducing the exposure of hydrophobic surface area (**Figure B.5A**).

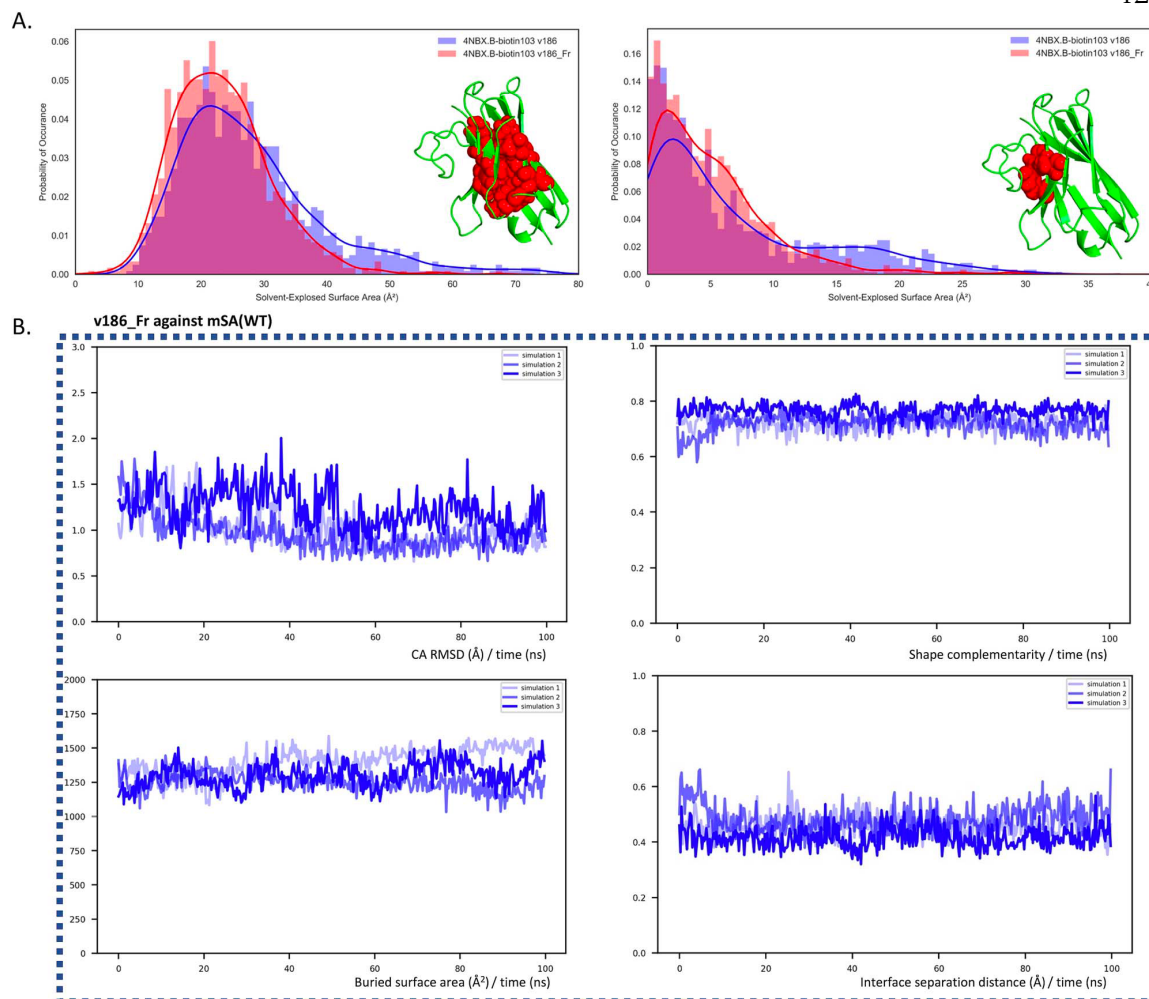


Figure B.5 MD simulation analysis of various models. (A) Solvent accessible surface area distribution from 100 ns MD simulations for two hydrophobic residue clusters shown as red spheres in the left and right panels. Data for v186 shown in blue; data for v186_Fr shown in red. **(B)** Analysis of CA backbone RMSD to average structure (upper left panel), interface shape complementarity (upper right panel), interface buried surface area (lower left panel) and interface separation distance (lower right panel) from 100 ns MD simulations shown in triplicate for the v186_Fr/mSA_{WT} complex. All MD simulation frames were used in the analysis.

To ensure that the prepared conjugates contain a single biotin-maleimide conjugation per nanobody molecule, we used intact-protein mass spectrometry (MS) to analyze a set of

variants. The most abundant species from deconvolution of the mass spectra had molecular weights within 3 Da of the expected values of mono-biotin conjugation (**Figure B.S6**). Subpopulations within ± 20 Da of the expected values were also observed, which could result from succinimide ring opening, ion adduct formation, pyroglutamate formation or other processes (**Figure B.S6**). These subpopulations could have been present in protein preparations before samples were subjected to mass spec analysis or could have arisen as part of the mass spec analysis itself or some combination of the two. In the first and third scenarios, the amount of active material could have been less than 100% of the measured protein quantity which could lead to inaccuracies in the SPR analysis.

To address a question of whether the final design, v186_Fr, would still bind TcdA, the native target of the parent nanobody scaffold, we examined the interface between the parent nanobody and TcdA in PDB structure 4NBX.B (chain B of PDB structure 4NBX). The site of biotin conjugation, A103 (A110 in 4NBX chain B PDB numbering – see **Table B.S6** for amino acid sequences for the protein templates used in this study), is directly in the interface between the nanobody and TcdA. A103 makes extensive van der Waals contacts with TcdA. Mutating the alanine at this position to cysteine and coupling it to biotin-C2-maleimide introduces considerable steric bulk at this site. Although we did not test for binding of v186_Fr to TcdA, modeling analysis suggests that no binding is to be expected.

Designing Directly Against mSA Yields Tight Binding Nanobody Small-Molecule Conjugates

Because the 4NBX.B nanobody scaffold described above was not obtained by directly docking nanobody scaffolds against mSA, we performed the docking, filtering, and rotamer screening steps using mSA and selected scaffold 2X89.A with biotin conjugated to position 57 (**Figure B.S7A** and **B top panel**). Similar to the 4NBX.B-biotin103 variants, the selected pose of 2X89.A was predicted to interact with mSA through a R-E interaction (in addition to other potential intermolecular H-bonds) (**Figure B.S7B top panel**). In order to ensure a single conjugation site, two cysteines at positions 33 and 104 that form an intra-CDR disulfide bond in 2X89.A were replaced by alanine residues. The initial conjugate, 2X89.A-CCAA-biotin57 binds to mSA_{WT} with a K_D of 0.8 ± 0.2 nM (**Figures**

B.6A and **B.S7B bottom panel**). 2X89.A-CCAA-biotin57 was shown to aggregate by SEC analysis (**Figure B.S7C**). To reduce aggregation, we constructed a sequence design pipeline that alternates between CDR and framework sequence optimization based on our previous experience designing 4NBX.B-biotin103 variants and applied the pipeline to 2X89.A-CCAA-biotin57 (**Figure B.S7D**). Six rounds of CDR sequence design and one round of framework sequence design resulted first in worsening and then ultimately improving the aggregation profile after 18 mutations were accumulated (**Figure B.S7E**), a trend similar to what we observed in the optimization of 4NBX.B-biotin103. The resulting K_D of 2.8 nM is reduced relative to the initial design, suggesting a possible tradeoff between improved protein stability/aggregation and binding affinity.

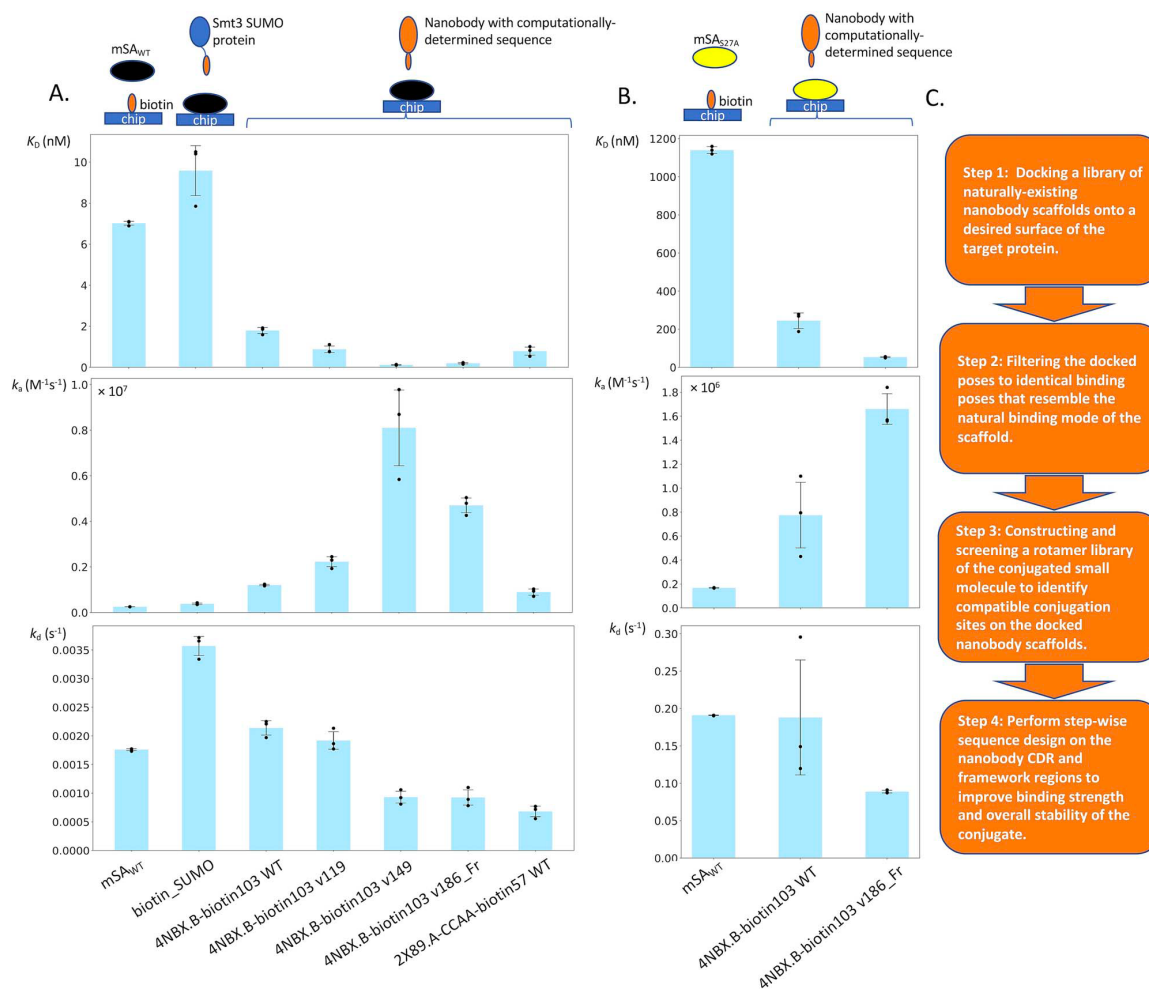


Figure B.6 Summary of results and design workflow. (A and B) Summary of SPR determined binding affinities and kinetic parameters for the various complexes studied in this report. Setups for the SPR binding experiments are depicted by the cartoon above the lanes. Individual data points are from biological triplicates. Error bars in figures show standard deviation. (C) Summary of design workflow for creating synergistically binding nanobody-small molecule conjugates.

Conclusions

We developed and validated a structure-based computational protein design workflow that relies on pre-existing ligand-receptor structures to engineer ligand-antibody hybrids with synergistic binding behavior. We believe this workflow will contribute to the expanding computational toolbox that optimizes protein-engaging small molecules for improved functions [41]. Using mSA/biotin as a model system, we showed that biotin/nanobody conjugates bind tighter to the mSA target than biotin alone. Our results showed that the affinity, kinetics and stability of the conjugates can be computationally designed and improved in a stepwise manner, which allows for a highly tunable optimization process. This approach allows the target-binding properties of the small molecule to be tuned through changes in the amino acid sequence of the conjugated antibody. The binding interface for the designed conjugates leverages the deep binding pockets common for small molecules and the broad contact interface common for antibodies [42, 43]. Combining these binding characteristics into a single molecular entity may offer solutions to challenges such as off-target binding of small molecule drugs and tackling undruggable targets in pharmaceutical development.

Testing the design strategy introduced in this study on therapeutically relevant targets is an obvious next step. In addition, examining the use of virtually docked small-molecule/target complexes and explicitly testing for improvement of binding specificity promise to be interesting avenues for future research. There are a number of computational approaches that could be used to improve the overall design strategy. Virtually recombining structural fragments was reported to improve affinity maturation of computationally designed

antibodies [28, 44, 45]. Algorithms that explicitly bias for the formation of hydrogen-bonding networks were also proven to be useful for improving the binding affinity and specificity of designed protein/protein interfaces [46]. Advanced loop-modeling methods and ensemble design could also facilitate a more accurate assessment of binding poses for the conjugates and potentially improve the engineering of binding specificity [47, 48].

Methods

Computational Design Workflow for Nanobody-Biotin Conjugates

The protein design software package TRIAD (Protabit LLC) together with PyMOL (Schrodinger) and OpenBabel [49] were used to perform computational protein design and analysis in this study. The detailed process and setup are described below.

A. Design process of 4NBX.B-derived nanobody-biotin conjugates

Searching optimal streptavidin-binding nanobody CDR conformations by docking and loop modeling

A monomeric streptavidin model, S45A/T90A/D180A, was prepared from the crystal structure of wild-type tetrameric streptavidin (PDB ID: 1MK5). A single subunit was extracted and standardized using TRIAD. The S45A/T90A/D128A mutations were then introduced using the TRIAD sequence-design module. Initially, we attempted optimizing nanobody/streptavidin binding conformations by protein-protein docking followed by loop modeling of the nanobody CDRs. We used a published nanobody structure (PDB ID: 5VNW, chain C) as the starting nanobody scaffold with all CDR residues replaced by alanine using TRIAD sequence design in a hope to avoid sequence bias [50]. The nanobody scaffold was docked onto a set of manually selected surface residues surrounding the binding pocket of the monomeric streptavidin model. Docking was performed using an FFT-based docking algorithm and the top 15 CDR binding poses were kept [51]. CDR loop modeling of each pose was then performed to attempt optimizing binding to the target using the TRIAD loop modeling module. Only the top solution for each pose was kept. Based on visual analysis, no

solutions had reasonable CDR conformations that would be beneficial for binding. The sequence of the monomeric streptavidin used for these calculations is listed below with the residues selected as docking targets surrounded by brackets:

2EAGITGTWY(NQLGS)TFIVTAGADGALTGT(YEAAVGNAESRY)VLTGRYDSAPA
TDGSGTALGWTVA(WKNNYRNAHSA)ATWSGQYVGGAEARINTQ(WLLTSGTTE
ANAWKSTLVGHATFT)KVK.

Searching optimal streptavidin-binding nanobody CDR conformations by docking native CDR sequences and conformations

154 nanobodies crystal structures with resolutions better than 3 Å, complete electron density, and structurally diverse target-binding CDR conformations were assembled from published nanobody/target complexes from the PDB (see section C below). The TRIAD surface-complementarity module was used to perform interface analysis on the 154 nanobody-target complex structures to collect statistics on separation distance, shape complementary, and buried interface area for use in evaluating designs (**Figure B.S1B**) [52]. CDRs for each nanobody were annotated following established nanobody CDR-mapping criteria [37] but with softer edge cutoffs so that the sampled geometries of nanobody approaching are not too stringent.

Each nanobody was docked to the previously selected binding pocket surface (defined above) using the monomeric streptavidin model and the annotated CDR loops. The top 15 poses for each docking trial were kept. Initially, we performed alanine-replacement on all CDR residues before docking. However, several CDR sequence design trials using the dock complexes resulted in the selection of mostly small amino acids, suggesting that the poly-alanine CDRs resulted in streptavidin/nanobody distances that were too close. Therefore, we decided to dock the 154 nanobody structures using their native CDR sequences. 2310 docked poses were generated and then filtered using three selection steps to identify the most native-like binding poses. Step one selected poses with interface separation distance, shape complementary score, and buried interface area within 1 standard deviation of the values

obtained for the 154 nanobody/target structure dataset (**Figure B.S1B**). This filter resulted in 231 poses. Step two then selected nanobody poses where the nanobody/streptavidin interface utilized >80% of the nanobody residues that participate in the native nanobody/target interface, where the interface residues were determined using a PyMOL script that evaluates changes in solvent accessibility between the docked and undocked structures (PyMOL Wiki). This filter resulted in 31 poses. The last step selected poses that directly blocked the biotin binding pocket and resulted in 7 final poses (**Figure B.S1C**). Binding pocket blockage was assessed using PyMOL to calculate the change in solvent accessible area between the bound and unbound complex for the following selected residues that are a subset of the docking target residues and closely surround the biotin molecule (highlighted in parentheses along with the full streptavidin sequence):

EAGITGTWYN(QLGS)TFIVTAGADGALTGTYEAA(VGNA)ESR(Y)VLTGRYDSAP
ATDGSGTALGWTVA(W)KNNYRN(AHS)AATWSGQYVGGAEARINTQ(W)L(L)T(S)
)GTTEANAWKSTLV(G)H(A)T(FT)KVK.

Rotamer library generation for biotin C2 maleimide side chain and conjugation plan determination

A rotamer library for the biotin-CH₂-CH₂-succinimide-S-CH₃ “side chain” was constructed using OpenBabel [49] and virtually screened against CDR amino acid locations on the 7 poses from above to find the optimal conjugation sites. In the rotamer library, the biotin portion remains fixed, while the CH₂-CH₂-succinimide-S-CH₃ portion is structurally diversified through torsion angle modifications. Screening was done by measuring the distance between the terminal carbon of the rotamer and the C β of respective CDR residues, steric clash between the rotamer and streptavidin, steric clash between the rotamer and the nanobody, and the angle of the rotamer’s terminal carbon with respect to the target C α attachment point on the nanobody. Rotamers that were <1 Å from the target C β , clashed with streptavidin by <1 unit, clashed with the nanobody by <15 units, and had attachment point angles between 100-120 degrees were kept for further evaluation. Measurements were performed only for CDR residues that are originally alanine, which is similar in size to

cysteine, as we hypothesized that making an alanine to cysteine mutation would be less likely to cause serious structural consequences for the nanobody. Alanine 103 on 4NBX.B was the only conjugation site that passed all of the filters, and the rotamer that clashed least with both streptavidin and the nanobody was selected for further processing. To prepare the final conjugation structure, excess atoms were removed and a bond was made between the C β of the biotin-CH₂-4CH₂-succinimide-S-CH₃ “side chain” and the C α of 4NBX.B at position 103. The conjugated 4NBX.B-biotin103 structure bound to S45A/T90A/D128A streptavidin was then relaxed using BIOGRAF [53] with force restraints placed to maintain the biotin-streptavidin hydrogen bonds and torsion angles of the aliphatic arm portion of biotin. To prepare the 4NBX.B-biotin103/mSA structure, the crystal structure of mSA (PDB ID: 4JNJ) was aligned to the modeled triple-mutation streptavidin structure, and the resulting complex with 4NBX.B-biotin103 was relaxed using BIOGRAF with the same force restraints as above [53].

4NBX.B-biotin103 sequence designs

CDR sequence design was first performed on the 4NBX.B-biotin103/mSA model. CDR residues were determined following established nanobody CDR-mapping criteria [37], and specifically by first aligning the 4NBX.B model with an example nanobody structure in agreement with that criteria (PDB ID: 5VNW, chain C), and then selecting the corresponding CDR residues on 4NBX.B. Residues 27-34 were selected as CDR1, 47-60 were selected as CDR2, and 98-111 were selected as CDR3. Single residue site-saturation mutation scans were performed on each CDR position using reduced sets of amino acids that were reported to be frequently used in each corresponding CDR position [37]. The wild-type amino acid was considered for each position. The amino acid sets for each position were as follows:

Table B.S1: 4NBX.B-biotin103 amino acid sets for site-saturation mutation scans.

Positions	Amino Acids
27	N, S, T, Y
28, 51	I
29	S, F
34	M

47	F, L
48	V
49, 60, 98	A
50, 55	A, G, S, T
52	A, D, G, Q, S, T
54	G
56	I, N, S, T
57, 58	T
59	N, Y
60, 111	Y
109	F, H, L, Y
30-33, 53, 99-102, 104-108, 110	A, R, N, D, Q, E, G, H, I, L, K, F, P, S, T, W, Y, V

Position 103 with the attached biotin “side chain” was left un-designed and the coordinates for all atoms were left unchanged. The Rosetta force field [54] option with covalent terms was used in TRIAD for design the calculations. The BIOGRAF-relaxed 4NBX.B-biotin103/mSA structure was used as the input design model. During the design calculations, residues that were within 10 Å from the position being evaluated were allowed to repack. Each rotamer optimization for the position being evaluated was initiated by random rotamer configurations and then repacked while the C α backbone was allowed to relax through Cartesian minimization to optimize the structure for different amino acid choices, which were then ranked given the energy scores of the corresponding modeled structures after iterative rotamer repacking and backbone relaxation. The chemical attributes of the biotin103 “side chain” were generated by TRIAD and then used during the design calculations of the energy scores. 10 runs with different random seeds were performed for each design calculation and averaged to reflect the final amino acids preference for each site that underwent single-site mutation calculations. Mutation choices with lower Rosetta energy unit than the WT amino acids were kept as designable mutations. Designability of each site was reflected by the sum of Rosetta energy unit differences of designable mutations from the corresponding WT amino acid choice. As the result, the output designable sites ranked by designability were as follows: 105, 109, 107, 104, 106, 32, 108, 31, and 56. Those sites were also separately grouped into two bins. Bin 1 contains sites that interact with the original target of 4NBX.B: sites 105, 104, 32, and 31, of which the order was ranked by designability. Bin

2 contains sites that do not interact with the original target of 4NBX.B: sites 109, 107, 106, 108, and 56, of which the order was ranked by designability.

Combinatorial designs with different choices of designable sites were performed in parallel.

Combinatorial design 1 was performed on all the 9 designable sites. Combinatorial design 2 was performed on the designable sites in bin 1 only. Combinatorial design 3 was performed on the designable sites in bin 1 and bin 2, with an exception that for bin 2 sites, only mutation choices that are different from WT amino acid with >1 Rosetta energy units were used. Combinatorial design 4 was performed on the designable sites in bin 1 and bin 2, with an exception that for bin 2 sites, only the top-ranked mutation choice by energy score was used. Combinatorial designs were performed with the same configurations as single-site mutation designs that were described before, with one difference: the output sequences from the 10 parallel design runs were re-ranked by threading the sampled sequences in each run individually onto the backbone of the input structure, followed by rotamer repacking and backbone Cartesian minimization. The TRIAD-modeled structures and Rosetta energy scores of the top 20 sequences of the re-ranked sequences were used to evaluate design results.

Structures of the top 20 sequences for the 4 combinatorial designs were analyzed by PyMOL to identify intermolecular H-bonds between mSA_{WT} and 4NBX.B-biotin103 variants, and intramolecular H-bonds within 4NBX.B-biotin103 variants, using a publicly available PyMOL script that relies on the “find_pairs” command module of PyMOL [55]. The goal was to find sequences with improved overall energy score, new intermolecular H-bonds with mSA_{WT}, and intramolecular H-bond profile comparable to 4NBX.B-biotin103 WT, as the imbalance of forming new interactions with targets and keeping the structural integrity was a common reason behind the failure of designing protein-protein interactions [56]. All combinatorial designs output sequences with improved energy scores, but only combinatorial design 2 output the top 20 sequences with an average number of intermolecular H-bond

higher than that of the 4NBX.B-biotin103 WT against mSA_{WT}. The top 20 sequences in combinatorial design 2 also had the highest average number of intramolecular H-bonds in the nanobodies among the 4 designs. Variant v119 was the top-ranked sequence in combinatorial design 2, and variant v149 had the highest number of predicted intermolecular H-bonds among the top 20 sequences (**Table B.S1**).

To improve the stability of v149, we hypothesized that a suitable method would optimize the protein structure while keeping the designed interactions contributed by R31/S32/A104/R105, the new anchoring spots, unchanged, without altering the target backbone structure too much. Therefore, we devised a sequential design workflow that creates stepwise local structural optimizations that compensate for the mutations built-up in previous steps. As the result, subsequent rounds of CDR sequence designs were performed on v149. For each round, single-point mutation scan was first performed on the CDR residues using the identical setup as the first round of design. The calculation results were processed in the same way as the first round of design, with an exception that only mutation choices that are different from WT amino acid with >1 Rosetta energy units were kept for all further combinatorial design calculations. Next, skipping the sites that were mutated in previous rounds of design, four combinatorial designs were performed on the designable sites reported by the single-point mutation scan calculation. Combinatorial design 1 was performed on all designable sites with the reported designable mutation choices. Combinatorial design 2 was performed only the designable sites in bin 1 with the reported designable mutation choices. Combinatorial design 3 was performed on top 5 designable sites with the reported designable mutation choices. Combinatorial design 4 was performed on 5 designable sites ranked by designability, but with a bias on sites in bin 1. In other words, sites in bin 2 were not used unless the number of sites in bin 1 was smaller than 5. All combinatorial design calculations were performed and processed under the same setup as the first round of design.

The second round of design was performed using the output structure of v149 from the first round of design as input. No improvement in the number of intermolecular H-bond formation was observed for the outputs of all the 4 combinatorial designs, while the numbers of

intramolecular H-bond were minimally different among the designs. Therefore, the design result with the biggest overall difference in energy score among the top 20 sequences against v149, combinatorial design 4, was chosen, and of which the sequence with the best energy score, v149 plus Y101L/R107F, was selected as the input structure for the third round of design. Combinatorial design 4 was performed with sites 27/59/101/107/110. In the third round, again no improvements in the number of intermolecular and intramolecular H-bonds were observed among the 4 combinatorial designs. So, the sequence with the best energy score, v186 (v149 plus Y101L/R107F/R56T/Y106K/D108A/Y110S), of combinatorial design 1 whose top 20 sequences showed biggest overall improvement in energy scores against the input sequence was selected. Combinatorial design 1 was performed with sites 27/29/56/106/108/110. A further round of CDR design was performed on v186 and all combinatorial design results returned sequences with worse energy score than v186.

Because v186 turned out to be even more prone to aggregation than v149, and based on MD simulation results of v149 against mSA_{WT}, we hypothesized that only mutating CDRs was not sufficient. Therefore, we proceeded to design the framework regions of v149. Because the frameworks of nanobodies are highly conserved [37], suitable sets of amino acid choices and locations would be crucial for the design calculation. Because the previous CDR designs were based on a published summary of nanobody CDR sequence diversity, we referred to the framework sequence used in that study for framework sequence design [37]. We aligned 4NBX.B-biotin103 v186 with chain C of 5VNW, identified framework sites where the two nanobodies differ, and performed a combinatorial design with the selected sites being one or the other amino acid choice. Site positions and sequence choices were as follows: 5-V/G, 12-A/V, 35-A/G, 37-F/Y, and 40-P/A. The v186 structure output by the third round of design was used as input, and the configurations and processing of design calculation were the same as the combinatorial CDR sequence designs described previously. The top-ranked sequence by energy score was v186_Fr (v186 plus A12V/F37Y) (Table S2). Framework design with identical amino acid sites, sequence choices, and calculation configurations was performed using v149 as input, and the design results were also reported in this study for comparison (Table B.S3).

As a comparison, we performed two rounds of CDR sequence design using the above-described configurations on v119. Combinatorial designs in both rounds failed to produce variants with new inter-molecular H-bond formation in the top 20 variants, so the combinatorial designs with best overall energy improvement than the input sequence were chosen, and the top-ranked sequences by energy score in those designs were used as input for further rounds of design and experimental testing. No improvements in kinetics and affinity were observed in these selected sequences (data not shown), in agreement with the unchanged inter-molecular H-bond profiles and with what we observed for v186 versus v149.

B. Design process of 2X89.A-derived nanobody-biotin conjugates

Docking, pose filtering, rotamer screening, and binding pose generation for 2X89.A-CCAA-biotin57 WT against mSA_{WT}

The 154 nanobody structures from PDB with native CDR sequences were docked against a manually selected set of surface residues around the biotin-binding pocket of mSA, in the same way as described in part A. The amino acids being docked against are highlighted below by parentheses with the rest of mSA sequence shown for reference:

GAEAGITGTWYN(QSG)STFTVTAGADGNLTGQY(ENRAQGTG)C(QNSP)YTLTGR
YNGTKLEWRVEWN(NSTENCH)SRTEWRGQYQGGAEARINTQWNLT(YEGGSGP
ATEQGQDT)FTKVK.

Filtering procedures of the docked poses were also identical to those introduced in part A, with an exception that the following residues (highlighted by parentheses) were selected as the target for binding pocket blockage filter:

GAEAGITGTWYN(QS)GSTFTVTAGADGNLTGQYENRAQGTGCQNSPYTLTGRYN
GTKLEWRVEWNNSTENCHSRTEWRGQYQGGAEARINTQWNLT(YEGGSGPATE
QGQDT)FTKVK.

6 poses that respectively comprised nanobodies 2X89.A, 3EBA.A, 4LHQ.B, 4OCL.C, 3V0A.C, and 4P2C.G passed the series of filters. Visual inspection of the poses revealed that the 4OCL.C and 3V0A.C binding poses showed significant contacts that are mediated by nanobody residues outside of the CDRs, so the corresponding two poses were discarded since these binding modes were potentially unrealistic. As the result, 4 final binding poses were kept for the evaluation of optimal conjugation sites (**Figure B.S7A**).

The biotin-CH₂-CH₂-succinimide-S-CH₃ rotamer library built in part A was used again for rotamer screening. Because the design process of the 4NBX.B-derived CDRxAbs demonstrated that our design capability allowed structural stability of the conjugates to be designed after binding synergy was designed, we did not put much emphasis on preserving structural integrity in the early stage of the design process of 2X89.A-derived CDRxAbs. Therefore, instead of screening only against alanine CDR residues, all CDR residues were screened by measuring the distance between the terminal carbon of the rotamer and the C β of respective CDR residues, steric clash between the rotamer and the streptavidin, steric clash between the rotamer and both proteins, and the angle of the rotamer terminal carbon approaching the respective attachment spot. Conjugation geometries that clashed with streptavidin by <0.5 unit, clashed with both proteins by <10 units, approached the attachment spot by 100-120 degree, and were <2 Å away from the C β of screened conjugation sites were kept. Only one rotamer that was screened against I57 of 2X89.A passed the filter. The final conjugation structure was prepared by BIOGRAF, under the same parameters as described in part A [53].

To remove the intra-CDR disulfide bond in the Biograf-relaxed structure, C33A/C104A mutations were introduced by TRIAD sequence design module to create the finalized model of 2X89.A-CCAA-biotin57/mSA for further sequence design optimization.

Subsequent sequence design of 2X89.A-CCAA-biotin57 conjugates

Summarizing the experience from the design process of the 4NBX.B-biotin103 conjugates, we gained the following insights into the sequence design principles of CDRxAbs:

1. Performing sequential rounds of design on limited sets of amino acid sites and choices that are recommended by iterative energetic and structural analysis allows functionally-improved CDRxAb mutants to be discovered without experimentally screening a large set of sequences.
2. New intermolecular interactions between the nanobody scaffold and the target can be engineered first before further mutations are introduced to optimize the structural integrity of the conjugates.
3. Simply mutating CDR residues is not sufficient for structural optimization of the conjugates, and mutated CDR residues likely need accommodation by introducing mutations in the β -barrel framework region.

Based on the above principles, we established a rudimentary sequence design pipeline and tested it on 2X89.A-CCAA-biotin⁵⁷ to create mutants that are less prone to aggregation. The pipeline is detailed below (**Figure B.S7D**):

1. The pipeline performs sequential rounds of sequence design that is either restricted on CDR residues or framework residues. The first round of design is performed on CDR residues. Residues that are mutated in previous rounds are kept from mutation in further rounds.
2. H-bonds are still the only intermolecular interactions that are explicitly evaluated after design calculations and biased towards for sequence selection, but other types of interactions can be evaluated if they are deemed to be crucial for specific scenarios.
3. CDR sequence design follows the procedures of designing the CDR loops of v149, as described in part A, with one exception: besides the four combinatorial designs, it is optional that two additional combinatorial designs can be performed in parallel, respectively on all identified sites and amino acid choices in bin 2, and on the top 5 designable sites with a bias on sites in bin 2. All combinatorial designs are evaluated together as described in part A.

4. Framework design follows the procedures of designing the framework of v186, as described in part A.
5. To evaluate the design results of CDR design, the following steps are used:
 - a). Sequences that showed worse energy score than the immediate parent sequence are discarded.
 - b). Sequences with the number of intermolecular H-bonds lower than the immediate parent sequence are discarded.
 - c). If no sequences survived filters a and b, perform framework design on the immediate parent sequence.
 - d). For sequences pass the filters, the sequence that has the highest number of intermolecular H-bonds and the best energy score among sequences that share the same number of intermolecular H-bonds is kept as input for the next round of CDR design.
6. To evaluate the design results of framework design, the following steps are used:
 - a). Sequences that showed worse energy score than the immediate parent sequence are discarded.
 - b). Sequences with the number of intermolecular H-bonds lower than the immediate parent sequence are discarded.
 - c). Sequences with the number of nanobody intramolecular H-bonds lower than the immediate parent by >1 are discarded.
 - d). If no sequences survived filters a-c, design fails.

e). For sequences pass the filters, the sequence that has the highest energy score is kept as input for the next round of CDR design.

The outputs after rounds 3, 5, 6, and 7, which are variants v37, v42, v20, and v5 were selected for experimental testing.

C. Nanobody sequences that are used in this study for docking and binding pose selection

Note: Sequences are represented by [PDB ID].[chain name]

1BZQ.K, 1JTT.A, 1KXQ.E, 1OP9.A, 1ZVH.A, 1ZVY.A, 2X89.A, 3EBA.A, 3JBC.7, 3JBE.7, 3JBF.7, 4GRW.E, 4I0C.C, 4W6W.B, 5BOP.A, 5C2U.B, 5FOJ.A, 5M13.B, 5TJW.K, 2XT1.B, 3K74.B, 3SN6.N, 4C57.C, 4EIZ.C, 4FHB.D, 4GRW.F, 4HEM.E, 4KML.B, 4LGP.B, 4N9O.B, 4NBY.B, 4NBZ.B, 4TVS.b, 5C3L.D, 5F1K.C, 5F1O.B, 5H8O.A, 5IMK.B, 5J57.B, 5JA8.B, 5JMO.C, 5KU2.7, 5LHN.B, 5NBD.C, 5O03.C, 5USF.C, 5UZ7.N, 5VXL.B, 5VXM.B, 4Z9K.B, 5IVN.A, 3QXT.A, 4DK3.A, 4GFT.B, 4KRL.B, 4LHQ.B, 4N1H.B, 4S10.A, 5OJM.K, 5UKB.a, 5VXK.B, 1I3U.A, 1RJC.A, 4GRW.H, 4IOS.D, 4OCL.C, 5BOZ.G, 5JDS.B, 5M2M.D, 1KXV.C, 1QD0, 1RI8.A, 1ZV5.A, 3JBD.7, 3K3Q.A, 3QXV.A, 3RJQ.B, 3STB.A, 4AQ1.B, 4CDG.C, 4I13.B, 4QO1.A, 4W6X.B, 4WEM.B, 4XT1.C, 5E0Q.A, 5G5R.B, 5GXB.B, 5LHR.B, 5LWF.C, 5O8F.K, 5OCL.B, 4C58.B, 2XXM.B, 3JBG.7, 4X7F.C, 5F21.B, 5F7K.C, 5HVG.B, 4LHJ.B, 4M3K.B, 5HVF.B, 5MJE.B, 1G6V.K, 2X6M.A, 3K1K.C, 3V0A.C, 4U3X.A, 5E5M.B, 5HGG.S, 5HM1.A, 5J56.B, 5JA9.A, 5MWN.N, 5O2U.B, 5OVW.G, 5TOK.D, 1KXT.B, 3CFI.C, 3K81.A, 4EIG.B, 4HEP.G, 4LGR.B, 4MQS.B, 4NBX.B, 4NC2.B, 4P2C.G, 4W6Y.B, 4YGA.B, 5DFZ.E, 5IP4.A, 5J1S.C, 5JQH.C, 5KTZ.7, 5KU0.7, 5KWL.7, 5L21.B, 5MY6.B, 5O02.C, 5O04.E, 5O0W.E, 5OCL.A, 5OMN.C, 5TOJ.D, 5UK4.a, 5VXJ.B, 4X7C.C, 3EZJ.B, 5F7L.B, 5M30.D, 5M94.B, 3P0G.B, 4LDE.B, 2BSE.D

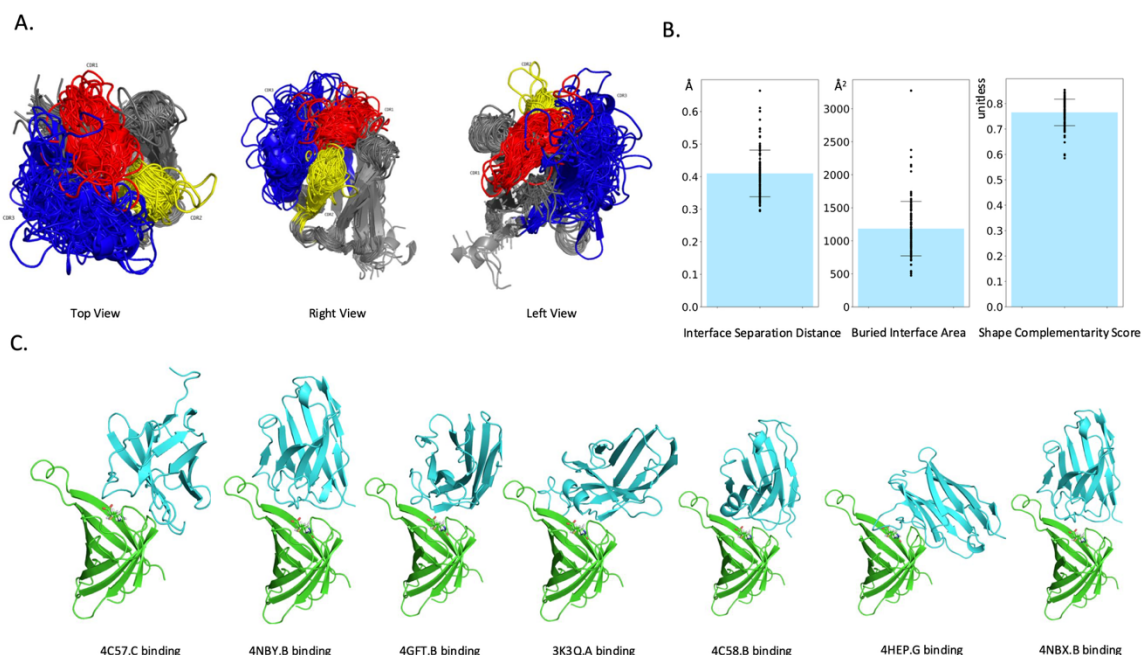


Figure B.S1 Searching for optimal CDR conformation against monomeric streptavidin model. (A) Whole structural alignment of 154 curated PDB nanobody scaffolds with diverse CDR conformations and sequences. (B) Interface statistics of naturally occurring nanobody-target complexes. Error bars represent standard deviations. (C) The final 7 docked poses of nanobody scaffolds that passed the filters selecting poses that most likely recapitulate the natural binding modes of the corresponding nanobody scaffolds. Streptavidin S45A/T90A/D180A is colored green, and nanobody scaffold is colored cyan.

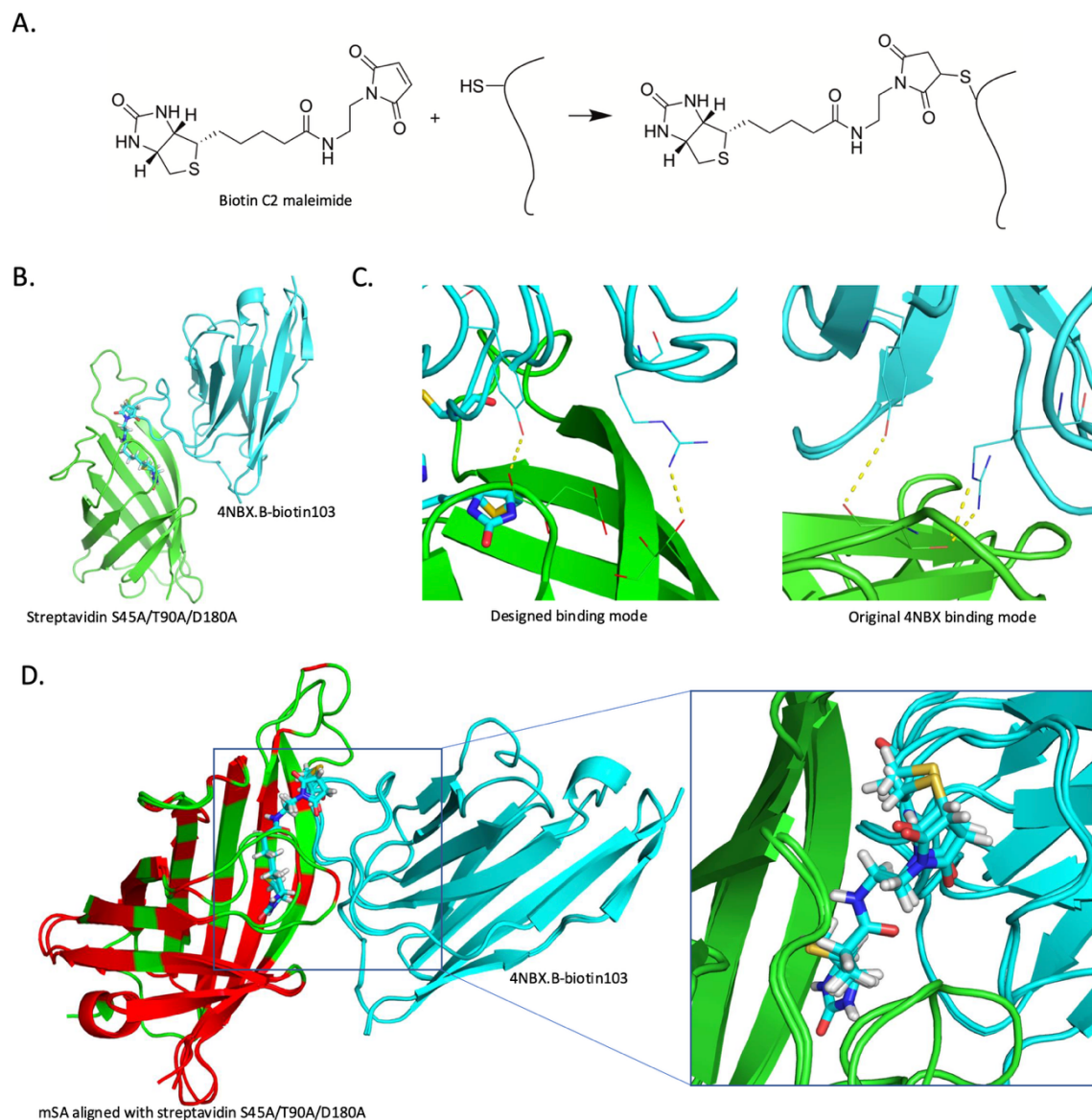


Figure B.S2 Identification of optimal conjugation strategy and finalized conjugate models. (A) Biotin conjugation was performed by biotin C2 maleimide with mutated cysteine residues. (B) Prepared structure of 4NBX.B-biotin103 in complex to streptavidin S45A/T90A/D180A. Streptavidin S45A/T90A/D180A is colored green, and nanobody scaffold is colored cyan. (C) The H-bond forming potential of Y112 and R27 in 4NBX.B nanobody was predicted to be recapitulated in the designed binding pose with the streptavidin model. Streptavidin S45A/T90A/D180A is colored green, and nanobody scaffold is colored cyan. Biotin103 side chain is shown as stick. Y112 and R27 together with their predicted H-

bond partners are shown as line. **(D)** Alignment results for prepared structures of 4NBX.B-biotin103 in complex with streptavidin S45A/T90A/D180A and mSA. Streptavidin models are colored green, and nanobody scaffolds are colored cyan. Residues that were identified by sequence alignment as pair-wise identical sequences are colored red. Biotin103 side chain from both models is shown as a stick.

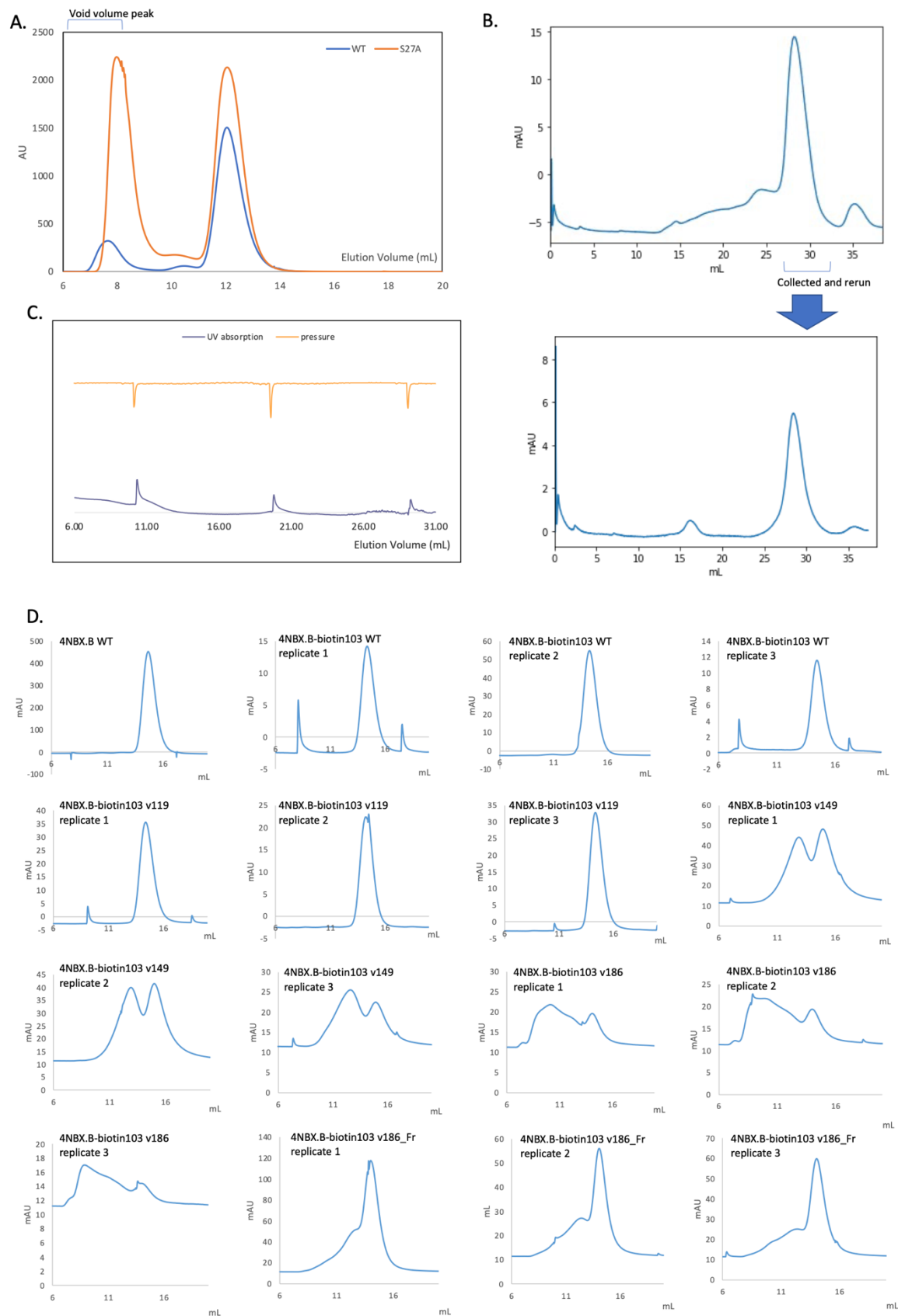


Figure B.S3 Additional Supporting SEC Traces. (A) SEC traces of mSA_{WT} and mSAS27A. (B) SEC rerun trace of collected monomeric fraction for 4NBX.B-biotin103 v186_Fr for an independent re-expression and purification. (C) Buffer only run of the FPLC that reveals instrument artifacts (i.e. spikes) that appear in many of the SEC traces and that are labeled with “*” in Figure 3A. A different, well-behaved FPLC system was used for (B). (D) Unnormalized SEC traces for proteins shown in **Figure B.4A**.

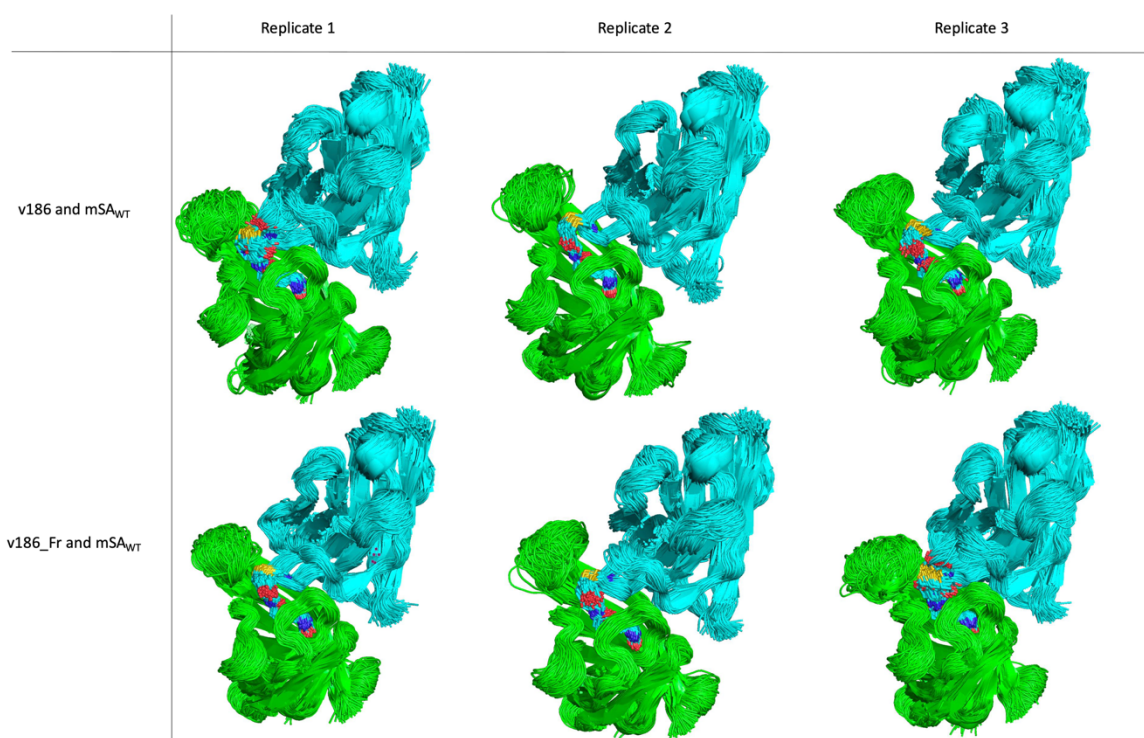


Figure B.S4 Summary of MD simulations performed for 4NBX.B-biotin103 v186 and v186_Fr against mSA_{WT}. For each simulation, 400 snapshots evenly spaced along the 100-ns timescale are aligned together. green: mSA_{WT}. cyan: nanobody-biotin conjugates. The biotin103 “residue” is shown as a stick.

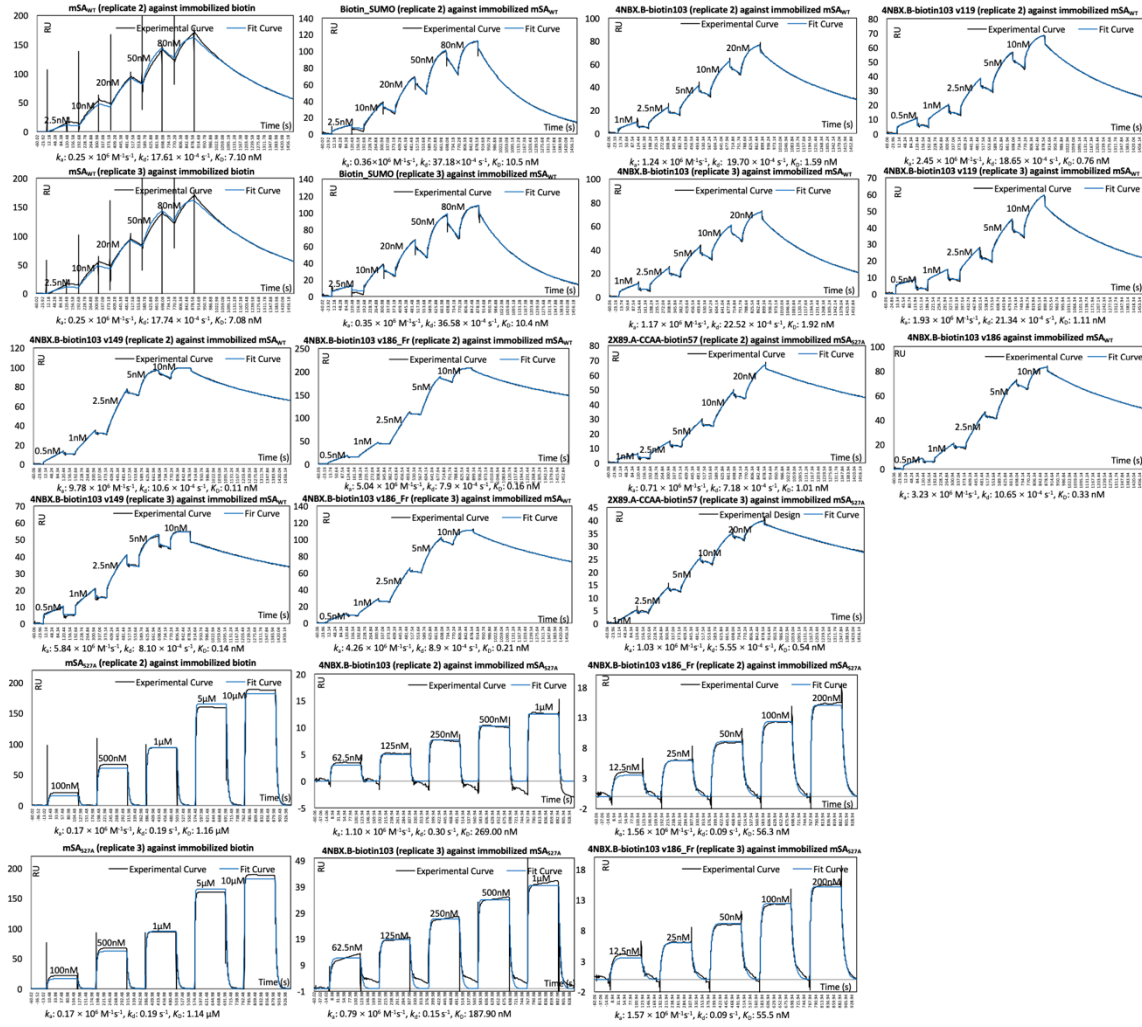


Figure B.S5 SPR measurements from the intermediate design variants. Biological replicates not shown in the main text were included for affinity and kinetics estimation.

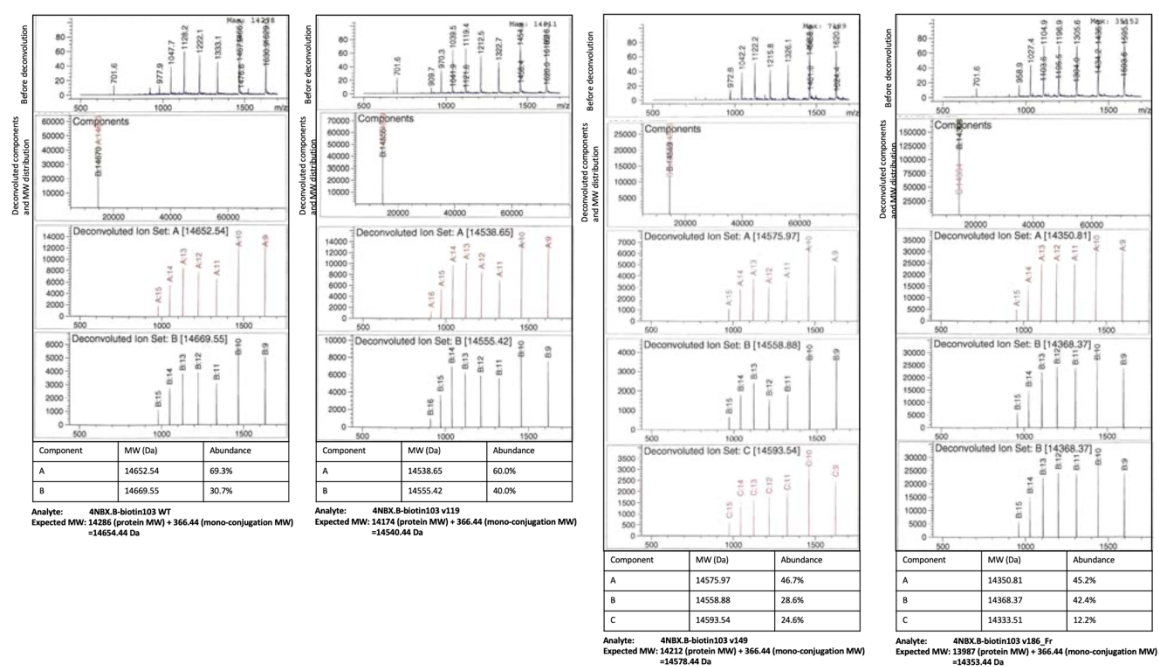


Figure B.S6 Intact-protein mass spectrometry (MS) confirmed mono-conjugated materials. MS deconvolution of nanobody-biotin conjugates returned MWs within 20 Da of the expected MW of mono-conjugated materials. The most abundant component for all four proteins is the mono-conjugated product (noting the ± 3 Da resolution of the mass spec instrument that was used). The differences between the expected and observed masses are 1.9 Da, 1.8 Da, 2.5 Da, and 2.6 Da, respectively. The expected mass is larger than the observed mass in all cases. The next most abundant component for three of the four cases has approximately +17 Da of added mass, presumably corresponding to succinimide ring opening: 17.0 Da, 16.8 Da, 17.6 Da (for the 24.6% component), and 17.6 Da, respectively. The final two components for the last two proteins (at 28.6% and 12.2% abundance, respectively) are approximately 17 Da lighter than the mass of the mono-conjugated protein: 17.1 Da and 17.3 Da, respectively. This mass difference could reflect N-terminal pyroglutamate formation given that all of the nanobodies used in this study have N-terminal glutamines (**Table B.S5**).

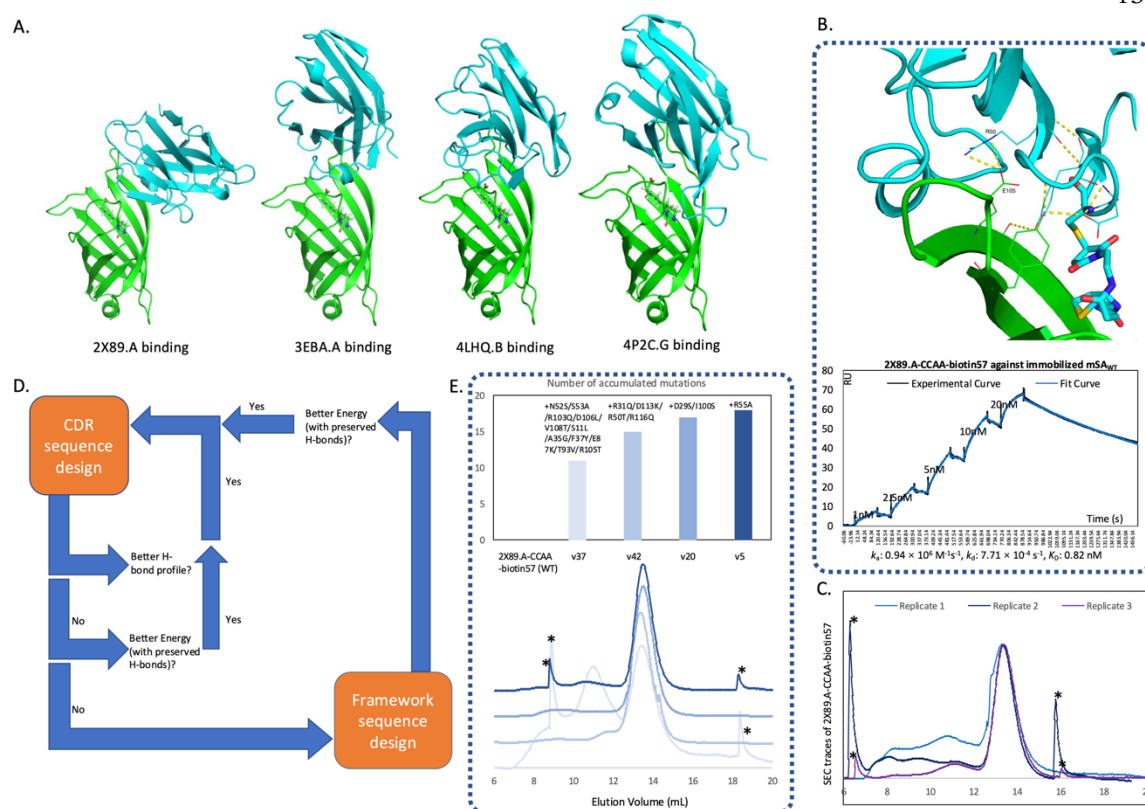


Figure B.S7 Designing and testing of a nanobody scaffold obtained by directly docking against mSA. (A) Finalized nanobody scaffolds and binding poses against mSA streptavidin. (B) Predicted H-bond formation profile and SPR binding curve for 2X89.A-CCAA-biotin57 WT against immobilized mSA_{WT}. Data from one of the triplicates is shown here, and data from the other two replicates is in Fig. S5. (C) Size-exclusion chromatography (SEC) traces of biological triplicates for 2X89.A-CCAA-biotin57, normalized by monomer peak height for better comparison of aggregates formation. * indicates peaks of sample-irrelevant instrument defect of the overall FPLC, please refer to figure S3C for more details. (D) A rudimentary sequence design pipeline that performs CDR and framework design in a step wise manner. (E) SEC traces and newly accumulated mutations for sequence-designed variants of 2X89.A-CCAA-biotin57 conjugates. Peak of the monomeric fractions were normalized to identical heights. * indicates peaks of sample-irrelevant instrument defect of the overall FPLC, please refer to figure S5C for more details.

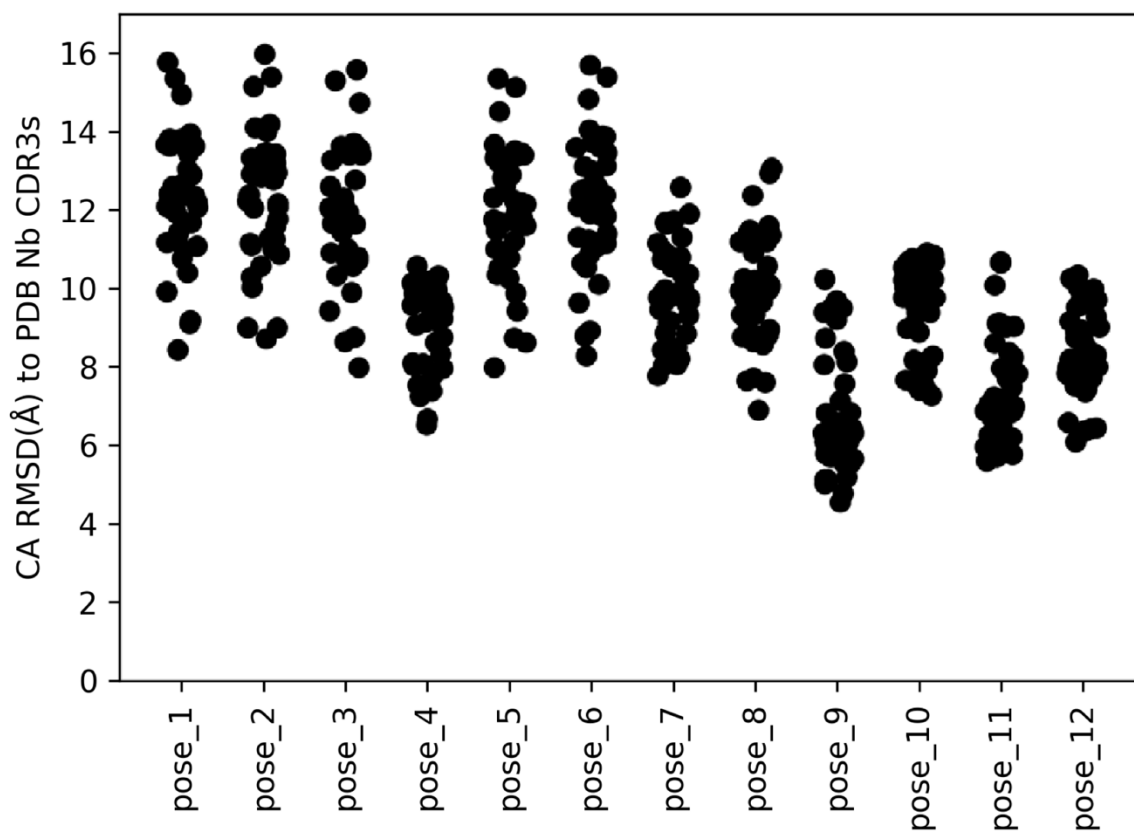


Figure B.S8 CDR3 loop modeling for 12 docked poses. 40 PDB nanobody X-ray crystal structures with less than 3 Å resolution that share the same CDR3 length as the remodeled pose were aligned using all amino acids, and then the CA RMSD between CDR3 amino acids were calculated for each aligned pair of structures.

Table B.S2: Top 20 sequence outputs from CDR design of site 31, 32, 104, and 105 on 4NBX.B-biotin103 WT. (Ranked by energy score)

Ranking	Energy Score	Note	Mutations: chain name and accepted residue
1	-457.58628	v119	B_31H+B_32A+B_104S+B_105H
2	-456.82531		B_31N+B_32S+B_104S+B_105S
3	-456.56855		B_31N+B_32S+B_104S+B_105A
4	-456.56059		B_31R+B_32A+B_104D+B_105R
5	-456.50215		B_31R+B_32A+B_104S+B_105Y
6	-456.42944		B_31N+B_32A+B_104S+B_105S
7	-456.25964		B_31Q+B_32A+B_104S+B_105Y

8	-456.13421		B_31H+B_32S+B_104S+B_105V
9	-456.11428		B_31N+B_32A+B_104S+B_105R
10	-455.81205		B_31N+B_32A+B_104S+B_105A
11	-455.7737		B_31N+B_32S+B_104S+B_105D
12	-455.67248		B_31R+B_32A+B_104D+B_105H
13	-455.6018		B_31Q+B_32A+B_104A+B_105R
14	-455.26629		B_31R+B_104S+B_105H
15	-455.13137		B_31H+B_32S+B_104D+B_105S
16	-455.10018		B_31N+B_32S+B_104A+B_105S
17	-455.05911	v149	B_31R+B_32S+B_104A+B_105R
18	-454.97734		B_31R+B_32S+B_104S+B_105Y
19	-454.93878		B_31K+B_32S+B_104A+B_105A
20	-454.90643		B_31Q+B_32S+B_104A+B_105A

Note: chain B refers to the nanobody

Table B.S3: Top 20 sequence outputs from framework design on 4NBX.B-biotin103 v186. (Ranked by energy score)

Ranking	Energy Score	Note	Mutations: chain name and accepted residue
1	-500.8748	v186_Fr	B_12V+B_37Y
2	-500.3636		B_37Y+B_40A
3	-500.24131		B_5Q+B_12V+B_37Y
4	-500.10634		B_12V+B_35G+B_37Y
5	-499.9846		B_12V+B_37Y+B_40A
6	-499.90705		B_12V
7	-499.22799		B_12V+B_35G
8	-499.15484		B_5Q+B_12V+B_35G+B_37Y
9	-499.02353		B_12V+B_35G+B_37Y+B_40A
10	-498.98295		B_5Q+B_12V+B_35G+B_37Y+B_40A
11	-498.78612		B_37Y
12	-498.6384		B_12V+B_35G+B_40A
13	-498.54874		B_35G+B_37Y
14	-498.53802		B_12V+B_40A
15	-498.48109		B_5Q+B_12V+B_40A
16	-498.31249		B_5Q+B_37Y
17	-498.04085		WT
18	-497.75764		B_5Q+B_12V
19	-497.68241		B_5Q
20	-497.59715		B_5Q+B_35G+B_37Y

Note: chain B refers to the nanobody

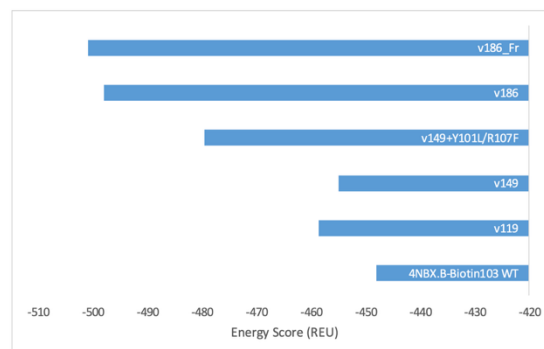
Table B.S4: Top 20 sequence outputs from framework design on 4NBX.B-biotin103 v149.
(Ranked by energy score)

Ranking	Energy Score	Note	Mutations: chain name and accepted residue
1	-482.29049		WT
2	-481.67807		B_37Y
3	-481.6627		B_5Q+B_12V
4	-481.65469	A12V/F37Y mutations	B_12V+B_37Y
5	-481.51508		B_12V+B_35G
6	-481.5122		B_5Q
7	-481.48205		B_35G
8	-481.38994		B_35G+B_37Y
9	-481.32699		B_5Q+B_37Y
10	-481.13866		B_40A
11	-481.12555		B_12V
12	-480.83661		B_35G+B_40A
13	-480.82825		B_12V+B_37Y+B_40A
14	-480.50684		B_5Q+B_35G+B_37Y
15	-480.44015		B_5Q+B_12V+B_40A
16	-480.42843		B_12V+B_35G+B_37Y
17	-480.3031		B_5Q+B_12V+B_37Y+B_40A
18	-480.19273		B_5Q+B_12V+B_35G+B_40A
19	-480.16999		B_5Q+B_35G+B_40A
20	-480.16866		B_12V+B_35G+B_37Y+B_40A

Note: chain B refers to the nanobody

Table B.S5: Summary of energy scores for proteins experimentally evaluated in this study.

Name	Energy score (REU)
4NBX.B-Biotin103 WT	-447.97681
v119	-458.58628
v149	-455.05911
v149+Y101L/R107F	-479.59553
v186	-498.04085
v186_Fr	-500.8748

**Table B.S6:** Amino acid sequences for the protein templates used in this study.

4NBX.B_A103C	QVQLQESGGGLAQAGGSLRLSCAASGRTFSMDPMAWFRQPPG KEREFVAAGSSTGRTTYADSVKGRFTISRDNKNTVYLQMN SLKPEDTAVYYCAAAPYGCNWRDEYAYWGQGTQVTVSSH HHHH
2X89.A_CCAA_I57C	QVQLQESGGGSVQAGGSLRLSCAASGYTDSRYAMAWFRQAP GKEREWVARINSGRDCTYYADSVKGRFTFSQDNAKNTVYLQ MDSLEPEDTATYYCATDIPLRARDIVAKGGDGFYWGQGTQV TVSSHHHHHH
mSA_WT	HHHHHHSQDLASAEAGITGTWYNQSGSTFTVTAGADGNLTGQ YENRAQGTGCQNSPYTLTGRYNGTKLEWRVEWNNSTENCHS RTEWRGQYQGGAEARINTQWNLTYEGGSGPATEQGQDTFTK VKPSAASGS DYKDDDDK
Smt3 SUMO with N-terminal cysteine	CLQDSEVNQEAKPEVKPEVKPETHINLKVS DGSSEIFFKIKKTP LRRLEAF AKRQGKEMDSLRF LYDGIRIQADQAPEDLDMEDN DIEAHREQIGGHHHHHH

Note: the numbering scheme of the above sequences follows 1-2-3-4... The first residue is numbered as 1.

Plasmids, Expression Cell Lines, and Cloning of Protein Variants

pRSET-mSA was a gift from Sheldon Park (Addgene plasmid # 39860) [34]. The mSA S27A mutation was created by site-directed mutagenesis using commercially available kits (NEB). 4NBX.B_C103 and 2X89.A_CCAA_C57 sequences were directly ordered from IDT and cloned into a pHen6c vector by Gibson assembly using commercially available reagents (NEB) [57]. The assembled pHen6c vectors harbor a PelB N-terminal signal sequence,

allowing for bacterial periplasmic expression [58]. Variants of 4NBX.B_C103 and 2X89.A_CCAA_C57 were created by mutagenic PCR and assembled into a pHen6c vector by Gibson assembly using commercially available reagents (NEB). 4NBX.B WT sequence with C103A mutation was created by site-directed mutagenesis using commercially available kits (NEB). Smt3 SUMO protein with an N-terminal cysteine was created from wild type Smt3 SUMO by mutagenic PCR and subcloned into a pY71A(lc) vector by Gibson assembly using commercially available reagents (NEB).

Expression and Purification of Wild-Type mSA Streptavidin and its S27A Variant

Expression, purification, and refolding of mSA variants followed published protocols with slight variations [34]. The expression plasmids were first transformed into *E. coli* BL21-Gold (DE3) chemically competent cells (Agilent), which were then grown overnight in LB with 100 µg/ml of ampicillin (amp100) at 37 °C on a shaker running at 250 rpm. 1 ml of the overnight culture was used to inoculate 300 ml of TB medium (2.3 g KH₂PO₄, 16.4 g K₂HPO₄, 12 g tryptone, 24 g yeast extract, 4 ml glycerol, dissolved in water to 1 L) supplemented with 2 mM MgCl₂, 0.1% glucose, and amp100. Inoculation was done at 37 °C and 250 rpm until an OD₆₀₀ of 1.5 to 2. Expression was induced by 1 mM IPTG at 28 °C and 250 rpm for 18 hours. Cells were then centrifuged at 4500 g for 15 minutes at 4 °C and protein extraction was then performed using 50 ml of chemical lysis buffer. The lysis buffer was composed of 1 × PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, pH 7.4), 1 × CellLytic B reagent (Sigma), 0.02 mg/ml DNase1, 0.2 mg/ml lysozyme, and 1 mM protease inhibitor AEBSF (Sigma). Pellets were resuspended in lysis buffer and nutated for 4 hours at room temperature. Cell lysate was then centrifuged at 15,000 g for 30 minutes at 4 °C, and the pellets was used to refold and purify the protein.

Pellets were first resuspended into 3 mL of 6 M guanidine hydrochloride in 1 × TBS (50 mM Tris, 150 mM NaCl, pH 8.0) and incubated under 37 °C for 30 minutes to solubilize the proteins. Un-dissolved materials were cleared by 15,000 g centrifugation for 5 minutes at 4 °C. Supernatants were then chilled on ice before dropwise addition into 40 ml of pre-chilled refolding buffer (50 mM Tris-HCl, 150 mM NaCl, 0.3 mg/ml D-biotin, 0.2 mg/ml oxidized

glutathione, 1 mg/ml reduced glutathione, pH 8.0) while stirring. The refolded mSA protein was then allowed to incubate on ice for an additional 2 hours with stirring before centrifugation at 15,000 g for 30 minutes at 4 °C to remove insoluble material. The supernatants were then supplemented with 20 mM imidazole and then loaded onto 1 ml bed volume of Ni-NTA agarose beads (Qiagen) that were pre-washed with 5 column volumes of 1 × PBS. Sample loading was performed by gravity flow. The column was then washed with 10 column volumes of 1 × PBS supplemented with 20 mM imidazole before 3 ml of 1 × PBS supplemented with 500 mM imidazole was used to elute the protein. The entire Ni-NTA purification process was performed at 4 °C. The 3 ml of purified protein was then dialyzed against fresh 1 L of 1 × PBS (pH 8.0) at 4 °C 3 times using Slide-A-Lyzer 10 kDa molecular-weight cutoff dialysis cassettes (Thermo). Each dialysis step was performed for at least 4 hours. The dialyzed mSA product was further purified at 4 °C using a Superdex 75 10/300 GL SEC column (GE) with 1 × PBS (pH 7.4) as running buffer. The fractions corresponding to monomeric mSA were collected for subsequent experiments.

Expression and Purification of Smt3 SUMO Protein with N-terminal Cysteine

The expression plasmid was first transformed into *E. coli* BL21-Gold (DE3) chemically competent cells (Agilent), which were then grown overnight in LB with amp100 and 1% glucose at 37 °C and 250 rpm. 1 ml of the overnight culture was then used to inoculate 300 ml of TB medium with 2 mM MgCl₂, 0.1% glucose, and amp100 at 37 °C and 250 rpm until an OD₆₀₀ of 1.5 to 2. Expression was then induced by 1 mM IPTG at 28 °C and 250 rpm for 18 hours. Cells were then pelleted at 4500 g for 15 minutes at 4 °C, resuspended in 50 ml of chemical lysis buffer supplemented with 5 mM 2-mercaptoethanol (BME) and incubated for 4 hours at room temperature to release the expressed protein. Lysate supplemented with 20 mM imidazole was cleared by centrifugation at 15,000 g at 4 °C for 30 minutes and the supernatant was then loaded onto 1 ml bed volume of Ni-NTA agarose beads (Qiagen) pre-washed with 5 column volumes of 1 × TBS (pH 7.3). Sample loading was performed by gravity flow. The loaded column was then washed with 5 column volumes of 1 × TBS (pH 7.3) supplemented with 20 mM imidazole and 5 mM BME and another 5 column

volumes of 1 × TBS (pH 7.3) supplemented with 20 mM imidazole. 3 ml of 1 × TBS (pH 7.3) supplemented with 500 mM of imidazole was used to elute the proteins. All Ni-NTA purification procedures were performed under 4 °C. Purified protein was then concentrated to ~0.5 ml using Amicon 10 kDa molecular-weight cutoff centrifuge filters (GE) and stored for subsequent experiments.

Expression and Purification of Nanobodies

The pHen6c expression plasmids were first transformed into *E. Coli* BL21-Gold (DE3) chemically competent cells (Agilent), which were then grown overnight in LB with amp100 and 1% glucose at 37 °C and 250 rpm. 1 ml of the overnight culture was then used to inoculate 300 ml of TB medium with 2 mM MgCl₂, 0.1% glucose, and amp100 at 37 °C and 250 rpm until an OD₆₀₀ of 1.5 to 2. Expression was then induced by 1 mM IPTG at 28 °C and 250 rpm for 18 hours. Cells were then pelleted under at 4,500 g for 15 minutes at 4 °C and resuspended in 12 ml of TES periplasmic extraction buffer (0.2 M Tris, 0.5 mM EDTA, 0.5 M sucrose, pH 8.0) (supplemented with 5 mM BME if the nanobody had a cysteine handle for conjugation) before incubation on ice with shaking at 32 rpm for 1 hour [58, 59]. 18 ml of 4x diluted TES buffer (supplemented with 5 mM BME if the nanobody had a cysteine handle) was then added to the cells which were incubated on ice at 32 rpm for another hour [59, 60]. After periplasmic extraction, the cells were pelleted at 15000 g at 4 °C for 30 minutes and the supernatant was supplemented with 20 mM of imidazole before loading onto 1 ml bed volume of Ni-NTA agarose beads (Qiagen) pre-washed with 5 column volumes of 1 × TBS (pH 7.3). Sample loading was performed by gravity flow. The loaded column was then washed with 10 column volumes of 1 × TBS (pH 7.3) supplemented with 20 mM of imidazole. 3 ml of 1 × TBS (pH 7.3) supplemented with 500 mM of imidazole was used to elute the proteins. For nanobodies with a cysteine handle, 5 mM BME was added to the first 5 column volumes of wash buffer and the elution buffer was supplemented with 5 mM TCEP. The elution buffer was incubated with Ni-NTA beads for 30 minutes before eluting the proteins. The entire Ni-NTA purification process was performed at 4 °C. Purified nanobodies with a cysteine handle in 3 ml of elution buffer were concentrated to ~0.5 ml

using Amicon 10 kDa molecular-weight cutoff centrifuge filters (GE) and stored for subsequent experiments. 4NBX.B WT was instead further purified using a Superdex 75 10/300 GL SEC column (GE) with $1 \times$ PBS (pH 7.4) as a running buffer and the fractions corresponding to the monomeric protein were collected for subsequent experiments.

Biotin C2 Maleimide Conjugation and Purification of Conjugates

Maleimide labeling on the solvent exposed cysteines of nanobodies followed a published protocol with some modifications [61]. Conjugation reactions were typically carried out in 1.5 to 2 ml reactions with 1 mM maleimide reagent and protein concentrations to ensure at least a 20-fold excess of maleimide reagent to protein. Purified nanobodies from storage were first incubated with another 5 mM TCEP supplement at 4 °C for 2 hours and then buffer exchanged into $1 \times$ TBS (pH 7.3) using HiTrap desalting columns (GE) at room temperature to remove TCEP. Thawed stock solutions (100 mM in DMSO) of biotin C2 maleimide (AnaSpec) were immediately added to the buffer-exchanged nanobodies to final concentration of 1 mM before the reaction mixture was nutated at 4 °C for 4 hours with a tinfoil cover to avoid light contact. Excess maleimide stock solutions were discarded. Biotin C2 maleimide was in >20-fold molar excess over the nanobody in the reaction mixture. The reaction mixture was then filtered using 0.2 μ m syringe filters (Thermo) to remove precipitated protein and then buffer exchanged to $1 \times$ TBS (pH 7.3) using PD-10 desalting columns (GE) at room temperature to remove excess maleimide reagent. The labeled nanobodies were further purified at 4 °C using a Superdex 75 10/300 GL SEC column (GE) with $1 \times$ TBS (pH 7.3) as the running buffer and the fractions corresponding to monomeric protein were collected for subsequent experiments. Maleimide labeling of Smt3 SUMO protein with an N-terminal cysteine followed the identical procedure described above.

Intact Protein Mass Spectrometry (MS) Workflow to Analyze Conjugation Efficiency

HPLC-MSD (HP, Agilent) was used to assess the labeling efficiency of the nanobody-biotin conjugates. Conjugates were first dried using a spin vacuum evaporator and then resuspended in 0.2% formic acid. A C3 HPLC column was used to purify the protein sample before MS

analysis. Before running samples, the column was first washed with isopropyl alcohol (IPA). Analysis of the conjugation efficiency of 4NBX.B-based conjugates was performed by deconvoluting each eluted sample HPLC peak using the following parameters: positive adduct ion +H 1.0079 Da, negative adduct ion -H -1.0079 Da, molecular weight cutoff 5000–80 000 Da, maximum charge 90, minimum peaks 5, ion PWHH 0.6 Da, molecular weight agreement 0.05%, noise cutoff 0, abundance cutoff 10%, molecular weight assignment cutoff 40%, and envelope cutoff 50%. Deconvolution was performed with ChemStation (Agilent). For each sample of interest, approximately 0.1 to 1 μ g of material was used for the analysis.

Surface Plasmon Resonance (SPR) Analysis of Binding Affinity and Kinetics

A Biacore T200 instrument (GE) was used to perform SPR analysis. 4NBX.B WT, 4NBX.B-biotin103 conjugates, and 2X89.A-CCAA-biotin57 conjugates were first buffer-exchanged into HBS-EP+ buffer (Teknova) using Amicon 10 kDa molecular-weight cutoff centrifuge filters (GE). Protein concentration was then determined by BCA assay using commercially available kits (Thermo). The calibration curve for the BCA assay was prepared using purified 4NBX.B WT, which was also buffer exchanged into HBS-EP+ but had concentrations determined by A_{280} readings using an extinction coefficient $30,035 \text{ M}^{-1}\text{cm}^{-1}$. For SPR analysis, biotin pentylamine (Thermo), mSA_{WT}, and mSA_{S27A} were separately immobilized on CM5 sensor chips (GE) using an EDC/NHS amine coupling kit following standard protocols (GE). Binding kinetics were measured by single-cycle kinetics experiments. Biotin pentylamine was immobilized at 7.5 mM concentration to reach a target surface density of ~ 200 resonance units (RUs) [62]. In order to compare how binding events changed in response to different surface densities, surfaces with three different densities of immobilized mSA_{WT} at ~ 1000 RU, ~ 2500 RU, and ~ 3000 RU were prepared under immobilization concentrations 0.1 μ M, 0.5 μ M, and 1 μ M, respectively. Immobilization of mSA_{S27A} was also performed at 0.01 μ M and 0.05 μ M concentrations with target surface density of ~ 200 and ~ 600 RU. The fitted affinities and kinetics of identical conjugates using different surfaces densities were not significantly different. Reference channels were either treated

with EDC/NHS using HBS-EP+ buffer alone or with 1 μ M 4NBX.B WT to assess non-specific interactions. No non-specific interactions were observed. All immobilization samples were dissolved in acetate buffer (pH 4.5).

Binding experiments were performed at 25 °C. HBS-EP+ was used as running buffer. The flow channels were first incubated in the running buffer before analytes at 5 different concentrates were consecutively injected at a flow rate of 30 μ L/min through both the reference channel and the sample channel containing the immobilized molecules of interest. After injections, the bound analytes were allowed to dissociate for 10 minutes in order to generate dissociation curves. HBS-EP+ buffer was then washed through both the reference and sample channels in order to regenerate the surfaces for the next binding experiments. Sensorgrams were processed with Biacore evaluation software using a 1:1 kinetics model. For immobilized mSA_{WT}, global fitting of bulk shifts was turned on to accommodate the observed bulk shifts before and after each injection event.

Molecular Dynamics (MD) Simulation Protocols

Molecular Dynamics simulations were carried out using ACEMD (Acellera) [63]. Each system was placed in a box with dimensions selected to allow 12 Å of clearance on each side. The systems were solvated using the TIP3P water model [64] and counter ions were added to bring the net charge to zero. The systems were then minimized for 500 steps and equilibrated for 5 ns before 100 ns production runs were carried out at 300 K. All experiments utilized the Amber ff14SB force field and a 4 femtosecond timestep [65]. Data from the equilibration run was not included in subsequent analysis, and where replicates were collected no part of the intermediate data was reused. Parameters for the biotin-CH₂-CH₂-succinimide-S-CH₃ ‘side chain’ were prepared using Antechamber and utilized RESP charges calculated with Gaussian 09 [66, 67]. Calculation of solvent accessible surface areas was performed using MDTraj, and hydrogen bonding was assessed using a tcl script written for VMD [68, 69].

Acknowledgements

We thank Monica C. Breckow for technical assistance; Dr. Jost G. Vielmetter and the Caltech Protein Expression Center for consultation with Biacore experimental design and data interpretation; and Dr. Barry D. Olafson and Paul Chang from Protabit LLC for consultation on the implementation and use of the TRIAD and BIOGRAF software programs.

Bibliography

1. Swinney DC. Biochemical mechanisms of drug action: what does it take for success? *Nature Reviews Drug Discovery*. 2004;3(9):801-8. 10.1038/nrd1500
2. Liu JKH. The history of monoclonal antibody development – Progress, remaining challenges and future innovations. *Annals of Medicine and Surgery*. 2014;3(4):113-6. 10.1016/j.amsu.2014.09.001
3. Hopkins AL, Groom CR. The druggable genome. *Nature Reviews Drug Discovery*. 2002;1(9):727-30. 10.1038/nrd892
4. Verdine GL, Walensky LD. The Challenge of Drugging Undruggable Targets in Cancer: Lessons Learned from Targeting BCL-2 Family Members. *Clin Cancer Res*. 2007;13(24):7264-70. 10.1158/1078-0432.CCR-07-2184
5. Ayoub MA, Crépieux P, Koglin M, Parmentier M, Pin J-P, Poupon A, Reiter E, Smit M, Steyaert J, Watier H, Wilkinson T. Antibodies targeting G protein-coupled receptors: Recent advances and therapeutic challenges. *MAbs*. 2017;9(5):735-41. 10.1080/19420862.2017.1325052
6. Dodd RB, Wilkinson T, Schofield DJ. Therapeutic Monoclonal Antibodies to Complex Membrane Protein Targets: Antigen Generation and Antibody Discovery Strategies. *BioDrugs*. 2018;32(4):339-55. 10.1007/s40259-018-0289-y
7. Valeur E, Guéret SM, Adihou H, Gopalakrishnan R, Lemurell M, Waldmann H, Grossmann TN, Plowright AT. New Modalities for Challenging Targets in Drug Discovery. *Angewandte Chemie International Edition*. 2017;56(35):10294-323. 10.1002/anie.201611914
8. Phillips GDL, Li G, Dugger DL, Crocker LM, Parsons KL, Mai E, Blättler WA, Lambert JM, Chari RVJ, Lutz RJ, Wong WLT, Jacobson FS, Koeppen H, Schwall RH, Kenkare-Mitra SR, Spencer SD, Sliwkowski MX. Targeting HER2-Positive Breast Cancer with Trastuzumab-DM1, an Antibody–Cytotoxic Drug Conjugate. *Cancer Research*. 2008;68(22):9280-90. 10.1158/0008-5472.CAN-08-1776
9. Feng Y, Zhu Z, Chen W, Prabakaran P, Lin K, Dimitrov DS. Conjugates of Small Molecule Drugs with Antibodies and Other Proteins. *Biomedicines*. 2014;2(1):1-13. 10.3390/biomedicines2010001
10. Hedrich WD, Fandy TE, Ashour HM, Wang H, Hassan HE. Antibody–Drug Conjugates: Pharmacokinetic/Pharmacodynamic Modeling, Preclinical Characterization, Clinical Studies, and Lessons Learned. *Clin Pharmacokinet*. 2018;57(6):687-703. 10.1007/s40262-017-0619-0
11. Zhao RY, Wilhelm SD, Audette C, Jones G, Leece BA, Lazar AC, Goldmacher VS, Singh R, Kovtun Y, Widdison WC, Lambert JM, Chari RVJ. Synthesis and Evaluation of Hydrophilic Linkers for Antibody–Maytansinoid Conjugates. *Journal of Medicinal Chemistry*. 2011;54(10):3606-23. 10.1021/jm2002958
12. Shen B-Q, Xu K, Liu L, Raab H, Bhakta S, Kenrick M, Parsons-Reponte KL, Tien J, Yu S-F, Mai E, Li D, Tibbitts J, Baudys J, Saad OM, Scales SJ, McDonald PJ, Hass PE, Eigenbrot C, Nguyen T, Solis WA, Fuji RN, Flagella KM, Patel D, Spencer SD, Khawli LA, Ebens A, Wong WL, Vandlen R, Kaur S, Sliwkowski MX, Scheller RH, Polakis P, Junutula JR. Conjugation site modulates the in vivo stability and

- therapeutic activity of antibody-drug conjugates. *Nature Biotechnology*. 2012;30(2):184-9. 10.1038/nbt.2108
13. Wang Y, Cheetham AG, Angacian G, Su H, Xie L, Cui H. Peptide–Drug Conjugates as Effective Prodrug Strategies for Targeted Delivery. *Adv Drug Deliv Rev*. 2017;110-111:112-26. 10.1016/j.addr.2016.06.015
 14. Su D, Kozak KR, Sadowsky J, Yu S-F, Fourie-O'Donohue A, Nelson C, Vandlen R, Ohri R, Liu L, Ng C, He J, Davis H, Lau J, Del Rosario G, Cosino E, Cruz-Chuh Jd, Ma Y, Zhang D, Darwish M, Cai W, Chen C, Zhou H, Lu J, Liu Y, Kaur S, Xu K, Pillow TH. Modulating Antibody–Drug Conjugate Payload Metabolism by Conjugation Site and Linker Modification. *Bioconjugate Chemistry*. 2018;29(4):1155-67. 10.1021/acs.bioconjchem.7b00785
 15. Chen S, Lovell S, Lee S, Fellner M, Mace PD, Boggyo M. Identification of highly selective covalent inhibitors by phage display. *Nature Biotechnology*. 2021;39(4):490-8. 10.1038/s41587-020-0733-7
 16. Gower CM, Thomas JR, Harrington E, Murphy J, Chang MEK, Cornella-Taracido I, Jain RK, Schirle M, Maly DJ. Conversion of a Single Polypharmacological Agent into Selective Bivalent Inhibitors of Intracellular Kinase Activity. *ACS Chemical Biology*. 2016;11(1):121-31. 10.1021/acschembio.5b00847
 17. Cheng AC, Doherty EM, Johnstone S, DiMauro EF, Dao J, Luthra A, Ye J, Tang J, Nixey T, Min X, Tagari P, Miranda LP, Wang Z. Structure-guided Discovery of Dual-recognition Chemibodies. *Scientific reports*. 2018;8(1). 10.1038/s41598-018-25848-0
 18. Tang J, Majeti J, Sudom A, Xiong Y, Lu M, Liu Q, Higbee J, Zhang Y, Wang Y, Wang W, Cao P, Xia Z, Johnstone S, Min X, Yang X, Shao H, Yu T, Sharkov N, Walker N, Tu H, Shen W, Wang Z. An Inhibitory Antibody against Dipeptidyl Peptidase IV Improves Glucose Tolerance in Vivo. *The Journal of Biological Chemistry*. 2013;288(2):1307-16. 10.1074/jbc.M112.396317
 19. Lewis AK, Harthorn A, Johnson SM, Lobb RR, Hackel BJ. Engineered protein-small molecule conjugates empower selective enzyme inhibition. *Cell Chemical Biology*. 2022;29(2):328-38.e4. 10.1016/j.chembiol.2021.07.013
 20. Hmila I, Abdallah R BA-B, Saerens D, Benlasfar Z, Conrath K, Ayeb ME, Muyldermans S, Bouhaouala-Zahar B. VHH, bivalent domains and chimeric Heavy chain-only antibodies with high neutralizing efficacy for scorpion toxin AahI'. *Molecular Immunology*. 2008;45(14):3847-56. 10.1016/j.molimm.2008.04.011
 21. Muyldermans S. Nanobodies: Natural Single-Domain Antibodies. *Annual Review of Biochemistry*. 2013;82(1):775-97. 10.1146/annurev-biochem-063011-092449
 22. Webster CI, Caram-Salas N, Haqqani AS, Thom G, Brown L, Rennie K, Yogi A, Costain W, Brunette E, Stanimirovic DB. Brain penetration, target engagement, and disposition of the blood–brain barrier-crossing bispecific antibody antagonist of metabotropic glutamate receptor type 1. *The FASEB Journal*. 2016;30(5):1927-40. 10.1096/fj.201500078
 23. Almagro JC, Teplyakov A, Luo J, Sweet RW, Kodangattil S, Hernandez-Guzman F, Gilliland GL. Second antibody modeling assessment (AMA-II). *Proteins: Structure, Function, and Bioinformatics*. 2014;82(8):1553-62. 10.1002/prot.24567

24. Fischman S, Ofra Y. Computational design of antibodies. *Current Opinion in Structural Biology*. 2018;51:156-62. 10.1016/j.sbi.2018.04.007
25. Dang LT, Miao Y, Ha A, Yuki K, Park K, Janda CY, Jude KM, Mohan K, Ha N, Vallon M, Yuan J, Vilches-Moure JG, Kuo CJ, Garcia KC, Baker D. Receptor subtype discrimination using extensive shape complementary designed interfaces. *Nature structural & molecular biology*. 2019;26(6):407-14. 10.1038/s41594-019-0224-z
26. Kundert K, Kortemme T. Computational design of structured loops for new protein functions. *Biological chemistry*. 2019;400(3):275-88. 10.1515/hsz-2018-0348
27. Lewis SM, Kuhlman BA. Anchored Design of Protein-Protein Interfaces. *PLoS ONE*. 2011;6(6):e20872. 10.1371/journal.pone.0020872
28. Fleishman SJ, Whitehead TA, Ekiert DC, Dreyfus C, Corn JE, Strauch E-M, Wilson IA, Baker D. Computational Design of Proteins Targeting the Conserved Stem Region of Influenza Hemagglutinin. *Science*. 2011;332(6031):816-21. 10.1126/science.1202617
29. Liu X, Taylor RD, Griffin L, Coker S-F, Adams R, Ceska T, Shi J, Lawson ADG, Baker T. Computational design of an epitope-specific Keap1 binding antibody using hotspot residues grafting and CDR loop swapping. *Scientific reports*. 2017;7:41306. 10.1038/srep41306
30. Qureshi MH, Yeung JC, Wu S-C, Wong S-L. Development and Characterization of a Series of Soluble Tetrameric and Monomeric Streptavidin Muteins with Differential Biotin Binding Affinities. *Journal of Biological Chemistry*. 2001;276(49):46422-8. 10.1074/jbc.M107398200
31. Nimrod G, Fischman S, Austin M, Herman A, Keyes F, Leiderman O, Hargreaves D, Strajbl M, Breed J, Klompus S, Minton K, Spooner J, Buchanan A, Vaughan TJ, Ofra Y. Computational Design of Epitope-Specific Functional Antibodies. *Cell reports*. 2018;25(8):2121-31.e5. 10.1016/j.celrep.2018.10.081
32. Jain N, Smith SW, Ghone S, Tomczuk B. Current ADC Linker Chemistry. *Pharmaceutical research*. 2015;32(11):3526-40. 10.1007/s11095-015-1657-7
33. Murase T, Eugenio L, Schorr M, Hussack G, Tanha J, Kitova EN, Klassen JS, Ng KKS. Structural basis for antibody recognition in the receptor-binding domains of toxins A and B from *Clostridium difficile*. *The Journal of Biological Chemistry*. 2014;289(4):2331-43. 10.1074/jbc.M113.505917
34. Lim KH, Huang H, Pralle A, Park S. Stable, high-affinity streptavidin monomer for protein labeling and monovalent biotin detection. *Biotechnology and Bioengineering*. 2013;110(1):57-67. 10.1002/bit.24605
35. Sheng W, Liao X. Solution structure of a yeast ubiquitin-like protein Smt3: The role of structurally less defined sequences in protein-protein recognitions. *Protein Science : A Publication of the Protein Society*. 2002;11(6):1482-91.
36. Liu F, Zhang JZH, Mei Y. The origin of the cooperativity in the streptavidin-biotin system: A computational investigation through molecular dynamics simulations. *Scientific reports*. 2016;6. 10.1038/srep27190
37. McMahon C, Baier AS, Pascolutti R, Wegrecki M, Zheng S, Ong JX, Erlandson SC, Hilger D, Rasmussen SGF, Ring AM, Manglik A, Kruse AC. Yeast surface display

- platform for rapid discovery of conformationally selective nanobodies. *Nature structural & molecular biology*. 2018;25(3):289-96. 10.1038/s41594-018-0028-6
38. Mitchell LS, Colwell LJ. Analysis of nanobody paratopes reveals greater diversity than classical antibodies. *Protein Engineering, Design and Selection*. 2018;31(7-8):267-75. 10.1093/protein/gzy017
 39. Privett HK, Kiss G, Lee TM, Blomberg R, Chica RA, Thomas LM, Hilvert D, Houk KN, Mayo SL. Iterative approach to computational enzyme design. *Proceedings of the National Academy of Sciences*. 2012;109(10):3790-5. 10.1073/pnas.1118082108
 40. Nisthal A, Wang CY, Ary ML, Mayo SL. Protein stability engineering insights revealed by domain-wide comprehensive mutagenesis. *Proceedings of the National Academy of Sciences*. 2019;116(33):16367-77. 10.1073/pnas.1903888116
 41. Wang Q, Meng F, Xie Y, Wang W, Meng Y, Li L, Liu T, Qi J, Ni X, Zheng S, Huang J, Huang N. In Silico Discovery of Small Molecule Modulators Targeting the Achilles' Heel of SARS-CoV-2 Spike Protein. *ACS Central Science*. 2023;9(2):252-65. 10.1021/acscentsci.2c01190
 42. Makley LN, Gestwicki JE. Expanding the Number of 'Druggable' Targets: Non-Enzymes and Protein-Protein Interactions. *Chemical Biology & Drug Design*. 2013;81(1):22-32. 10.1111/cbdd.12066
 43. Nam DH, Rodriguez C, Remacle AG, Strongin AY, Ge X. Active-site MMP-selective antibody inhibitors discovered from convex paratope synthetic libraries. *Proceedings of the National Academy of Sciences of the United States of America*. 2016;113(52):14970-5. 10.1073/pnas.1609375114
 44. Voigt CA, Martinez C, Wang Z-G, Mayo SL, Arnold FH. Protein building blocks preserved by recombination. *Nature Structural Biology*. 2002;9(7):553-8.
 45. Baran D, Pszolla MG, Lapidoth GD, Norn C, Dym O, Unger T, Albeck S, Tyka MD, Fleishman SJ. Principles for computational design of binding antibodies. *Proceedings of the National Academy of Sciences*. 2017;201707171. 10.1073/pnas.1707171114
 46. Boyken SE, Chen Z, Groves B, Langan RA, Oberdorfer G, Ford A, Gilmore JM, Xu C, DiMaio F, Pereira JH, Sankaran B, Seelig G, Zwart PH, Baker D. De novo design of protein homo-oligomers with modular hydrogen-bond network-mediated specificity. *Science*. 2016;352(6286):680-7. 10.1126/science.aad8865
 47. Mandell DJ, Coutsiaris EA, Kortemme T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nature Methods*. 2009;6(8):551-2. 10.1038/nmeth0809-551
 48. Allen BD, Nisthal A, Mayo SL. Experimental library screening demonstrates the successful application of computational protein design to large structural ensembles. *Proceedings of the National Academy of Sciences*. 2010;107(46):19838-43. 10.1073/pnas.1012985107
 49. O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. *Journal of Cheminformatics*. 2011;3(1):33. 10.1186/1758-2946-3-33
 50. Mou Y, Huang P-S, Hsu F-C, Huang S-J, Mayo SL. Computational design and experimental verification of a symmetric protein homodimer. *Proceedings of the*

- National Academy of Sciences of the United States of America. 2015;112(34):10714-9. 10.1073/pnas.1505072112
51. Huang P-S, Love JJ, Mayo SL. Adaptation of a fast Fourier transform-based docking algorithm for protein design. *Journal of computational chemistry*. 2005;26(12):1222-32. 10.1002/jcc.20252
 52. Lawrence MC, Colman PM. Shape complementarity at protein/protein interfaces. *Journal of Molecular Biology*. 1993;234(4):946-50. 10.1006/jmbi.1993.1648
 53. Mayo SL, Olafson BD, Goddard WA. DREIDING: a generic force field for molecular simulations. *J Phys Chem*. 1990;94(26):8897-909. 10.1021/j100389a010
 54. Alford RF, Leaver-Fay A, Jeliaskov JR, O'Meara MJ, DiMaio FP, Park H, Shapovalov MV, Renfrew PD, Mulligan VK, Kappel K, Labonte JW, Pacella MS, Bonneau R, Bradley P, Dunbrack RL, Das R, Baker D, Kuhlman B, Kortemme T, Gray JJ. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of Chemical Theory and Computation*. 2017;13(6):3031-48. 10.1021/acs.jctc.7b00125
 55. Find pairs PyMOL Wiki2011 [Available from: https://pymolwiki.org/Find_pairs.
 56. Huang P-S, Love JJ, Mayo SL. A de novo designed protein-protein interface. *Protein Science*. 2007;16(12):2770-4. 10.1110/ps.073125207
 57. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*. 2009;6(5):343-5. 10.1038/nmeth.1318
 58. Demeestere D, Dejonckheere E, Steeland S, Hulpiau P, Haustraete J, Devoogdt N, Wichert R, Becker-Pauly C, Wouterghem EV, Dewaele S, Imschoot GV, Aerts J, Arckens L, Saeys Y, Libert C, Vandenbroucke RE. Development and Validation of a Small Single-domain Antibody That Effectively Inhibits Matrix Metalloproteinase 8. *Molecular Therapy*. 2016;24(5):890-902. 10.1038/mt.2016.2
 59. Hand K, Wilkinson MC, Madine J. Isolation and purification of recombinant immunoglobulin light chain variable domains from the periplasmic space of *Escherichia coli*. *PLOS ONE*. 2018;13(10):e0206167. 10.1371/journal.pone.0206167
 60. Hansen SB, Laursen NS, Andersen GR, Andersen KR. Introducing site-specific cysteines into nanobodies for mercury labelling allows de novo phasing of their crystal structures. *Acta Crystallographica Section D Structural Biology*. 2017;73(10):804-13. 10.1107/S2059798317013171
 61. Pleiner T, Bates M, Trakhanov S, Lee C-T, Schliep JE, Chug H, Böhning M, Stark H, Urlaub H, Görlich D. Nanobodies: site-specific labeling for super-resolution imaging, rapid epitope-mapping and native protein complex isolation. *eLife*. 2015;4:e11349. 10.7554/eLife.11349
 62. Magalhães MLB, Czekster CM, Guan R, Malashkevich VN, Almo SC, Levy M. Evolved streptavidin mutants reveal key role of loop residue in high-affinity binding. *Protein Science : A Publication of the Protein Society*. 2011;20(7):1145-54. 10.1002/pro.642
 63. Doerr S, Harvey MJ, Noé F, De Fabritiis G. HTMD: High-Throughput Molecular Dynamics for Molecular Discovery. *Journal of Chemical Theory and Computation*. 2016;12(4):1845-52. 10.1021/acs.jctc.6b00049

64. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*. 1983;79(2):926-35. 10.1063/1.445869
65. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*. 2015;11(8):3696-713. 10.1021/acs.jctc.5b00255
66. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling*. 2006;25(2):247-60. 10.1016/j.jmgm.2005.12.005
67. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, Li X, Caricato M, Marenich A, Bloino J, Janesko BG, Gomperts R, Mennucci B, Hratchian HP, Ortiz JV, Izmaylov AF, Sonnenberg JL, Williams-Young D, Ding F, Lipparini F, Egidi F, Goings J, Peng B, Petrone A, Henderson T, Ranasinghe D, Zakrzewski VG, Gao J, Rega N, Zheng G, Liang W, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Throssell K, Montgomery Jr. JA, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Keith T, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Millam JM, Klene M, Adamo C, Cammi R, Ochterski JW, Martin RL, Morokuma K, Farkas O, Foresman JB, Fox DJ. *Gaussian 09 Revision A.02*. Gaussian, Inc.; 2016.
68. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. 1996;14(1):33-8. 10.1016/0263-7855(96)00018-5
69. McGibbon Robert T, Beauchamp Kyle A, Harrigan Matthew P, Klein C, Swails Jason M, Hernández Carlos X, Schwantes Christian R, Wang L-P, Lane Thomas J, Pande Vijay S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal*. 2015;109(8):1528-32. 10.1016/j.bpj.2015.08.015