De novo design of protein binders against scorpion toxins with Claude

Claude Science[1], Nathaniel Horwitz[2], James van Geelen[3], Kat Kajderowicz[4], Amir Shanehsazzadeh[1,*]

[1]Anthropic [2]Hunterbrook [3]Citrini Research [4]Novus Capital

[*]Correspondence: ashanehsazzadeh@anthropic.com

October 20, 2026

Abstract

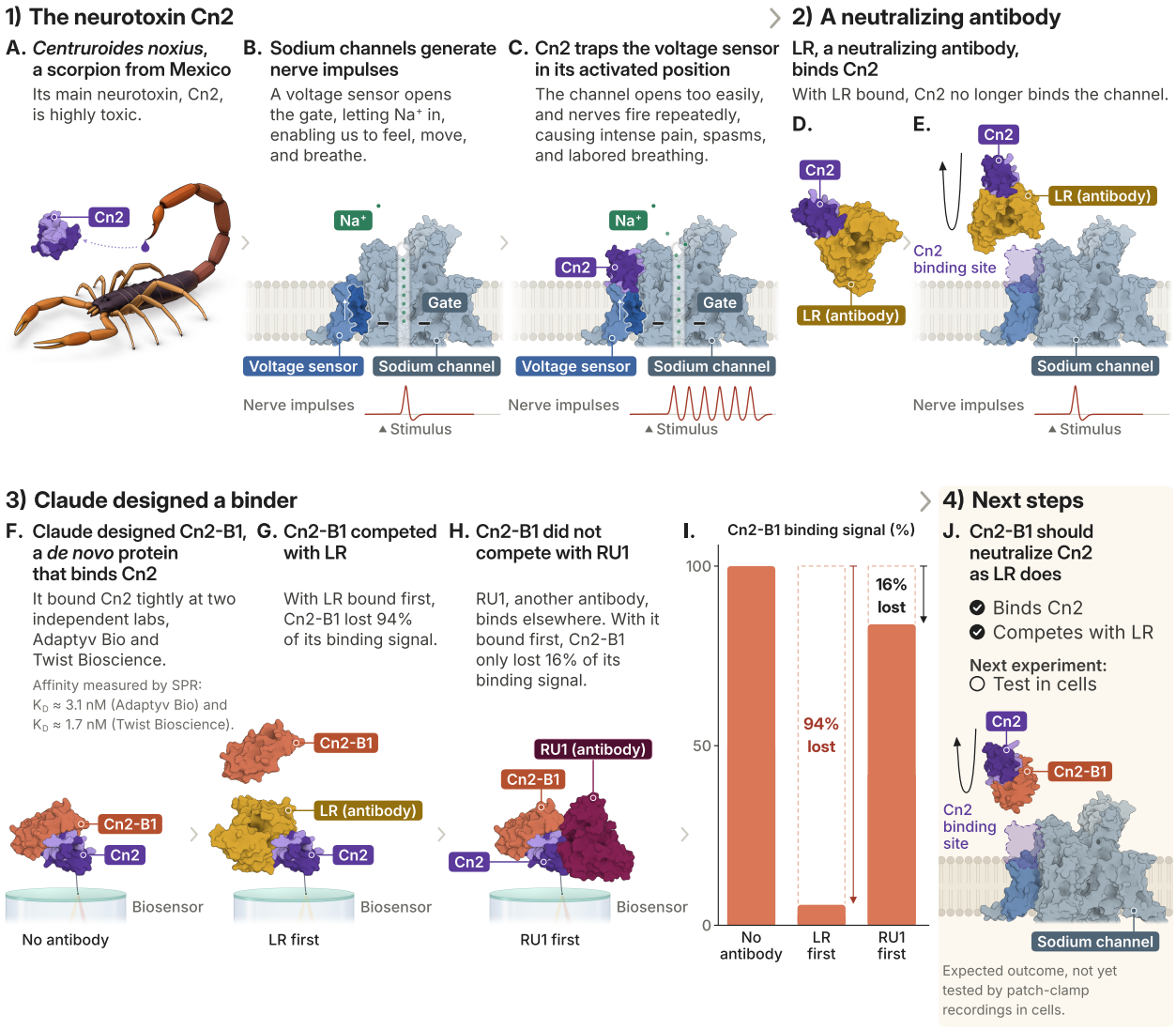
Scorpion stings are estimated to kill more than 3,250 people a year, with children the most severely affected. Here we evaluated whether Claude can design binders against medically relevant scorpion toxins. Working from an initial prompt and four follow-ups, Claude chose epitope residues on five sodium-channel toxins that we had selected (Cn2, Ts1, AaH II, AaH I, and LqhIII), generated 318,484 designs, and selected 180, to which we added 31 more with its help, for experimental characterization by Adaptyv Bio and Twist Bioscience. Of the designs tested, 17 bound a toxin they were made against and met our quality criteria. Seven designs bound Cn2, four bound AaH II, four bound AaH I, two bound LqhIII (measured at one lab only), and none bound Ts1. The tightest Cn2 binder, Cn2-B1, had single-digit nanomolar affinity or better, and a competition assay showed that it competed with LR, an antibody fragment that targets Cn2 at its channel-binding site. Claude also tried to design the multi-toxin binders our prompt called for, but none of the 151 designs made against two or more toxins met our criteria against more than one. We have not yet tested whether any binder neutralizes its toxin, either in functional assays or *in vivo*. This work shows that Claude can design protein binders with potential therapeutic relevance for scorpion envenomation, suggesting an increasingly important role for AI systems in treating neglected diseases.

Introduction

Scorpions are estimated to sting more than 1.2 million people a year and to kill more than 3,250 victims, with children most impacted [1–4]. In the Buthidae, the family that includes most scorpions dangerous to humans [5], the most medically relevant toxins act on voltage-gated sodium (Na_V) channels: β-toxins bind receptor site 4 and α-toxins bind site 3 [6]. Severe systemic effects arise downstream, from the neurotransmitters, including catecholamines, released in response [2], and it is these effects that a neutralizing binder must prevent. Most stings occur in resource-poor regions, where access to medical care is limited [7, 8]. In this report, Claude runs a binder-design campaign against these toxins with considerable autonomy, from choosing epitope residues to selecting designs to test.

Antivenoms generally rely on antibodies or antibody fragments from immunized animals. Treatments can work when given early [2, 9], but vary in composition [5, 8], are costly [2, 3], and are often limited in supply and in the venoms they recognize [10, 11]. Access differs by region: in Brazil and Mexico, antivenoms are produced domestically and reach victims through public health services [12–14], whereas cost and availability limit their use in North Africa, the Middle East, and rural India [3, 15, 16]. Recombinant antibody fragments and nanobodies against individual scorpion toxins have protected mice from those toxins and from whole venom [11, 17–20], with one human antibody fragment neutralizing an estimated 13 neurotoxins from nine Mexican species [21]. All of these came from immunized animals or libraries (Section S6), and to our knowledge, *de novo* designed binders against scorpion toxins have not been described.

Our approach was inspired by Vázquez Torres et al., who designed neutralizing proteins against three-finger toxins of elapid snakes [22]. Through *de novo* design and optimization, they produced high-affinity binders against a long-chain and a short-chain α-neurotoxin, with dissociation constants (K_D) of 1.9 nM and 0.9 nM, as measured by surface plasmon resonance (SPR). Each binder fully protected mice when given at tenfold molar excess 15 min after three median lethal doses of its toxin. They argued that stable, easily made proteins could yield cheaper, more accessible antivenoms [22]. Nanobodies are also a promising modality for antivenoms. Ahmadi et al. immunized an alpaca and a llama with the venoms of 18 African elapid snakes. They combined eight of the resulting nanobodies into a defined mixture that, premixed with venom, prevented lethality in mice for 17 of the species [23]. More recently, Overath et al. targeted



Graphical abstract. Claude designed a *de novo* protein binder against the scorpion neurotoxin Cn2. A) Drawing of *Centruroides noxius*; Cn2, the major toxic component of its venom [66], is shown as a surface from PDB 25IH. **B, C)** Cryo-electron microscopy structure of human $Na_V1.6$ with Cn2 bound (PDB 25IH; Cn2 left out in B) [44]. Cn2, a β-toxin, binds receptor site 4, on the voltage sensor of domain II. The motion of sensor and gate, and the Na^+ ions, are schematic and follow the voltage-sensor trapping mechanism [6]. Symptoms in C are those of severe *Centruroides* stings [9]. Nerve-impulse traces are schematic, not recordings. **D, E)** Crystal structure of LR, a human single-chain variable fragment (scFv) reported to neutralize Cn2 and whole venom in mice, bound to Cn2 (PDB 4V1D; RU1, bound to the same Cn2, left out) [18]. In E, the complex is placed on the channel structure by superposing the toxins (Cα root-mean-square deviation, RMSD, 0.5 Å; nothing relaxed). In that pose LR overlaps the channel (812 of its 1,741 atoms within 3 Å of a channel atom), so the pair is drawn backed off from the empty site (dashed outline). **F to H)** Competition biolayer interferometry (BLI) at Twist Bioscience (Figure 3B): Cn2 on the biosensor exposed to buffer (F), LR (G), or RU1 (H), then to Cn2-B1, a 104-residue protein designed with RFdiffusion3. Its complex with Cn2 is a predicted model (ESMFold2); no structure of it has been solved. K_D by SPR, approximately 3.1 nM at Adaptyv Bio and 1.7 nM at Twist Bioscience, both estimated to be upper limits (Table 1), which the panel shows with the approximately equal sign (≈). Antibodies are placed from PDB 4V1D by superposing the toxins and drawn as the fragments in that structure; biosensor tip and tether are schematic and not to scale. **I)** Measured: Cn2-B1's binding signal after each antibody as a percentage of its signal after buffer, with that run's IgG control subtracted from both (Figure 6D; Table S22; Section S4.1). RU1, which binds a separate epitope, reduced LR's own signal by 14%, so a loss of 16% is about what an antibody bound elsewhere on Cn2 costs. **J)** The predicted complex of F placed on the channel structure by superposing the toxins (Cα RMSD 0.5 Å). In that pose Cn2-B1 overlaps the channel (423 of its 834 atoms within 3 Å of a channel atom), as LR does in E, and covers 16 of the 20 residues of Cn2's channel-binding site; LR covers 12 in the crystal structure (Figure 6A). Expected outcome, not yet tested by patch-clamp recordings in cells.

α-cobratoxin by designing the binding loops of single-domain antibodies, a scaffold with clinical precedent. Four of the 46 tested designs bound, and two fully protected mice from the purified toxin and partly from whole venom [24].

Both design studies relied on experts to choose the target, epitope, tools, and filters. Generative tools [25–29] can now design binders from target structures [30], so these choices may be the bottleneck, and AI agents have begun to make them. In the Virtual Lab, agents guided by a researcher designed 92 nanobodies against SARS-CoV-2 variants, some of them functional [31], and agents have since designed binders against multiple targets [32–34]. Yet on 76 design tasks, agents beat fixed pipelines but trailed experts [35], and another agent was tested only computationally [36]. In our earlier report [37], Claude ran campaigns against 16 targets from a single prompt, but each design had a single target. The present study began when Citrini Research, a financial investment research firm, asked whether Claude could enable a group of motivated nonexperts to work on a genuine problem in human health [38]. Here we turned to the toxins of scorpion envenomation, which has been described as a neglected tropical disease [39]. Our prompt specified the targets, binding site, and design sets; listed candidate tools; and gave rules for novelty, scoring, and selection. Claude made the other choices autonomously (Methods).

We chose five toxins that span both toxin families (α and β) and four regions of the world (Methods; Figure 1A). They are the β-toxins Cn2 (Mexico) and Ts1 (Brazil), the α-toxins AaH II and AaH I (North Africa), and the α-like toxin LqhIII (Middle East). Antibody fragments and nanobodies have been reported against Cn2, Ts1, AaH II, AaH I, and AahI′, a natural variant of AaH I that differs at one residue [18–20, 40, 41]. To our knowledge, none has been raised against LqhIII, although serrumab, a fragment that binds Ts1, is reported to reduce LqhIII's effect on Na_V channels [40] (Section S6). We asked whether Claude could design binders against the individual toxins, and whether a single design could bind several of them. Since the toxins belong to four different species of scorpions, envenomation victims would not be exposed to all five simultaneously. We therefore used them as the hardest test of multi-toxin design rather than as a required product profile. The practical need is regional: antivenoms must cover related toxins from several species, and binders that each address multiple toxins would enable a simpler cocktail. In Mexico, for example, a product has to cover Cn2, Css2, Cll1, Cll2, Cl13, and related *Centruroides* toxins [21]. Such binders are hard to design: the five toxins we selected share one fold but only 26 to 62% pairwise sequence identity. Only 14 aligned positions, eight of them cysteines belonging to the shared fold's disulfide bonds, are identical in all five (Figure 1B,C).

Adaptyv Bio and Twist Bioscience measured binding in three ways: 1) SPR at Adaptyv Bio with the design on the chip, 2) SPR at Twist Bioscience with the design captured on the chip as an Fc fusion, and 3) reverse biolayer interferometry (BLI) at both labs with the toxin on the sensor and the design in solution. Of the 211 designs tested, 17 bound one of four toxins, and none bound Ts1. The tightest Cn2 binder, Cn2-B1, has a single-digit nanomolar K_D or tighter at both labs, and it competes with LR, an antibody fragment reported to neutralize Cn2 and whole venom in mice [18, 42, 43]. No design met our criteria for binding against more than one toxin, which could reflect Claude's capabilities, our prompting, the available design tools, or the feasibility of such binders. The 17 binders have not been tested for neutralization in a cellular environment, so we do not claim to have produced an antivenom, but we consider this work an early step toward collaborating with AI systems to address envenomation.

Methods

Targets and epitopes

We chose five toxins using three criteria: 1) they come from species whose stings cause severe envenomation in Mexico, Brazil, North Africa, or the Middle East [1]; 2) they together cover both main families of scorpion toxins that act on Na_V channels [6] (Figure 1A); and 3) they seemed obtainable as recombinant protein for binding assays.

From our prompt's reading list, Claude chose published structures for four toxins [44–47] (Figure 1D). We found no published structure of AaH I, so Claude used predicted models (Table S9). We chose to target each toxin's channel-binding site, defined as the toxin residues with a heavy atom within 5 Å of a channel. The site came from toxin-channel complexes of Cn2 (PDB 25IH) [44], AaH II (6NT4) [48], and LqhIII (7K18 [47] and 7XSU [49]; solid outlines in Figure 1E). Claude transferred the site by superposition from Cn2 to Ts1 and from AaH II and LqhIII to AaH I (dashed outlines), and added annotated residues to the sites of AaH II and Ts1 (Section S1.1). Before initiating design, Claude selected five amino acid hotspots in each site to target its designs against; some designs against multiple toxins were generated with other hotspot lists (Table S5; Section S1.1).

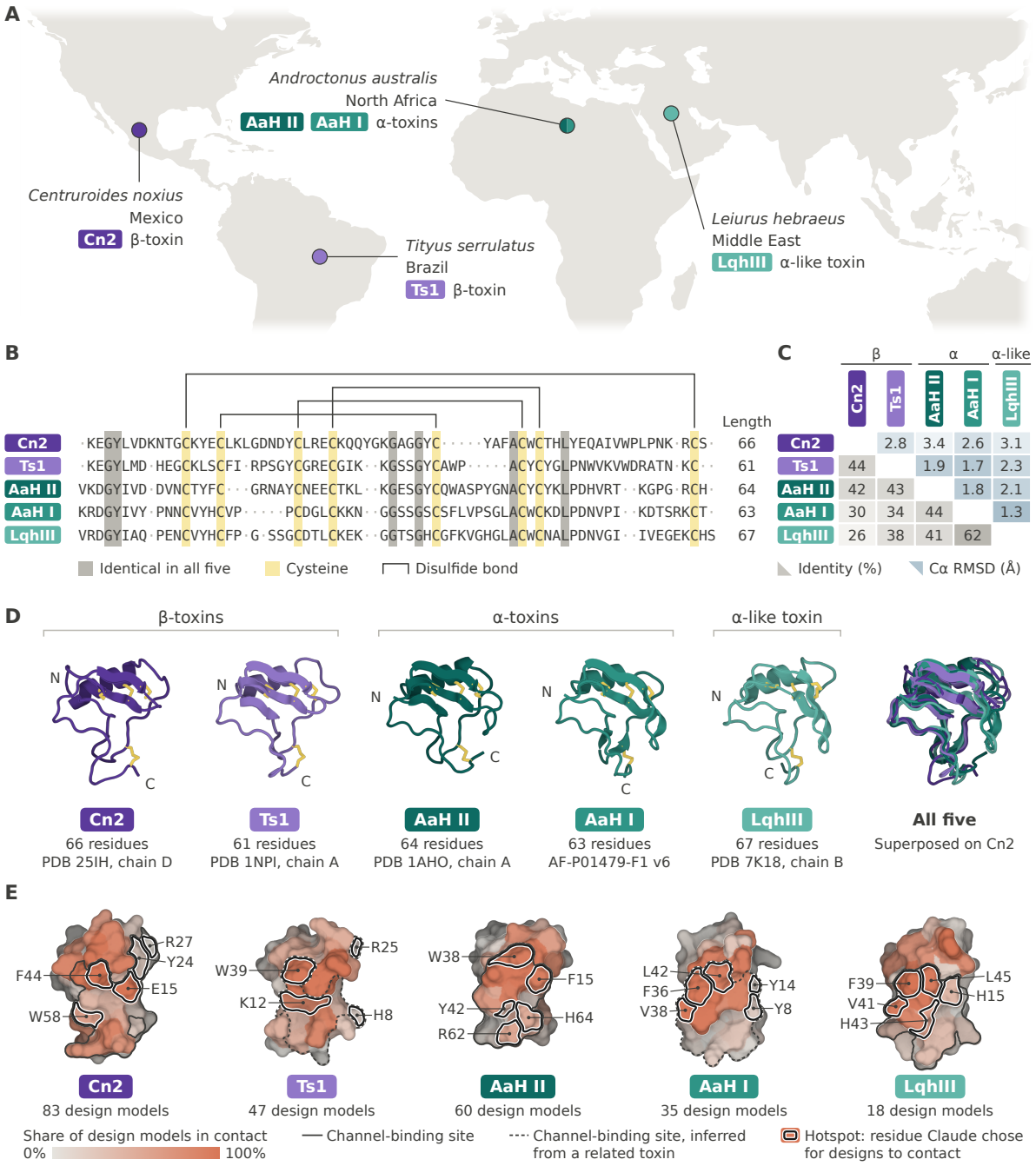


Figure 1. The five targets are related sodium-channel toxins from four scorpion species. A) Species, region, and toxin family; pins mark a country or region, not a range. **B)** Mature chains, sequences aligned; dots mark gaps. **C)** Below the diagonal, sequence identity over the shorter chain in pairwise global alignments; above it, Cα RMSD between the structures used for design, over 56 to 64 aligned residues. **D)** Structures used for design, and all five superposed on Cn2; AaH I is a predicted model; disulfide bonds in yellow. **E)** As in D, rotated by 90° and colored by the share of design models in which each residue makes contact (Table S5); one of Ts1's five hotspots, I29, is hidden.

Design campaign

An internal research model, working within Claude Science, ran the campaign, starting from our prompt and with four follow-ups (Figure 2A,B; Section S2.1). Our prompt specified the targets, the design sets and their sizes, the rules for each design, the protein design and prediction tools, the scores and ranking formulas, and the novelty and diversity

rules. The follow-ups extended the campaign and mainly directed Claude toward multi-toxin binders. Claude chose the toxin structures and hotspots, how much to use each design tool, the score thresholds, the strategies for binding multiple toxins, and the fusions' domain order and linker (Table S4). Claude produced a final list of 180 designs 64 hours after our initial prompt, and we later worked with it to add an additional 31 designs, making it a total of 211 designs.

Claude used 12 of the design tools listed in our prompt to generate backbones, or backbones and sequences: RFdiffusion [25], RFdiffusion3 [29], FreeBindCraft [27, 50], BoltzGen [28], PXDesign [51], Proteina-Complexa [52], Genie 3 [53], Mosaic [54], ESMFold2 Inversion [55], FoldCraft [56], BoltzDesign1 [57], and Protein Hunter [58]. For designing sequences on fixed backbones, Claude used SolubleMPNN [59] and SolubleCaliby, a variant of Caliby [60]. Claude then filtered out designs that resembled a toxin or a known protein, had cysteines or other sequence liabilities, made fewer than eight interface contacts, or were predicted as monomers with low confidence (Table S1). After testing, we compared all 211 designs with the five toxin structures, and none resembled a toxin in fold (Section S2.2).

Using three predictors (ESMFold2-Fast, ESMFold2 [55, 61], and Protenix v2 [62]), Claude co-folded each design with each toxin it was made against, and computed two scores for each predicted complex. The first, ipSAE [63], is a confidence score for the predicted interface. The second, self-consistency DockQ (scDockQ), is the DockQ score [64, 65] between the predicted complex and the design model, the complex that Claude kept as the design's reference pose (Section S1.1). A design passed the score thresholds against a toxin if, with all three predictors, its ipSAE reached a threshold and its scDockQ was at least 0.23. Before any design was scored, Claude set an ipSAE threshold per toxin and per predictor. Each threshold was at least 0.15 and at least the 95th percentile of 28 to 31 decoys (unrelated or scrambled proteins) and negative controls (Figure S2; Table S1). Claude generated 318,484 designs, screened 553,052 design-toxin pairs with one random seed of each predictor, and ranked 28,451 pairs with at least five seeds (Figure 2A).

Claude ranked the designs by the prompt's formulas, principally a weighted z-score of the six scores. A design against multiple toxins was ranked by its lowest per-toxin value. Claude then selected designs under the prompt's diversity rules, which capped the number from any one backbone, structural cluster, or design tool. In our fourth follow-up, we updated these rules for the all-five single domains to give the 30 places to those that passed the score thresholds against the most toxins (Sections S1.1 and S2.1).

The 211 tested designs fall into five groups (Figure 2C): 60 single-toxin designs (12 per toxin), 30 double β-toxin designs, 30 triple α-toxin designs, and 91 designs against all five toxins (59 single domains and 32 fusions of two domains). Each of the campaign's 30 fusions joined a domain made against Cn2 and Ts1 to one made against AaH II, AaH I, and LqhIII through a rigid helical linker. In a second Claude Science session run in parallel to and after the campaign, we added 31 designs against all five toxins: 2 fusions and 29 single domains built on the two backbones of the campaign's all-five single domains (Section S1.1).

Experimental characterization

Two contract research organizations (CROs), Adaptyv Bio and Twist Bioscience, measured binding by SPR, injecting toxin from low to high concentration without regenerating the surface (Figure 3A). Adaptyv Bio expressed the designs cell-free, immobilized them, and tested each against the toxins it was made against (665 pairs), and seven designs against additional toxins (133 to 140 designs measured per toxin). Each pair had two to four replicates and up to eight concentrations from 0.3 to 1,000 nM. Some replicates against Cn2, Ts1, AaH II, and AaH I were measured by BLI instead. Twist Bioscience expressed each design as a human Fc fusion in HEK cells and captured the designs on an anti-human Fc chip. They then ran 210 of the 211 designs against Cn2, Ts1, AaH II, and AaH I at six concentrations from 4.1 to 1,000 nM, once per pair.

The two teams screened the designs against commercial recombinant Cn2, Ts1, and AaH II from different suppliers, one per lab (Table S11; Section S3.3). Adaptyv Bio also ran its screen against Cn2 and AaH II from Twist Bioscience's supplier (Section S3.7). Adaptyv Bio expressed AaH I and LqhIII in-house, and Twist Bioscience's supplier custom-made their AaH I and LqhIII. Twist Bioscience did not receive LqhIII in time, so only Adaptyv Bio measured binding against it. Our positive controls (Table S12; Section S3.4) were antibodies 9004G, LR, and RU1 (Cn2), serrumab (Ts1), 4C1 (AaH II), and 9C2 (AaH I) [18, 40, 66, 67].

We define the main screen as each lab's SPR measurement of all designs. To test whether binding depended on which partner was immobilized, both teams ran reverse BLI, with the toxin on the sensor and the design in solution. Adaptyv Bio ran reverse BLI against all five toxins, and Twist Bioscience against Cn2, Ts1, AaH II, and AaH I.

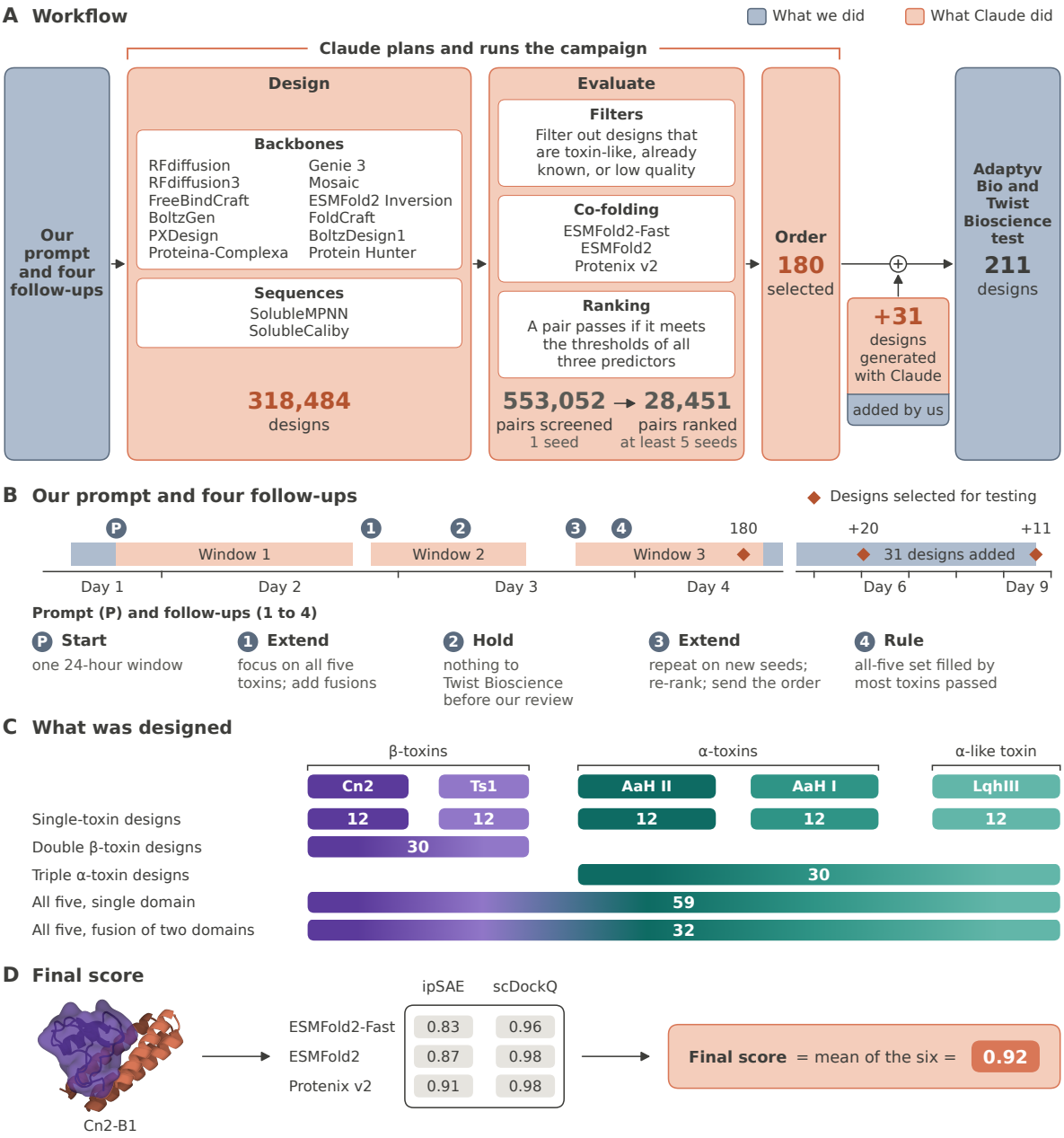


Figure 2. In response to our prompt and four follow-ups, Claude generated 318,484 designs and selected 180, and we worked with it to add 31 more. A) The 180 designs make 510 design-toxin pairs; 496 passed the score thresholds; selection did not require a pass against every intended toxin (Figure S2; Table S1). **B)** Windows, Claude's working periods. The axis is compressed from the afternoon of Day 4. **C)** Tested designs by intended toxins (colors) and format. **D)** Final score of Cn2-B1: the mean of six scores, ipSAE (interface confidence) and scDockQ (match to the design model).

Twist Bioscience used competition BLI to test where the designs bind Cn2 (Figure 3B). Sensor-bound Cn2 was exposed to an antibody, or to buffer in a reference run, and then to a design. A lower design signal after the antibody indicated competition. LR and RU1, published as fragments, neutralize Cn2 and bind non-overlapping epitopes, as shown by the crystal structure of Cn2 bound to both fragments simultaneously (PDB 4V1D) [18]. In a single run of the same assay, Twist Bioscience tested 22 designs on AaH I with 9C2, a neutralizing antibody [41] whose complex with the toxin has not been solved [67]. In Adaptyv Bio's competition assay, sensor-bound LR or 9C2 was exposed to 200 nM of its toxin premixed with 1 µM design, and lower capture than with toxin alone indicated competition (Section S4.1). To assess specificity, Twist Bioscience tested the designs in SPR for binding against CLEC12A, a human protein unrelated to the

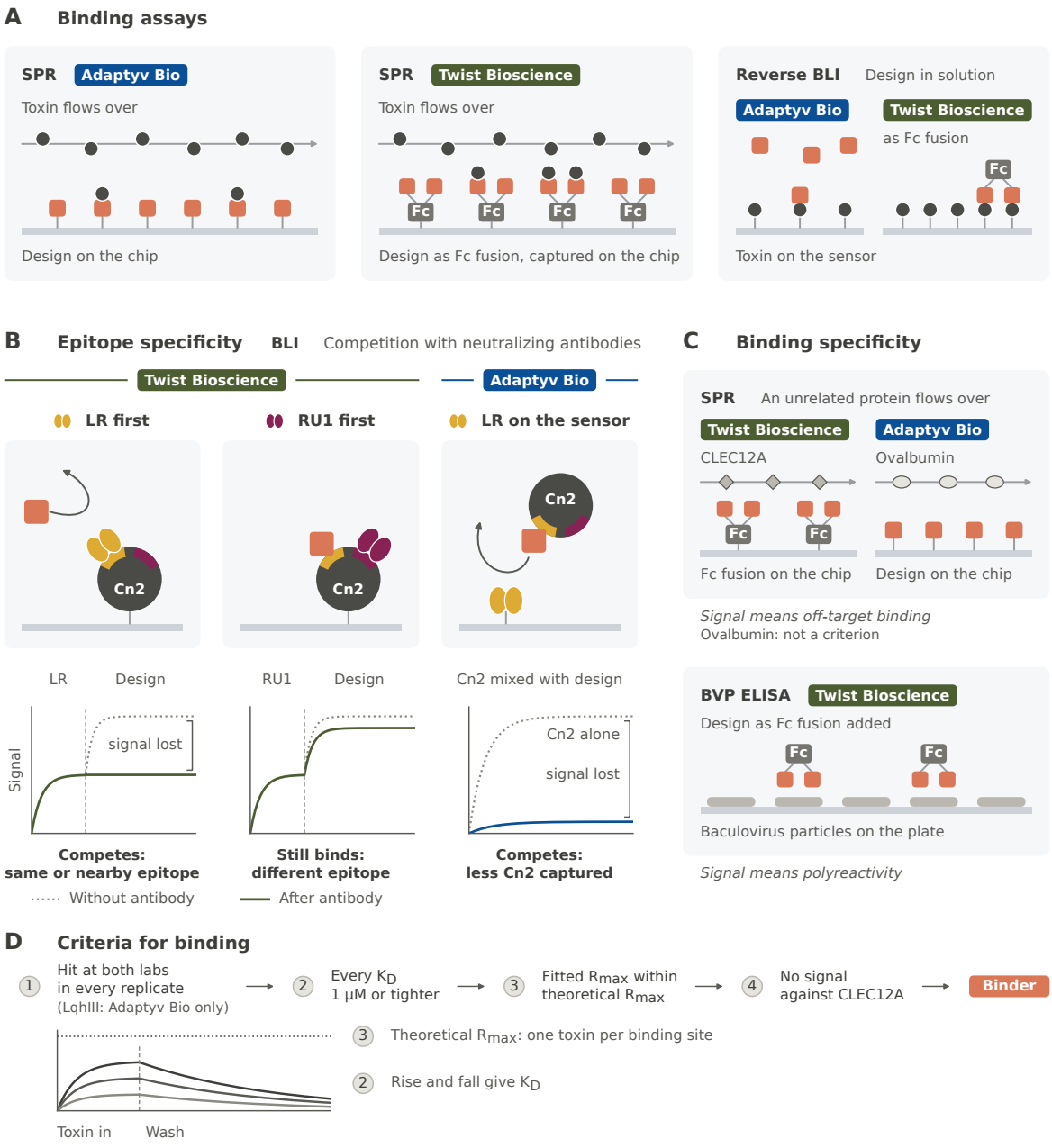


Figure 3. Adaptyv Bio and Twist Bioscience tested whether, where, and how specifically the designs bind. A) SPR and reverse BLI at Adaptyv Bio and Twist Bioscience. **B)** Competition with the neutralizing antibodies LR and RU1 on Cn2; the example competes with LR. Twist Bioscience added the antibody to sensor-bound Cn2 and then the design; Adaptyv Bio captured LR on the sensor and added Cn2 mixed with the design (Figure S22). Both teams ran their assay on AaH I with 9C2, for which no structure with the toxin has been solved. **C)** Binding against CLEC12A and ovalbumin, two unrelated proteins, and baculovirus particles; ovalbumin is not a criterion. **D)** Criteria for labeling a design a binder. For LqhIII, which only Adaptyv Bio measured, three conditions suffice: a hit at Adaptyv Bio in every replicate, a reported K_D of 1 μM or tighter, and criterion 4.

toxins. They also assessed the polyreactivity of 51 designs by baculovirus particle (BVP) ELISA [68] (Figure 3C). The Adaptyv Bio team also tested 52 designs, 48 of the 49 that were hits against any toxin at either lab and four others, in SPR for binding against ovalbumin, an egg-white protein unrelated to the toxins (Section S4.4).

We define a hit as a result at one lab indicating that a design binds a toxin or an unrelated protein (CLEC12A or ovalbumin; off-target binding). The design, or the design-toxin pair, is then a hit at that lab. Adaptyv Bio identified hits by their own analysis (Section S1.2), and for Twist Bioscience's data Claude applied a fixed rule: the response at

1 μM had to exceed the run's median for that toxin by 5 robust standard deviations. Unless stated otherwise, hits and K_D values come from the main screen. A hit against a toxin alone was insufficient for us to declare binding, so after seeing the data we set four criteria for labeling a design a binder (Figure 3D). We define a binder of a toxin as a design with 1) a hit against the toxin at both labs, in every replicate, 2) every K_D from the main screen at 1 μM or tighter, 3) a fitted maximum response (fitted R_{max}) at Twist Bioscience within the theoretical R_{max} (the largest response that one toxin molecule per binding site can give), and 4) no signal above background against CLEC12A in either of Twist Bioscience's two SPR runs. For LqhIII, measured only at Adaptyv Bio, a binder needed a hit there in every replicate, a K_D of 1 μM or tighter, and criterion 4. We define the success rate of a group of designs as the share that we labeled binders.

All K_D values come from 1:1 fits: for the designs, Adaptyv Bio's from their own fits, which our fits reproduced, and Twist Bioscience's from our fits (Section S1.2). The exception is Cn2-B1, which dissociated too slowly for its off-rate to be resolved in two of the three measurements (two replicates at Adaptyv Bio and one at Twist Bioscience), and whose traces did not fit a single-site model. At each lab we therefore report an upper limit on its K_D that uses no rate constant and is the loosest of our estimates (Section S3.9).

Results

Interface scores fell as designs targeted more toxins

Even after selection on their *in silico* scores, multi-toxin designs scored lower than single-toxin designs, which, consistent with typical protein design campaigns, had very high scores. As measured with ESMFold2-Fast, the median ipSAE per toxin was 0.84 to 0.89 for single-toxin designs and 0.74 to 0.81 for double β-toxin and triple α-toxin designs. For the two groups against all five toxins, it was 0.39 to 0.78. The single-toxin designs had the highest median for every toxin and predictor, though many ranges overlapped (Figure 4). With the three predictors, the three antibody fragments against Cn2 [18, 66] scored 0.81 to 0.91, above every Cn2 decoy (at most 0.70). Fragments against the other toxins [40, 41, 67, 69, 70] separated less clearly from decoys (Table S26; Section S5.2).

Seventeen designs each met our criteria for binding against one of four toxins

Seventeen of the 211 designs tested met the criteria for binding: seven against Cn2, four against AaH II, four against AaH I, and two against LqhIII (Figure 5; Table 1). All 17 binders are predicted to be helical proteins without β-strands, comprising 16 single domains of 74 to 118 residues with three to seven helices and one 234-residue, two-domain fusion. In their predicted complexes, the binders each bury 1,656 to 2,739 Å^2 and contact three to five of their toxin's five hotspots (Table S17). Binders among the single-toxin designs contact more toxin residues than the other single-toxin designs (medians of 24 and 19), but this difference rests on only 11 binders (Section S3.11; Figure S16).

Of the 35 design-toxin pairs that were hits at both labs, 15 met all four criteria and 20 did not (Figure 3D; Tables S13 and S14). Of the 20 that did not meet all four criteria, 15 were ruled out due to a signal above background against CLEC12A. Of the rest, three had a K_D above 1 μM against AaH II at Adaptyv Bio, and one exceeded the theoretical R_{max} against Ts1. The last, design 9a8cdc7a against Cn2, was a hit in only one of three replicates at Adaptyv Bio and so did not meet criterion 1.

In reverse BLI, with the toxin on the sensor, all 17 binders were hits against their toxin at Adaptyv Bio. At Twist Bioscience, where we read the assay only as hit or no hit, 12 of the 15 binders tested were hits (6 of 7 against Cn2, 4 of 4 against AaH II, and 2 of 4 against AaH I). Against AaH II at Adaptyv Bio, 10 of the 11 other designs tested were not hits, including two of the three with a K_D above 1 μM. Against AaH I, however, 19 of the 21 designs tested were hits (Section S3.10).

In the control against ovalbumin at Adaptyv Bio, Cn2-B1 and AaHI-B4 were hits, while the other 15 binders, including both LqhIII binders, were not (Table 1; Section S4.4). Ovalbumin is not among the four criteria, and requiring no hit against ovalbumin as well would leave 15 of the 17 binders (13 of the 15 pairs in Table S16).

Control antibodies bound Cn2 (9004G, LR, and RU1, which binds a separate epitope) and AaH I (9C2) at both labs, consistent with folded toxin in these preparations (Section S3.4). Serrumab gave a signal against Ts1 at Adaptyv Bio, but so did antigen-binding fragments (Fabs) of four antibodies raised against other toxins (Figure S4), so this control is inconclusive (Section S3.4). No specific binding was detected for 4C1 against AaH II at either lab or for Twist

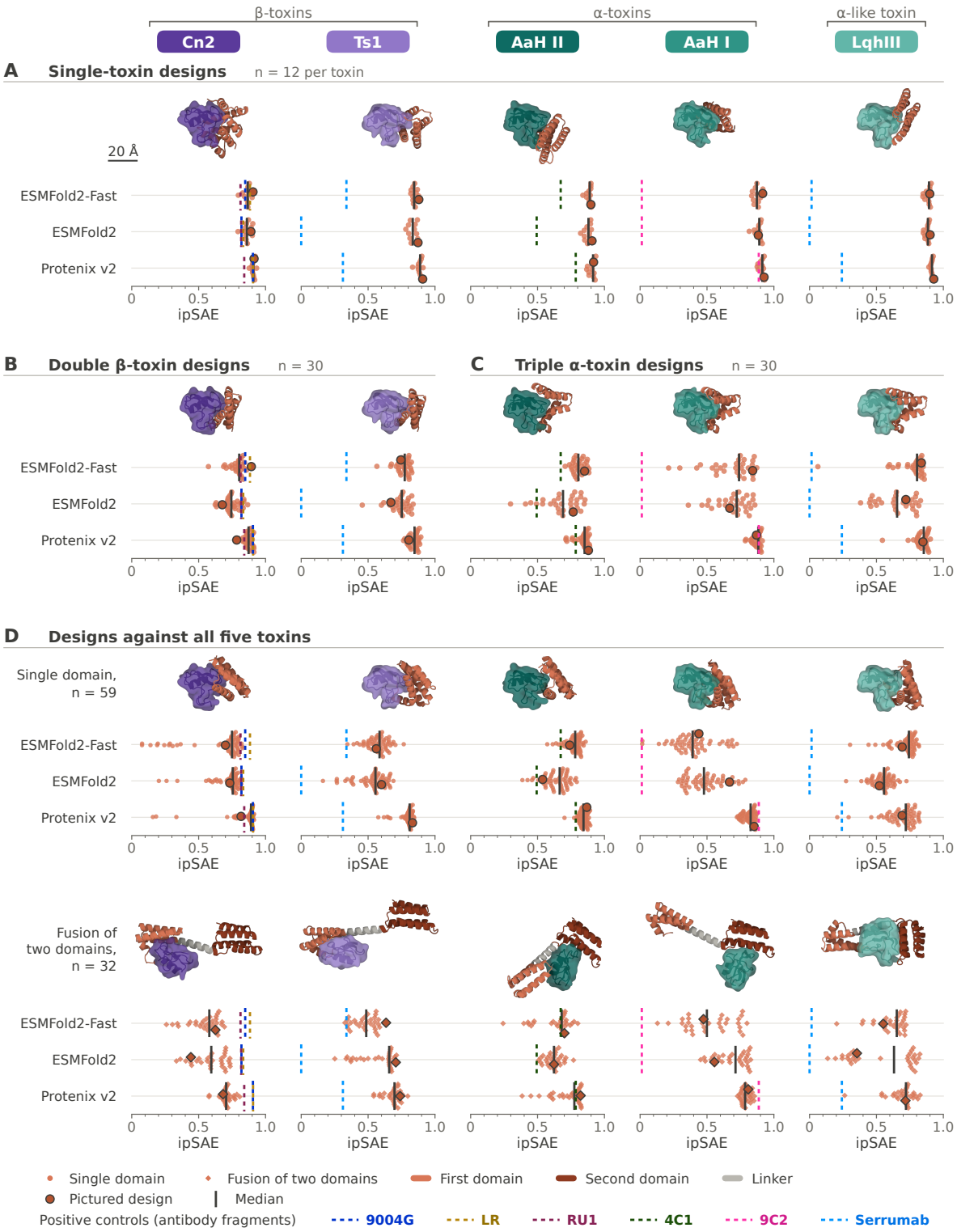


Figure 4. Single-toxin designs scored highest with all three structure predictors. A to D) Single-toxin, double β-toxin, and triple α-toxin designs, and designs against all five toxins. Images, ESMFold2 models of each group's top design by final score (for multi-toxin designs, the lowest over their toxins), at one scale. Strip plots, ipSAE, best of five seeds, one point per design; bars, medians. Dashed lines, positive controls (antibody fragments): 9004G, LR, and RU1 (Cn2), 4C1 (AaH II), 9C2 (AaH I), and serrumab (Ts1, LqhIII); only the first four have crystal structures of their complexes.

Table 1. Seventeen designs met the criteria for binding against one toxin each, 15 of them hits at both labs. Designs are named by toxin and ranked by K_D at Twist Bioscience (for LqhIII, Adaptyv Bio). K_D by SPR: at Adaptyv Bio, mean ± standard deviation of the fitted replicates (two; three for Cn2-B3, Cn2-B5, and the LqhIII binders; four for AaHII-B4); at Twist Bioscience, one measurement. ≤, upper limit; ≈, approximate. Sensorgrams, Figure S8; rate constants, Table S18.

				K_D (nM)	
Design	Made against	Design tool	Residues	Adaptyv Bio	Twist Bioscience
Cn2					
Cn2-B1‡	Cn2	RFdiffusion3	104	≤3.1[a]	≤1.7[a,†]
Cn2-B2	Cn2	Mosaic	81	5.6 ± 1.8	3.4[†]
Cn2-B3	Cn2	Mosaic	103	162 ± 45	40
Cn2-B4	Cn2	Genie 3	118	112 ± 13	48
Cn2-B5	Cn2, Ts1	PXDesign	110	89 ± 28	50
Cn2-B6	Cn2, Ts1	PXDesign	97	63 ± 17	54
Cn2-B7	Cn2	Genie 3	74	350 ± 76	≈580[b]
AaH II					
AaHII-B1	AaH II	RFdiffusion3	95	78 ± 1	1.7[†]
AaHII-B2	three α-toxins	RFdiffusion3	74	176 ± 9	3.8[†]
AaHII-B3	three α-toxins	RFdiffusion3	74	200 ± 8	3.9[†]
AaHII-B4	AaH II	Genie 3	109	21 ± 3	6.3
AaH I					
AaHI-B1	AaH I	RFdiffusion3	75	≈6.5 ± 4.6[c,†]	26
AaHI-B2	AaH I	Proteina-Complexa	93	≈63 ± 76[c]	58
AaHI-B3	all five	PXDesign, Genie 3 (fusion)	234	≈59 ± 59[c]	100
AaHI-B4‡	three α-toxins	Genie 3	107	≈90 ± 87[c]	180
LqhIII (measured at Adaptyv Bio only)					
LqhIII-B1	LqhIII	Proteina-Complexa	99	4.3 ± 1.8	not measured
LqhIII-B2	LqhIII	Genie 3	78	41 ± 19	not measured

[a]Off-rate not resolved. [b]Near the highest concentration tested. [c]Mean of two replicates, one by BLI except for AaHI-B2, that differ 3- to 13-fold. [†]Below the lowest concentration tested (4.1 nM at Twist Bioscience; for AaHI-B1, 31.6 nM in the BLI replicate: 3.2 nM, against 9.7 nM by SPR). [‡]A hit against ovalbumin (Section S4.4). Two pairs of designs share a backbone first made against AaH I: AaHII-B2 and AaHII-B3, and AaHI-B4 and one domain of AaHI-B3. The backbone of Cn2-B5 was first made against Ts1.

Bioscience's anti-Ts1 antibody, and none of the nine fragments tested against LqhIII, all raised against other toxins, was a hit (Table S11).

Cn2 binders competed with neutralizing antibody LR

Cn2-B1 and Cn2-B2 were the tightest Cn2 binders, both having single-digit nanomolar affinity or better: Cn2-B1 had K_D values of approximately 3.1 nM at Adaptyv Bio and 1.7 nM at Twist Bioscience, both estimated to be upper limits, and Cn2-B2 had 5.6 and 3.4 nM (Figure 6; Table 1). Both dissociated slowly, with off-rates below 10^{-3} s^{-1} at both labs, compared with 4.9×10^{-3} s^{-1} or faster for the other five Cn2 binders, whose K_D values were 40 nM or more (Table S18). The two labs' K_D values differed by a factor of 1.2 to 4.0. In reverse BLI at Adaptyv Bio, 12 of the 13 designs tested against Cn2 were hits, and Cn2-B1 again ranked first, at about 6.3 nM compared with 40 to 2,120 nM for the other 10 with a fitted K_D (Section S3.10).

LR, a neutralizing antibody whose epitope overlaps the site we chose to target, competed with the six Cn2 binders that could be assessed. With LR bound to Cn2 first, all six lost 87 to 95% of their binding signal, while the seventh, Cn2-B7, gave too little binding signal to assess even without antibody. RU1 binds a separate non-overlapping epitope and reduced the signal of five binders by only 15 to 22%, close to the 14% by which it reduced the LR signal. Cn2-B2 lost 66% of its binding signal, consistent with its predicted complex (ESMFold2). In that model, Cn2-B2 overlaps RU1 (Figure S19) and contacts 6 of the 17 residues of RU1's epitope, more than any other Cn2 binder (0 to 4; Figure 6A,D; Table S22; Section S4.1). Seven designs that we did not label binders showed the same pattern, losing 58 to 88% of their signal after LR and at most 36% after RU1. Some may bind Cn2 very weakly at the targeted site (Figure S20). Competition thus places every binder that could be assessed on or near LR's epitope.[1]

The Cn2 binders were not hits against Ts1, AaH II, AaH I, or CLEC12A in SPR (Section S4.3). Of the seven, only Cn2-B1 was a hit against ovalbumin, reaching 13% of the theoretical R_{max} at 1 µM ovalbumin without a plateau, a level it reached against Cn2 at about 0.5 nM, a concentration about 2,000 times lower. Ovalbumin is acidic (isoelectric point

[1]In Adaptyv Bio's competition assay, Cn2-B2 also competed with LR (5 and 18% of the capture seen with toxin alone). With Cn2-B1, no capture was seen, but its traces fell below the baseline, and the assay cannot assess the weaker binders (Section S4.1).

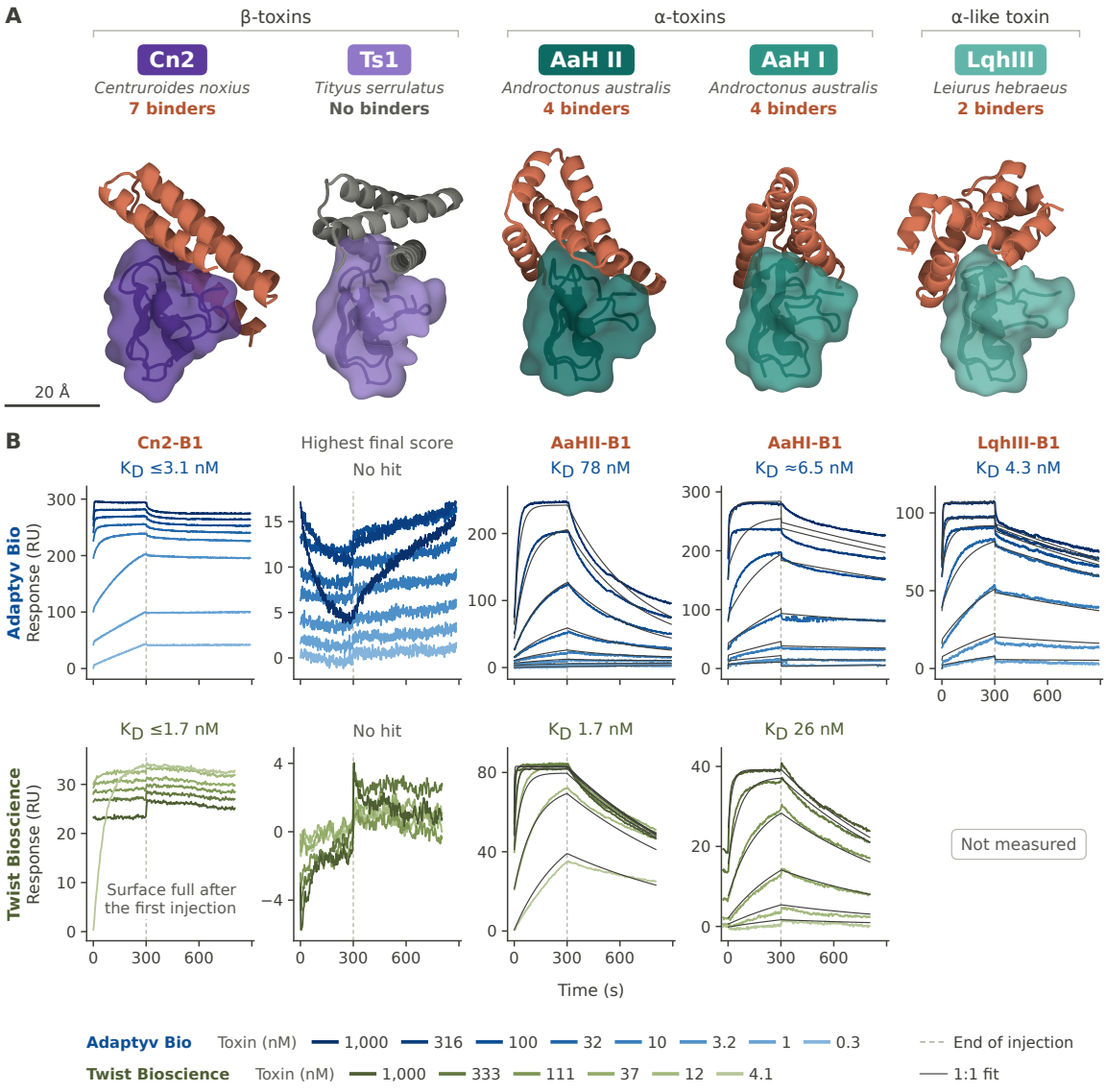


Figure 5. Claude's designs included binders against four of the five toxins, but none met the criteria against Ts1. A) Predicted complexes (ESMFold2): toxin as surface; design as ribbon, orange for the binder that Table 1 lists first, gray for the Ts1 design with the highest final *in silico* score. **B)** SPR sensorgrams (design on the chip; RU, response units), one replicate per lab, each on its own scale; without regeneration, each injection starts where the last ended. K_D values are those of Table 1 and can differ from the fit of the replicate shown; ≤, upper limit (no fit drawn); ≈, approximate. Twist Bioscience did not measure binding against LqhIII.

4.5) [71], and Cn2-B1 is the only binder that is near neutral rather than acidic (calculated isoelectric point 7.8), so charge may contribute to this response (Figure S27; Section S4.4). Twist Bioscience's BVP ELISA for polyreactivity, scaled to the highest-reading control antibody, placed five of the seven, including Cn2-B1 (29%), within the range of 12 reference nanobody-Fc fusions (28 to 94%), with Cn2-B2 (98%) and Cn2-B3 (108%) above it (Figure 6E; Section S4.2).

Ten more designs bound AaH II, AaH I, or LqhIII

For all ten binders, hits recurred across toxin preparations or replicates. Four designs bound AaH II, and none of the 73 designs against other toxins that could be assessed at Twist Bioscience was a hit against it. The four AaH I binders bound two independently produced toxin preparations, one per lab. Binding against LqhIII was measured only at Adaptyv Bio, where each of the two LqhIII binders bound in all three replicates.

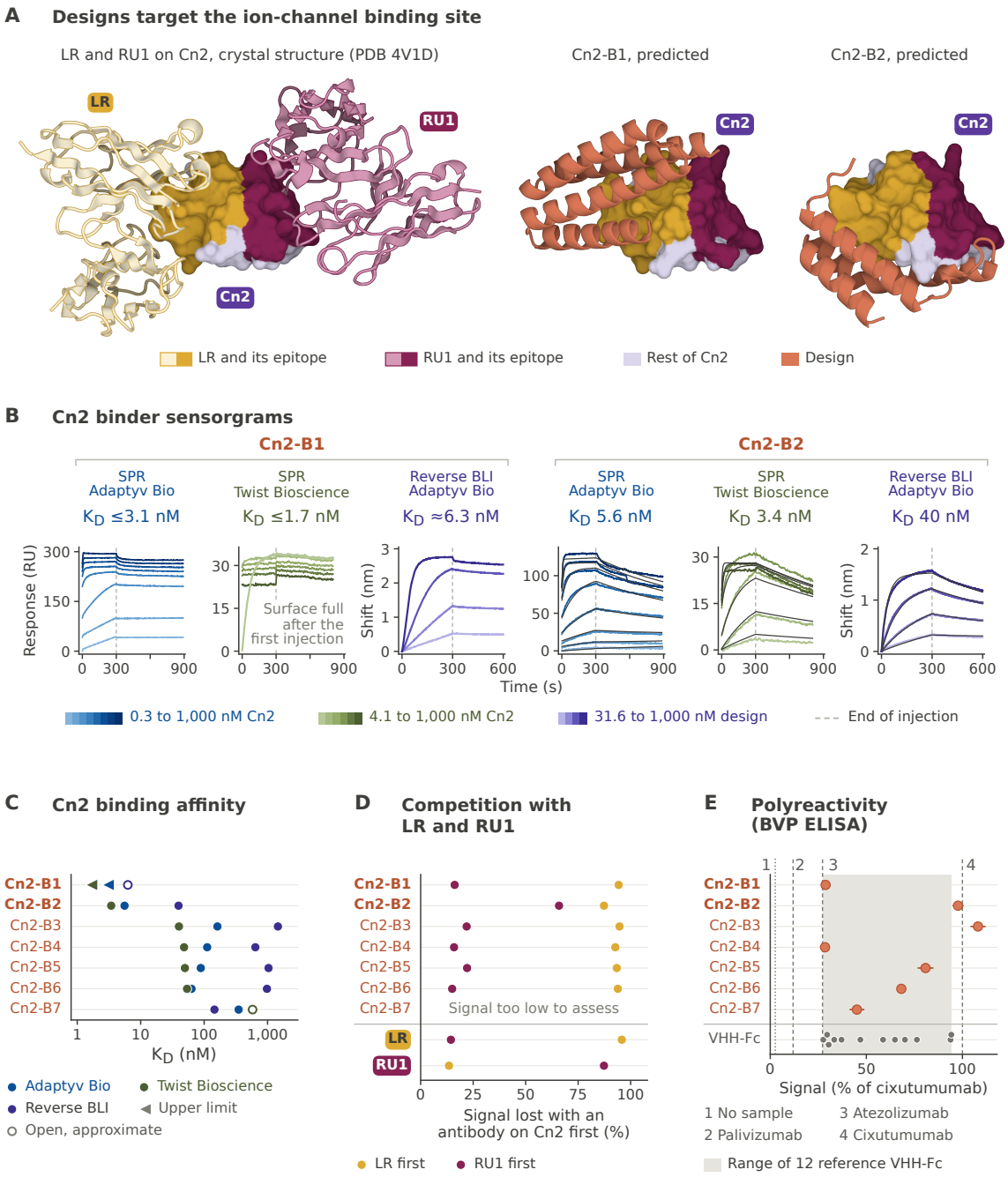


Figure 6. Cn2 binders, two with a single-digit nanomolar K_D or tighter, competed with neutralizing antibody LR. A) Crystal structure of LR and RU1 bound to Cn2, and predicted complexes (ESMFold2). Of the 20 residues of Cn2's channel-binding site, Cn2-B1 covers 16, Cn2-B2 9, LR 13 (12 in the structure shown, 13 when pooled with its parent 9004G), and RU1 6 (contacts within 4.5 Å; Section S4.1). **B)** Sensorgrams, one replicate per assay, each under its K_D (≤, upper limit; ≈, approximate); thin lines, 1:1 fits of Cn2-B2. **C)** K_D by assay. **D)** Competition BLI at Twist Bioscience: signal lost with LR or RU1 bound to Cn2 first; bottom, the antibodies as samples. **E)** BVP ELISA at Twist Bioscience: mean and range of two readings; gray dots and band, 12 reference nanobody-Fc fusions (VHH-Fc) and their range; numbered lines, no sample and three control antibodies.

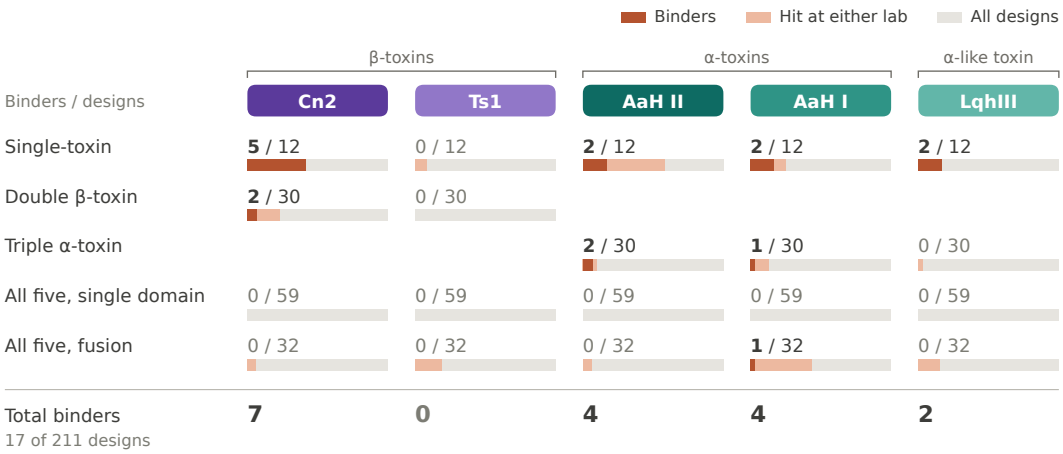


Figure 7. Single-toxin designs were labeled binders most often, and each binder met the criteria against one toxin. For each design group and each toxin it was made against, gray bars show all its designs against that toxin, pale orange the share that were hits at either lab in the main screen (binders included), and dark orange the share labeled binders. Counts give binders / designs. Empty cells mark toxins a group was not made against. Totals count each of the 17 binders once. Only Adaptyv Bio measured binding against LqhIII, so the two LqhIII binders are not directly comparable with the rest.

The neutralizing antibody 9C2 competed with AaHI-B1 and AaHI-B2, which lost 80% (range, 74 to 83%) and 89% (range, 80 to 93%) of their signal, respectively, when 9C2 was bound to AaH I first (Section S4.1). Twist Bioscience's run could not assess AaHI-B3 and AaHI-B4, which were not hits in reverse BLI there. In Adaptyv Bio's competition assay, 9C2 also competed with AaHI-B1 (6 and 11% of the capture seen with toxin alone), AaHI-B2 (25 and 37%), and AaHI-B4, the last in one of two runs. Because reads of toxin alone in these runs ranged from 33 to 106% of their median, only AaHI-B1 stands well clear of them.

Both LqhIII binders were hits in reverse BLI, as was one of the other eight designs tested. Neither LqhIII binder was a hit in the control against ovalbumin, and of all ten binders only AaHI-B4 was. Against AaH I, AaHI-B4 reached the level it gave at 1 μM ovalbumin at a concentration about 20 times lower (one SPR replicate; Section S4.4).

Between labs, K_D values differed 3- to 51-fold for AaH II binders and 1.1- to 4.1-fold for AaH I binders, although Adaptyv Bio's two replicates against AaH I differed 3- to 13-fold (Table 1; Section S3.2). The gaps for AaH II lay mainly in on-rates. They remained with toxin from Twist Bioscience's supplier (Section S3.7) and may reflect the designs' different formats and surfaces (Section S3.6; Figure S10). In size-exclusion chromatography at Twist Bioscience, the main peak held 90.8 to 100% of the signal for all 17 binders (Table S7), which argues against aggregation there, but we have no comparable measurement for Adaptyv Bio's cell-free material.

No design met the criteria for binding against Ts1

The few hits against Ts1 may reflect nonspecific signal. At Twist Bioscience, all five hits exceeded the theoretical R_{max}, our limit, with fitted values 1.5 to 4.2 times it. At Adaptyv Bio, Fabs of four antibodies raised against other toxins (9004G, LR, RU1, and 9C2) gave signals against Ts1, with mean apparent K_D values of 18 to 66 nM, at or above serrumab's 12 to 17 nM (Figure S4). In the main screen, the single-chain variable fragments (scFvs) of 9004G, LR, and RU1 and both formats of 4C1 gave none, and Adaptyv Bio attributes the Fabs' signals to the reformatting of these antibodies as Fabs with human constant domains, not to recognition of Ts1. In reverse BLI against Ts1, all 10 designs tested at Adaptyv Bio were unresolved, and all 14 tested at Twist Bioscience gave a signal, including six that were not hits against it in SPR at either lab, while serrumab did not bind. We therefore do not read these signals as binding (Section S3.10). Weak binding could have been missed at Adaptyv Bio, where half the replicates against Ts1 gave responses below −5 response units (RU) at 1 μM (Section S3.4). Both labs' Ts1 was expressed in *E. coli*, where its four disulfide bonds may not form correctly [72]. Its C-terminal tag may have occluded the C-terminal residue that 11 of 12 single-toxin and 28 of 30 double β-toxin designs contact in predicted complexes, although the AaH I and LqhIII constructs from Adaptyv Bio also carry C-terminal tags and designs bound them (Table S11; Section S3.3).

No design met the criteria for binding against multiple toxins

None of the 151 multi-toxin designs met the criteria for binding against more than one toxin (Figure 7). The share of designs labeled binders fell with the number of toxins, from 11 of 60 single-toxin designs (18%) to 5 of 60 double β-toxin and triple α-toxin designs (8%) and 1 of 91 all-five designs, the fusion AaHI-B3. Three of the five binders among the double- and triple-toxin designs (Cn2-B5, AaHII-B2, and AaHII-B3) came from backbones first made against a related toxin (Ts1 or AaH I; Table 1).

Nine designs were hits against multiple toxins in the main screen. Seven were fusions, all of which were also hits against CLEC12A (Figure S24A), suggesting mostly off-target binding. Among the binders, AaHII-B4 was a hit against Ts1 at both labs and may bind it, but its fitted R_{max} at Twist Bioscience was 4.2 times the theoretical R_{max}, so we label it a binder of AaH II only (Sections S3.8 and S4.3).

Discussion

Of the 211 designs tested, 17 met our criteria for binding, including at least two of 12 single-toxin designs against each of four of the five toxins. Two Cn2 binders had single-digit nanomolar affinity or better by SPR at both labs. The neutralizing antibody LR, whose epitope overlaps the site we chose to target, competed with all Cn2 binders that could be assessed, whereas RU1, with a separate epitope, had little effect on the majority (Table S22). The final *in silico* score (Figure 2D) tracked success rates across design sets but did not pick out binders within a set (Section S5). In hindsight, the mean of its ipSAE terms did so modestly, in line with a meta-analysis of tested binders [73], but with so few binders an advantage over the final score could not be established.

Published antibody fragments and nanobodies remain the standard to beat for designed binders: three have reported sub-nanomolar K_D values of 0.11 to 0.70 nM against Cn2 or AaH II [19, 21, 43], compared with approximately 3.1 nM at Adaptyv Bio and 1.7 nM at Twist Bioscience, both estimated to be upper limits, for Cn2-B1. Those against Cn2, AaH II, AaH I, and its variant AahI′ protected mice in challenge studies [18–20, 41, 42] (Table S27). Our success rate, 11 binders among 60 single-toxin designs (18%, 0 to 42% per toxin), is consistent with our earlier protein design campaigns [37].

Each of the 17 binders met our criteria against a single toxin (Figure 7). This could reflect Claude's current capabilities, our prompting, the available design tools, or the biological feasibility of multi-toxin binders. Just 1 of the 91 all-five designs was a binder, but only against one toxin. The multi-toxin designs were also a limited test: the 59 all-five single domains share two backbones; the fusions joined domains that were not labeled binders as single domains or were never tested as such; and the 30 double β-toxin designs depended on Ts1, against which no design met the criteria (Section S1.1). In predicted complexes, typically fewer than half the positions a design contacts in every intended toxin carry the same residue, so one surface must accommodate different residues (Section S3.11). Predicted alone with ESMFold2, only 1 of the campaign's 30 all-five single domains came within 2.0 Å of its design model, compared with 107 of the 120 designs against one to three toxins (Figure S28). Because 28 of those 30 share one backbone, this is close to a single case, but it suggests that selecting on co-folding scores against five toxins may favor binding surfaces that are not pre-formed. Two backbones gave three binders against a relative of the toxin they were first designed against, so backbones can transfer between related toxins. Other routes with precedent include binder cocktails [23, 74], consensus toxins as antigens [75, 76] and design targets [22], and linking proven binders, as in the two linked nanobodies that protected mice from scorpion venom [20]. Odin-Multi [77] and BindCraft2 [78] now also demonstrate multi-target binder design capabilities.

Limitations

- We measured binding but tested no design for neutralization *in vitro* or efficacy *in vivo* (Table S28). The toxins were recombinant proteins of undocumented activity, not toxin purified from venom. Positive controls worked at both labs for only Cn2 and AaH I (Section S3.4).
- The affinities should be read as preliminary, since each K_D rests on one concentration series at Twist Bioscience and on two to four fitted replicates at Adaptyv Bio (mean and standard deviation in Table S18). The SPR series were run without regenerating the surface, some show baseline drift, and surface effects could distort the rate constants (Section S3.6), so firmer values will need cleaner sensorgrams from an optimized assay with independent replicates.
- The four AaH I binders meet the theoretical R_{max} criterion at the toxin mass we derived from Twist Bioscience's values, but not at the lower mass of the mature toxin (Table S10).

- Competition could only be assessed for Cn2 and the neutralizing antibodies LR and RU1, and for three of the four AaH I binders and the neutralizing antibody 9C2. The runs on AaH I used only this one antibody, which has no solved structure on the toxin, so they place no design on a known site (Section S4.1).
- LR also competed with seven designs that gave weak responses against Cn2 and did not meet the criteria for binding, suggesting some may bind Cn2 very weakly (Figure S20). Our label is a definition, and calls near its edge could go either way.
- With the toxin on the sensor, all 17 binders were hits at Adaptyv Bio but only 12 of the 15 tested were hits in this assay at Twist Bioscience. Ts1 aside, 28 design-toxin pairs that we do not label binders were also hits (Section S3.10).
- Off-target binding was tested against two unrelated proteins, CLEC12A and ovalbumin, and two binders, including the tightest, were hits against ovalbumin (Section S4.4). Polyreactivity was tested by BVP ELISA for all 17 binders, each in one run at one concentration.
- Immunogenicity, serum half-life, production cost, thermal stability, and the benefit of selection over unselected single-toxin designs were not measured.
- We also ran no design baseline, such as experts or a fixed pipeline using the same tools and resources, so we cannot separate Claude's contribution from that of its tools or say how much time or cost the approach saves over a conventional campaign.

Future directions

Cn2-B1, which binds with single-digit nanomolar affinity or better and competes with the neutralizing antibody LR, is one of our strongest candidates to advance toward a therapy. We expect it to neutralize Cn2 in the same way as LR, by keeping the toxin from its channel, although this has yet to be confirmed experimentally. First, any binder that moves forward should be shown to bind toxin purified from venom, which would also reveal whether the absence of Ts1 binders reflects the designs or the recombinant toxin. Second, in patch-clamp recordings, the binder must stop a β-toxin from shifting Na_V activation or an α-toxin from slowing inactivation without altering the currents itself [6]. Third, it should be tested in mice, premixed with toxin, then with venom, and finally given after venom, as published binders were [8, 79]. Multi-toxin binding can be tested without new designs, by measuring the 17 binders against close relatives of their targets. Beyond binding and neutralization, a treatment must also be safe, stable, affordable, and available where stings occur. Future campaigns could put such properties into the prompt as design goals alongside binding. Envenomation is clinically complex; reaching patients will also take the work of clinicians, funders, and policymakers.

In a single campaign, Claude designed binders against four of five scorpion toxins, including two Cn2 binders with single-digit nanomolar affinity or better, but binding multiple toxins with one design was unsuccessful. Our campaign relied on cloud compute and two CROs (Section S2.1), using on the order of 10,000 H100 GPU hours, and Claude went from our initial prompt to the final list of 180 designs in 64 hours with four follow-ups from us. Compute on that scale is more than most researchers can access, and we have worked to reduce that cost substantially. In separate work [80],[2] Claude optimized more than 30 open-source biomolecular models, including the three structure predictors and most of the design tools used here, making structure prediction about 4 times faster on average with no measurable change in accuracy. With these models, a simplified prompt, and one GPU for 24 hours, Claude produced designs whose *in silico* scores were comparable to those of earlier campaigns that could use about 100 times as many GPU hours. That comparison rests on *in silico* scores alone, and the present campaign used neither advance, but future campaigns can draw on both.

We hope that this demonstration of Claude's capabilities, and the 211 designs and binding data we are releasing, will help accelerate further research into antivenom design.

Data availability

The data release holds the sequences of all 211 designs, how each was made, its predicted structures and scores as a monomer and with each toxin, and its interfaces with its intended toxins. The release also holds both labs' binding measurements, hits, kinetic fits, and sensorgrams; the competition, BVP ELISA, ovalbumin, reverse BLI, and second-supplier data; the five toxins; and the positive controls in both formats. The release is available under a CC BY 4.0 license at `https://huggingface.co/datasets/Anthropic/scorpion-toxin-binder-design`.[3]

[2] See also `https://www.anthropic.com/research/claude-uplifts-biomolecular-modeling`.
[3] See also `https://www.anthropic.com/research/claude-scorpion-toxin-binder-design`.

Acknowledgments

We thank the teams at Adaptyv Bio and Twist Bioscience for expression, purification, and binding measurements performed as contract research, and for answering questions about assay conditions and controls. We thank Susana Vázquez Torres (Spanish National Cancer Research Centre, CNIO) and Timothy P. Jenkins (Technical University of Denmark) for discussions, advice, and review, and Judy Tse and Colby Souders (Twist Bioscience); Gijs de Bruin, Daniel Nakhaee-Zadeh Gutierrez, Xavier Pierrat, Ron Gill, and Elizaveta Kozlova (Adaptyv Bio); and Daniel Bedinger (Carterra) for reviewing this report. We thank Kyle Zinn and Felix Enocksson (Citrini Research) and Stephen Malina, Nathan Frey, Richard Shuai, Vincent Fan, Janani S. Iyer, Xander Balwit, Caitlin Love, Stuart Ritchie, Rebecca Hiscott, Heather Whitney, and Lindsay Blohm (Anthropic) for helpful discussions and review.

Competing interests

Claude Science and Amir Shanehsazzadeh are affiliated with Anthropic, Nathaniel Horwitz with Hunterbrook, James van Geelen with Citrini Research, and Kat Kajderowicz with Novus Capital. Adaptyv Bio and Twist Bioscience were contracted and paid by Anthropic and Citrini Research to carry out the experimental validation described in this report. The authors from Hunterbrook, Citrini Research, and Novus Capital helped conceive, plan, and execute the project, including the choice of indication, the design goals, and the prompt. They did not contribute to the experimental measurements of the designed proteins or their analysis.

Use of AI tools

Claude designed the binders, helped us draft our prompt and four follow-ups, analyzed the data, drew the figures, and drafted this report under our supervision. Claude also wrote to Twist Bioscience about expression formats and sent the sequences, from one of our accounts, in emails marked as written by Claude. Before the campaign, we used Claude to generate the first lists of candidate indications, which informed our choice of envenomation as the indication.

References

[1] J.-P. Chippaux and M. Goyffon. Epidemiology of scorpionism: A global appraisal. *Acta Tropica*, 107(2):71–79, 2008. doi:10.1016/j.actatropica.2008.05.021.

[2] J.-P. Chippaux. Emerging options for the management of scorpion stings. *Drug Design, Development and Therapy*, 6:165–173, 2012. doi:10.2147/DDDT.S24754.

[3] T. P. Jenkins et al. Terrestrial venomous animals, the envenomings they cause, and treatment perspectives in the Middle East and North Africa. *PLOS Neglected Tropical Diseases*, 15(12):e0009880, 2021. doi:10.1371/journal.pntd.0009880.

[4] G. K. Isbister and H. S. Bawaskar. Scorpion envenomation. *The New England Journal of Medicine*, 371(5):457–463, 2014. doi:10.1056/NEJMra1401108.

[5] S. Ahmadi et al. Scorpion venom: Detriments and benefits. *Biomedicines*, 8(5):118, 2020. doi:10.3390/biomedicines8050118.

[6] W. A. Catterall et al. Voltage-gated ion channels and gating modifier toxins. *Toxicon*, 49(2):124–141, 2007. doi:10.1016/j.toxicon.2006.09.022.

[7] J. M. Gutiérrez et al. PLOS Neglected Tropical Diseases broadens its coverage of envenomings caused by animal bites and stings. *PLOS Neglected Tropical Diseases*, 15(6):e0009481, 2021. doi:10.1371/journal.pntd.0009481.

[8] A. H. Laustsen et al. Biotechnological trends in spider and scorpion antivenom development. *Toxins*, 8(8):226, 2016. doi:10.3390/toxins8080226.

[9] L. V. Boyer et al. Antivenom for critically ill children with neurotoxicity from scorpion stings. *The New England Journal of Medicine*, 360(20):2090–2098, 2009. doi:10.1056/NEJMoa0808455.

[10] World Health Organization. Rabies and envenomings: A neglected public health issue. Report of a consultative meeting, World Health Organization, Geneva, 10 January 2007. Technical report, World Health Organization, Geneva, 2007. ISBN 978-92-4-156348-2. https://www.who.int/publications/i/item/9789241563482.

[11] S. M. Alonso Villela et al. Production of recombinant scorpion antivenoms in *E. coli*: Current state and perspectives. *Applied Microbiology and Biotechnology*, 107(13):4133–4152, 2023. doi:10.1007/s00253-023-12578-1.

[12] F. Hui Wen et al. Snakebites and scorpion stings in the Brazilian Amazon: Identifying research priorities for a largely neglected problem. *PLOS Neglected Tropical Diseases*, 9(5):e0003701, 2015. doi:10.1371/journal.pntd.0003701.

[13] H. W. Fan et al. Situación de los laboratorios públicos productores de antivenenos en América Latina. *Revista Panamericana de Salud Pública*, 43:e92, 2019. doi:10.26633/RPSP.2019.92.

[14] J.-P. Chippaux et al. Factors involved in the resilience of incidence and decrease of mortality from scorpion stings in Mexico. *Toxicon*, 188:65–75, 2020. doi:10.1016/j.toxicon.2020.10.011.

[15] H. S. Bawaskar and P. H. Bawaskar. Efficacy and safety of scorpion antivenom plus prazosin compared with prazosin alone for venomous scorpion (*Mesobuthus tamulus*) sting: Randomised open label clinical trial. *BMJ*, 342:c7136, 2011. doi:10.1136/bmj.c7136.

[16] R. Kumar. An update on epidemiology and management practices of scorpion envenomation in India. *Journal of Family Medicine and Primary Care*, 11(9):4932–4935, 2022. doi:10.4103/jfmpc.jfmpc_2300_21.

[17] L. Riaño-Umbarila et al. State of the art on the development of a recombinant antivenom against Mexican scorpion stings. *Toxicon*, 257:108306, 2025. doi:10.1016/j.toxicon.2025.108306.

[18] L. Riaño-Umbarila et al. Optimal neutralization of *Centruroides noxius* venom is understood through a structural complex between two antibody fragments and the Cn2 toxin. *The Journal of Biological Chemistry*, 291(4):1619–1630, 2016. doi:10.1074/jbc.M115.685297.

[19] R. Ben Abderrazek et al. Identification of potent nanobodies to neutralize the most poisonous polypeptide from scorpion venom. *Biochemical Journal*, 424(2):263–272, 2009. doi:10.1042/BJ20090697.

[20] I. Hmila et al. A bispecific nanobody to provide full protection against lethal scorpion envenoming. *The FASEB Journal*, 24(9):3479–3489, 2010. doi:10.1096/fj.09-148213.

[21] L. Riaño-Umbarila et al. Generation of a broadly cross-neutralizing antibody fragment against several Mexican scorpion venoms. *Toxins*, 11(1):32, 2019. doi:10.3390/toxins11010032.

[22] S. Vázquez Torres et al. *De novo* designed proteins neutralize lethal snake venom toxins. *Nature*, 639(8053):225–231, 2025. doi:10.1038/s41586-024-08393-x.

[23] S. Ahmadi et al. Nanobody-based recombinant antivenom for cobra, mamba and rinkhals bites. *Nature*, 647(8090):716–725, 2025. doi:10.1038/s41586-025-09661-0.

[24] M. D. Overath et al. *De novo* designed single-domain antibodies protect against lethal cobra venom neurotoxicity *in vivo*. bioRxiv preprint, 2026. doi:10.64898/2026.09.01.748349.

[25] J. L. Watson et al. *De novo* design of protein structure and function with RFdiffusion. *Nature*, 620(7976):1089–1100, 2023. doi:10.1038/s41586-023-06415-8.

[26] J. Dauparas et al. Robust deep learning-based protein sequence design using ProteinMPNN. *Science*, 378(6615):49–56, 2022. doi:10.1126/science.add2187.

[27] M. Pacesa et al. One-shot design of functional protein binders with BindCraft. *Nature*, 646(8084):483–492, 2025. doi:10.1038/s41586-025-09429-6.

[28] H. Stark et al. BoltzGen: Toward universal binder design. bioRxiv preprint, 2025. doi:10.1101/2025.11.20.689494.

[29] J. Butcher et al. *De novo* design of all-atom biomolecular interactions with RFdiffusion3. bioRxiv preprint, 2025. doi:10.1101/2025.09.18.676967.

[30] L. Cao et al. Design of protein-binding proteins from the target structure alone. *Nature*, 605(7910):551–560, 2022. doi:10.1038/s41586-022-04654-9.

[31] K. Swanson et al. The Virtual Lab of AI agents designs new SARS-CoV-2 nanobodies. *Nature*, 646(8085):716–723, 2025. doi:10.1038/s41586-025-09442-9.

[32] M. Jeon et al. Agentic campaign control for high-throughput *de novo* binder design. bioRxiv preprint, 2026. doi:10.64898/2026.09.22.753604.

[33] Latent Labs Team. Latent-Y: A lab-validated autonomous agent for *de novo* drug design. arXiv:2603.29727, 2026. doi:10.48550/arXiv.2603.29727.

[34] Muni. Finding three sub-nanomolar binders, agentically. https://muni.bio/research/closing-the-autoresearch-loop, 2026.

[35] J. Kim and P. A. Romero. Benchmarking and behavioral characterization of LLM agents for protein design. bioRxiv preprint, 2026. doi:10.64898/2026.05.06.723381.

[36] X. Xu et al. ProteinMCP: An agentic AI framework for autonomous protein engineering. *Protein Science*, 35(4):e70547, 2026. doi:10.1002/pro.70547.

[37] Claude Science and A. Shanehsazzadeh. Autonomous *de novo* protein binder design with Claude. Technical report, Anthropic, August 2026. https://www-cdn.anthropic.com/30bf50e22a01388bb29bf077ee3f244531594b7a.pdf. Data and prompts: https://huggingface.co/datasets/Anthropic/claude-protein-binder-design.

[38] Citrini Research. Citrini Therapeutics. https://www.citriniresearch.com/p/citrini-therapeutics, 2026.

[39] E. A. Hernández Muñoz et al. Scorpion sting envenomation: a neglected tropical disease in the shadow of global health priorities: an urgent call to action. *BMJ Global Health*, 10(11):e020682, 2025. doi:10.1136/bmjgh-2025-020682.

[40] M. B. Pucca et al. Serrumab: A novel human single chain-fragment antibody with multiple scorpion toxin-neutralizing capacities. *Journal of Immunotoxicology*, 11(2):133–140, 2014. doi:10.3109/1547691X.2013.809175.

[41] C. Devaux et al. Construction and functional evaluation of a single-chain antibody fragment that neutralizes toxin AahI from the venom of the scorpion *Androctonus australis hector*. *European Journal of Biochemistry*, 268(3):694–702, 2001. doi:10.1046/j.1432-1327.2001.01923.x.

[42] L. Riaño-Umbarila et al. Exploiting cross-reactivity to neutralize two different scorpion venoms with one single chain antibody fragment. *The Journal of Biological Chemistry*, 286(8):6143–6151, 2011. doi:10.1074/jbc.M110.189175.

[43] L. Riaño-Umbarila et al. Full neutralization of *Centruroides sculpturatus* scorpion venom by combining two human antibody fragments. *Toxins*, 13(10):708, 2021. doi:10.3390/toxins13100708.

[44] X. Fan et al. Diverse binding poses of agonistic neurotoxins on human $Na_V1.6$. *Nature*, 656(8126):250–259, 2026. doi:10.1038/s41586-026-10661-x.

[45] C. B. Pinheiro et al. Structural analysis of *Tityus serrulatus* Ts1 neurotoxin at atomic resolution: Insights into interactions with Na^+ channels. *Acta Crystallographica Section D*, 59(3):405–415, 2003. doi:10.1107/S090744490202111X.

[46] G. D. Smith et al. *Ab initio* structure determination and refinement of a scorpion protein toxin. *Acta Crystallographica Section D*, 53(5):551–557, 1997. doi:10.1107/S0907444997005386.

[47] D. Jiang et al. Structural basis for voltage-sensor trapping of the cardiac sodium channel by a deathstalker scorpion toxin. *Nature Communications*, 12(1):128, 2021. doi:10.1038/s41467-020-20078-3.

[48] T. Clairfeuille et al. Structural basis of α-scorpion toxin action on Na_V channels. *Science*, 363(6433):eaav8573, 2019. doi:10.1126/science.aav8573.

[49] D. Jiang and W. A. Catterall. Cardiac sodium channel in complex with LqhIII. Protein Data Bank, entry 7XSU, 2023. doi:10.2210/pdb7XSU/pdb.

[50] cytokineking. FreeBindCraft: BindCraft with optional PyRosetta bypass. Code repository, https://github.com/cytokineking/FreeBindCraft, 2026.

[51] M. Ren et al. PXDesign: Fast, modular, and accurate *de novo* design of protein binders. bioRxiv preprint, 2025. doi:10.1101/2025.08.15.670450.

[52] K. Didi et al. Scaling atomistic protein binder design with generative pretraining and test-time compute. International Conference on Learning Representations (ICLR), oral, 2026. doi:10.48550/arXiv.2603.27950.

[53] Y. Lin et al. Fast and ultra-capable protein design: Advancing the frontier through atomistic SE(3)-equivariance with Genie 3. bioRxiv preprint, 2026. doi:10.64898/2026.05.01.722168.

[54] N. Boyd, S. Guns, and Escalante Bio. Mosaic. Code repository, https://github.com/escalante-bio/mosaic, 2025.

[55] S. Candido et al. Language modeling materializes a world model of protein biology. bioRxiv preprint, 2026. doi:10.64898/2026.06.03.729735.

[56] K. R. Rustamov and A. Y. Baev. Fold-conditioned *de novo* binder design via AlphaFold2-Multimer hallucination. bioRxiv preprint, 2025. doi:10.1101/2025.07.02.662497.

[57] Y. Cho et al. BoltzDesign1: Inverting all-atom structure prediction model for generalized biomolecular binder design. bioRxiv preprint, 2025. doi:10.1101/2025.04.06.647261.

[58] Y. Cho et al. Protein Hunter: Exploiting structure hallucination within diffusion for protein design. bioRxiv preprint, 2025. doi:10.1101/2025.10.10.681530.

[59] C. A. Goverde et al. Computational design of soluble and functional membrane protein analogues. *Nature*, 631(8020):449–458, 2024. doi:10.1038/s41586-024-07601-y.

[60] R. W. Shuai et al. Ensemble-conditioned protein sequence design with Caliby. bioRxiv preprint, 2025. doi:10.1101/2025.09.30.679633.

[61] Biohub. ESMFold2 and ESMFold2-Fast: Model weights and the esm inference package. Model repositories, https://huggingface.co/biohub/ESMFold2 and https://huggingface.co/biohub/ESMFold2-Fast; code, https://github.com/Biohub/esm, 2026.

[62] Y. Zhang et al. Protenix-v2: Broadening the reach of structure prediction and biomolecular design. bioRxiv preprint, 2026. doi:10.64898/2026.04.10.717613.

[63] R. L. Dunbrack. Rēs ipSAE loquuntur: What's wrong with AlphaFold's ipTM score and how to fix it. bioRxiv preprint, 2025. doi:10.1101/2025.02.10.637595.

[64] S. Basu and B. Wallner. DockQ: A quality measure for protein-protein docking models. *PLOS ONE*, 11(8):e0161879, 2016. doi:10.1371/journal.pone.0161879.

[65] C. Mirabello and B. Wallner. DockQ v2: Improved automatic quality measure for protein multimers, nucleic acids, and small molecules. *Bioinformatics*, 40(10):btae586, 2024. doi:10.1093/bioinformatics/btae586.

[66] J. C. Canul-Tec et al. Structural basis of neutralization of the major toxic component from the scorpion *Centruroides noxius* Hoffmann by a human-derived single-chain antibody fragment. *The Journal of Biological Chemistry*, 286(23):20892–20900, 2011. doi:10.1074/jbc.M111.238410.

[67] I. P. Fabrichny et al. Structural insights into antibody sequestering and neutralizing of Na^+ channel α-type modulator from Old World scorpion venom. *The Journal of Biological Chemistry*, 287(17):14136–14148, 2012. doi:10.1074/jbc.M111.315382.

[68] I. Hötzel et al. A strategy for risk mitigation of antibodies with fast clearance. *mAbs*, 4(6):753–760, 2012. doi:10.4161/mabs.22189.

[69] M. Mousli et al. A recombinant single-chain antibody fragment that neutralizes toxin II from the venom of the scorpion *Androctonus australis hector*. *FEBS Letters*, 442(2–3):183–188, 1999. doi:10.1016/S0014-5793(98)01647-0.

[70] M. B. Pucca et al. Serrumab: A human monoclonal antibody that counters the biochemical and immunological effects of *Tityus serrulatus* venom. *Journal of Immunotoxicology*, 9(2):173–183, 2012. doi:10.3109/1547691X.2011.649220.

[71] M. M. Pereira et al. Single-step purification of ovalbumin from egg white using aqueous biphasic systems. *Process Biochemistry*, 51(6):781–791, 2016. doi:10.1016/j.procbio.2016.03.002.

[72] Z. O. Shenkarev et al. Recombinant production and structure-function study of the Ts1 toxin from the Brazilian scorpion *Tityus serrulatus*. *Doklady Biochemistry and Biophysics*, 484:9–12, 2019. doi:10.1134/S1607672919010034.

[73] M. D. Overath et al. Predicting experimental success in *de novo* binder design: A meta-analysis of 3,766 experimentally characterised binders. bioRxiv preprint, 2025. doi:10.1101/2025.08.14.670059.

[74] M. Benard-Valle et al. *In vivo* neutralization of coral snake venoms with an oligoclonal nanobody mixture in a murine challenge model. *Nature Communications*, 15(1):4310, 2024. doi:10.1038/s41467-024-48539-z.

[75] A. Damsbo et al. Discovery of broadly neutralizing V_HHs against short-chain α-neurotoxins using a consensus toxin as an antigen. *mAbs*, 17(1):2522838, 2025. doi:10.1080/19420862.2025.2522838.

[76] S. Cardoso-Arenas et al. Immunogenicity of two representative American consensus scorpion neurotoxins from the genera *Tityus* and *Centruroides*. *PLOS Neglected Tropical Diseases*, 20(2):e0013982, 2026. doi:10.1371/journal.pntd.0013982.

[77] V. Brasas et al. Specificity-driven protein binder design with Odin-Multi. bioRxiv preprint, 2026. doi:10.64898/2026.09.08.749745.

[78] PacesaLab. BindCraft2: Easy to use suite for designing a wide range of protein binder modalities. Code repository, https://github.com/PacesaLab/BindCraft2; laboratory page, https://www.pacesalab.com/bindcraft, 2026.

[79] M.-F. Martin-Eauclaire et al. Serotherapy against voltage-gated sodium channel-targeting α-toxins from *Androctonus* scorpion venom. *Toxins*, 11(2):63, 2019. doi:10.3390/toxins11020063.

[80] Claude Science et al. Accelerating open-source biomolecular models with Claude. Technical report, Anthropic, September 2026. https://www-cdn.anthropic.com/e96b5807039a88168733d9687afe41dfbbd5de13.pdf. Code: https://github.com/anthropics/uplifting-biomolecular-modeling.

[81] B. E. Suzek et al. UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*, 31(6):926–932, 2015. doi:10.1093/bioinformatics/btu739.

[82] M. Steinegger and J. Söding. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11):1026–1028, 2017. doi:10.1038/nbt.3988.

[83] Y. Zhang and J. Skolnick. Scoring function for automated assessment of protein structure template quality. *Proteins: Structure, Function, and Bioinformatics*, 57:702–710, 2004. doi:10.1002/prot.20264.

[84] C. Zhang et al. US-align: Universal structure alignments of proteins, nucleic acids, and macromolecular complexes. *Nature Methods*, 19(9):1109–1115, 2022. doi:10.1038/s41592-022-01585-1.

[85] J. Jumper et al. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021. doi:10.1038/s41586-021-03819-2.

[86] Y. Zhang and J. Skolnick. TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research*, 33(7):2302–2309, 2005. doi:10.1093/nar/gki524.

[87] A. Pintar et al. Solution structure of toxin 2 from *Centruroides noxius* Hoffmann, a β-scorpion neurotoxin acting on sodium channels. *Journal of Molecular Biology*, 287(2):359–367, 1999. doi:10.1006/jmbi.1999.2611.

[88] I. Polikarpov et al. Crystal structure of neurotoxin Ts1 from *Tityus serrulatus* provides insights into the specificity and toxicity of scorpion toxins. *Journal of Molecular Biology*, 290(1):175–184, 1999. doi:10.1006/jmbi.1999.2868.

[89] D. Housset et al. Crystal structure of toxin II from the scorpion *Androctonus australis* Hector refined at 1.3 Å resolution. *Journal of Molecular Biology*, 238(1):88–103, 1994. doi:10.1006/jmbi.1994.1270.

[90] I. Krimm et al. NMR structures and activity of a novel α-like toxin from the scorpion *Leiurus quinquestriatus hebraeus*. *Journal of Molecular Biology*, 285(4):1749–1763, 1999. doi:10.1006/jmbi.1998.2418.

[91] J. A. Huntington and P. E. Stein. Structure and properties of ovalbumin. *Journal of Chromatography B: Biomedical Sciences and Applications*, 756(1–2):189–198, 2001. doi:10.1016/S0378-4347(01)00108-6.

[92] P. Rice et al. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics*, 16(6):276–277, 2000. doi:10.1016/S0168-9525(00)02024-2.

[93] I. Amaro et al. Isolation and characterization of a human antibody fragment specific for Ts1 toxin from *Tityus serrulatus* scorpion. *Immunology Letters*, 139(1–2):73–79, 2011. doi:10.1016/j.imlet.2011.05.002.

Supplementary Information

De novo design of protein binders against scorpion toxins with Claude

S1 Extended methods **20**

S1.1 Design campaign . . . 20

S1.2 Experimental characterization . . . 20

S2 Campaign details **23**

S2.1 We set the rules and Claude executed the campaign . . . 23

S2.2 Nearly every selected pair passed the thresholds, and no design resembles a toxin . . . 25

S3 Experimental details and full results **27**

S3.1 Production and coverage of the designs . . . 27

S3.2 Adaptyv Bio's fits and replicates . . . 27

S3.3 The toxins and the theoretical $R_{\max}$ . . . 28

S3.4 Positive controls . . . 29

S3.5 Hits across all 1,055 design-toxin pairs . . . 30

S3.6 Agreement between the two labs . . . 33

S3.7 SPR against toxins from a second supplier . . . 34

S3.8 The number of binders under other definitions . . . 36

S3.9 The 17 binders in detail . . . 36

S3.10 Reverse BLI . . . 42

S3.11 Predicted interfaces and binding . . . 43

S4 Epitope and specificity assays **51**

S4.1 Competition with reference antibodies . . . 51

S4.2 BVP ELISA . . . 56

S4.3 Unintended binding . . . 58

S4.4 SPR against ovalbumin . . . 60

S5 *In silico* scores and measured binding **65**

S5.1 Designs predicted as monomers . . . 65

S5.2 Co-folded positive controls . . . 67

S6 Published antibodies and nanobodies against the five toxins **67**

S7 Limitations of this study and what would resolve them **69**

Guide. Table S6 lists the ten design sets, and Set 10 holds the 31 designs added later. The main text groups the sets into single-toxin designs (Sets 1 to 5), double β-toxin designs (Set 6), and triple α-toxin designs (Set 7). Designs against all five toxins are single domains (Set 8 and 29 of Set 10) or fusions (Set 9 and 2 of Set 10). The main screen is each lab's surface plasmon resonance (SPR) measurement of all designs, and the hits come from it. Binders have names such as Cn2-B1 (Table 1), and all other designs go by their eight-character UUIDs from the data release (`https://huggingface.co/datasets/Anthropic/scorpion-toxin-binder-design`; column `uuid`). Table S17 and `design.csv` in the release map each name to its UUID. In the release, only the column `binder_against` marks a design as a binder and names its toxin. For criterion 4, a signal against CLEC12A, an unrelated protein, is a hit or a weak or near-background signal in either SPR run at Twist Bioscience (Section S1.2). Ovalbumin (OVA in the data release), tested at Adaptyv Bio, is not among the criteria, and hits against it are flagged (Section S4.4). Unless stated otherwise, intervals are 95% intervals, from a bootstrap over designs or, for proportions of designs or pairs, by the Wilson method.

S1 Extended methods

S1.1 Design campaign

Target sites. For Cn2, the channel complex that defined the site includes the channel's $\beta 1$ subunit. Claude added residues from UniProt annotations to the sites of AaH II (K58, R62, and H64) and Ts1 (K12, W39, and W54), and chose five hotspots per toxin (Table S5). Some designs of Sets 6 to 8, which target multiple toxins, were generated with other hotspot lists. The sequence identities and Cα root-mean-square deviations (RMSDs) in Figure 1C come from pairwise global alignments (BLOSUM62; a gap of n residues costs $11 + (n - 1)$; end gaps free; among alignments of equal score, the one with the lowest RMSD).

Scoring. Each predictor received a toxin as its full mature chain, with the number of seeds as in Table S1. The design model is the complex of a design and a toxin that Claude kept as the reference pose for scDockQ. Every selected design came from one or more rounds of redesign on a design tool's complex, a backbone remade by partial diffusion, or an earlier design's predicted complex, so the design model can be such a prediction, and scDockQ then measures agreement with it. If the design model lacked an intended toxin, that toxin was superposed onto the toxin in the model before scDockQ was computed. In Set 8, nearly all designs on one backbone share one design model, so their scDockQ measures agreement with that shared pose.

Ranking and diversity. The ranking used three keys in order: 1) a complete set of seeds, 2) an scDockQ pass, and 3) the weighted z-score of the six scores. The diversity rules were: no repeated sequence; no two designs within five substitutions of each other; at most 2 designs from one backbone and 3 from one structural cluster in a set of 12, or 3 and 4 in a set of 30; at most half of a set from one design tool; and at least three tools per set. A rule could be relaxed stepwise if a set could not be filled, as in Set 7, where four backbones supplied four designs each. For designs made against multiple toxins, the z-score of the third key was the soft minimum $-\tau \ln \Sigma_t \exp(-z_t/\tau)$ over the intended toxins t, with $\tau = 0.1$. We removed the diversity caps for Set 8 in our fourth follow-up (Section S2.1). The spacing of at least 10 substitutions between designs was Claude's own rule, and the final 30 designs keep it.

Sequence design and fusions. Multistate SolubleMPNN, which designs one sequence for several structures at once, generated 16 of the 30 designs in Set 7 and 28 of the 30 in Set 8, but none in Set 6 (Table S1). The fusions of Set 9 joined domains that had not been shown to bind on their own. In 29 of the 30, the first domain was one of six designs of Set 6, and none of the six was labeled a binder as a single domain. No second domain came from Set 7, and none was tested as a single domain.

Designs added later. Of the 31 designs of Set 10, 20 had passed the score thresholds against all five toxins in the campaign's rounds of sequence redesign but were not selected. They are the only unselected designs we tested, so no unselected single-toxin design was tested. Nine are variants of 005ebdae and cc1c8edb, chosen by co-folding and molecular dynamics to improve binding against LqhIII, and two are fusions of those two designs, one in each domain order. None of the 31, and none of the 30 designs of Set 8, was a hit at either lab in the main screen.

S1.2 Experimental characterization

Designs. The 211 designs consisted of 179 single domains of 55 to 120 amino acid residues and 32 fusions of 174 to 234 amino acid residues. Table S2 shows how Adaptyv Bio and Twist Bioscience produced them, and Table S3 how they measured them.

SPR at Twist Bioscience. Claude identified the hits from Twist Bioscience's data under our direction, after Adaptyv Bio's hits were known. A design was assessable if at least 50 response units (RU) of it were captured. The rule for hits in the Methods requires a response 5 robust standard deviations (SD) above the median, and one robust SD is 1.48 times the median absolute deviation of that toxin's responses in the run. Responses 3 to 5 SD above the median were labeled weak or ambiguous. Against CLEC12A in the second run (the run against AaH I), responses about 2 to 3 SD above the median were labeled near-background. Neither label is a hit. Each response is the mean over about the last 12 s of an injection, minus the mean curve of that injection's negative-control spots (trastuzumab; buffer; and 9C2 in the run against Cn2, Ts1, and AaH II). In the run against AaH I, each spot's adjacent reference region had already been subtracted. The baseline was set once per series, without regeneration, so the response at 1 µM is cumulative. A baseline reset before each injection would count only the signal added at 1 µM, so Cn2-B1, which saturates at lower concentrations, would no longer be a hit.

SPR at Adaptyv Bio. SPR was run on a Carterra LSA-XT with CMDP chips. Strep-Tactin XT was amine-coupled to each chip at 50 µg/mL in 10 mM sodium acetate, pH 4.5. Designs were captured through their twin-Strep tag for 750 s, followed by 600 s of buffer. In the main screen, the median capture of design spots ranged from 271 to 868 RU per run and was about 500 to 700 RU for each toxin, and Table S7 gives the binders' capture levels. The running buffer was HBSTE (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, and 0.05% Tween-20, pH 7.4), the flow rate was 50 µL/min, and the chip was held at 25 °C. Each series ran from the lowest to the highest concentration without regeneration. A replicate is one spot measured through one series, and the replicates of a pair did not always lie in separate runs: both replicates lay in one run for AaHII-B1 and AaHII-B2, two of the four for AaHII-B4, and two of the three for each LqhIII binder. Signals from reference spots with the same surface chemistry but no design, and from the leading buffer injection, were subtracted (double referencing), and a linear drift correction was made on a buffer segment.

Table S1. Settings of the design campaign, from filters to sequence design. A filter removed every design that met its criterion. Seeds are counted per structure predictor, and a window is one of Claude's three working periods. Sequence labels count the 180 campaign designs, and the fusion settings apply to the 30 fusions of Set 9.

Step	Criterion or term	Setting	Tool
Filter	Resemblance to a known protein	More than 60% identity over 50% coverage to UniRef90 [81] or to known binders	MMseqs2 [82]
Filter	Resemblance to a reference chain	At least 30% local identity over at least 40 aligned residues	MMseqs2
Filter	Counting of matches	A UniRef90 match counted only at an E-value of 0.001 or less, and a match to a chain longer than 300 residues only if it scored better than shuffled sequences	MMseqs2
Filter	Resemblance to a toxin	TM-score [83] of 0.5 or more (run on 64 of the 180 selected designs; Section S2.2)	US-align [84]
Filter	Sequence liabilities	Any cysteine, a run of five identical residues, or a hydrophobic or low-complexity stretch	
Filter	Monomer confidence	Mean pLDDT [85] (the predictor's per-residue confidence, 0 to 100) below 70, design predicted as a monomer with one seed	ESMFold2-Fast
Filter	Interface contacts	Fewer than 8 contacts (backbone and Cβ atoms within 5 Å) in the design model	
Scoring	Seeds	One to screen, five to rank, and ten to rank in the third window	Three structure predictors
Scoring	ipSAE term	Lower of the two directional values from the best seed, with a threshold per toxin and predictor (Table S26): at least the highest of the 95th percentile of 28 to 31 decoys and negative controls; 0.01 above the highest scrambled, shuffled, or arbitrary-helix decoy; and 0.15. Thresholds were fixed before any design was scored	ipSAE
Scoring	scDockQ term	Best seed's complex against the design model, passing at 0.23 or more	DockQ
Scoring	Missing score	By the campaign's rules, if the predicted toxin lost its native fold or any of its four disulfides	
Scoring	Final score	Mean of the six terms, ipSAE and scDockQ from each predictor, each 0 to 1; reported but not used to rank	
Ranking	Seed split	From the third window on, a pass on seeds 0 to 4 and again on seeds 5 to 9	
Ranking	z-score weights	ipSAE terms 4 and scDockQ terms 1, comparable only within a pool	
Ranking	Multiple toxins	Soft minimum over the intended toxins, $\tau = 0.1$	
Diversity	Structural cluster	Designed chains at a TM-score of 0.90 or more	
Diversity	Set 8	At least 10 substitutions between any two designs, in place of the caps	
Sequence	Label and designs	SolubleMPNN, 80; multistate SolubleMPNN, 44; SolubleCaliby, 24; fusion of two domains, 30; interface graft, 2 (5b454168 and cc1c8edb)	
Fusion	Linker and length	AEAAAKEAAAKEAAAKA, giving 174 to 234 residues	

BLI at Adaptyv Bio. Biolayer interferometry (BLI) was run on a Gator Bio Pivot at room temperature in HBSTE2 (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, and 0.2% Tween-20). In the main screen, designs were loaded on Strep-Tactin XT probes, and each cycle consisted of 60 s baseline, 120 s loading, 180 s baseline, 300 s association, and 300 s dissociation. Reference probes were loaded with a cell-free reaction that contained no design. Reverse BLI (Section S3.10) used the same instrument, buffer, and association and dissociation times, but the His-tagged toxin was captured on HIS XT probes (20 µg/mL for 120 s) and the purified design was in solution. Here the reference probes were loaded with buffer. For competition (Section S4.1), the antibody, a single-

Table S2. Production of the designs at the two labs. Table S7 gives the results.

Lab	Construct and expression	Purification and quality control
Adaptyv Bio	Design with C-terminal GFP11 and twin-Strep tags, by cell-free expression in an *E. coli* system; for competition, 8×His replaces twin-Strep	Soluble yield by a split-GFP assay (range 2.5 to 25 µM); clarified reaction, without a purity check of individual designs, in the main screen; purified by the tag for reverse BLI, SPR against ovalbumin, and competition
Twist Bioscience	Design, GGGGS linker, and human IgG1 hinge and Fc, expressed in HEK cells in 1 mL cultures	One-step Protein A purification; concentration by absorbance at 280 nm; purity by capillary electrophoresis with SDS (CE-SDS); main peak by analytical size-exclusion chromatography (SEC)

Table S3. Assays at the two labs. For binding, conditions are those of the measurements used to identify hits. The text gives Adaptyv Bio's surface chemistry, buffers, and referencing. Their SPR against the second supplier's toxins and against ovalbumin is described in Sections S3.7 and S4.4.

Assay, lab	Format	Method	Concentrations and timing	Replicates and hits
Binding, Adaptyv Bio	Design captured by its twin-Strep tag on Strep-Tactin XT; toxin in solution	High-throughput SPR (HT-SPR; Carterra LSA-XT) on CMDP chips, 25 °C, double referenced; median capture per run 271 to 868 RU; 39 of the 1,713 series, against Cn2, Ts1, AaH II, and AaH I, by biolayer interferometry (BLI; Gator Bio Pivot), room temperature. 1:1 fits with a drift slope and bulk offsets (Section S3.2)	SPR: eight concentrations, 0.3 to 1,000 nM in half-log steps (seven, from 1 nM, or five, from 10 nM, in some AaH I and LqhIII replicates); association about 300 s, dissociation about 590 s. BLI: 31.6 to 1,000 nM; 300 s each. Low to high, no regeneration	Two to four per pair; K_D, the arithmetic mean of the fitted replicates (standard deviations, Table S18). Hits identified by Adaptyv Bio, by the rule in the text
Binding, Twist Bioscience	Fc fusion, a homodimer carrying two copies of the design, captured on an anti-human Fc chip, one spot per design; toxin in solution	HT-SPR (Carterra LSA-XT) on HC30M chips, 25 °C; median capture 137 to 209 RU; two runs (Cn2, Ts1, and AaH II; AaH I), CLEC12A in both. 1:1 fits by us (Section S1.2)	Six concentrations, 4.1 to 1,000 nM in threefold steps; association 5 min, dissociation about 8 min. Low to high, no regeneration	One per pair. Hits identified by Claude, by the rule in the text
Competition, Twist Bioscience	Toxin on the biosensor; antibody at 150 nM or buffer; then design	BLI in tandem, 25 °C; polyclonal human IgG as negative control (none in the run on AaH I)	Toxin loading 90 s, antibody or buffer 300 s, design about 90 s	One per design and antibody. Assessed by us (Section S4.1)
Competition, Adaptyv Bio	scFv of LR or 9C2 on a Strep-Tactin XT probe; toxin premixed with His-tagged design in solution	BLI (Gator Bio Pivot), room temperature; toxin alone as the reference	200 nM toxin and 1 µM design, premixed overnight at 4 °C; association 300 or 600 s, dissociation 300 s	Two or three per design. Calls by Adaptyv Bio, ratios by us (Section S4.1)
BVP ELISA, Twist Bioscience	Fc fusions on baculovirus particles (BVP)	ELISA, absorbance at 450 nm; two runs (33 and 18 designs); reference panel on another plate on the same day	One concentration per design, 0.02 to 0.13 mg/mL; controls 0.1 mg/mL	One well on each of two plates; each design in one run
Reverse BLI, Adaptyv Bio	His-tagged toxin captured on a HIS XT probe; purified design in solution	BLI (Gator Bio Pivot), room temperature; reference probes loaded with buffer. Fits by Adaptyv Bio	31.6 to 1,000 nM in half-log steps; association and dissociation 300 s each	One to three per design and toxin. Hits identified by Adaptyv Bio (Section S3.10)
Reverse BLI, Twist Bioscience	His-tagged toxin captured on the biosensor; Fc fusion in solution	BLI (Octet HTX), 25 °C; sensors regenerated between cycles; reference sensors loaded with buffer	0.55 to 400 nM in threefold steps; association 90 s, dissociation 600 s	One per design and toxin. Hits identified by Twist Bioscience, at 0.05 nm or more (Section S3.10)

chain variable fragment (scFv) with a twin-Strep tag, was loaded on Strep-Tactin XT probes from a diluted cell-free reaction. The design kept its GFP11 tag and carried an 8×His tag in place of the twin-Strep tag, which would itself bind the probe.

Toxins at Adaptyv Bio. Cn2, Ts1, and AaH II were bought from Biomatik, and a second preparation of each from Cusabio, Twist Bioscience's supplier (Table S11). AaH I and LqhIII were made in-house in an *E. coli* cell-free system (12 h, room temperature) with DsbC, oxidized glutathione, and the chaperones DnaK, DnaJ, and GrpE. Each construct comprised an N-terminal methionine, the toxin, and C-terminal GFP11 and 8×His tags, with masses of 11.2 kDa for AaH I and 11.3 kDa for LqhIII, and the AaH I construct held 64 toxin residues, one more at the C-terminus than the 63-residue isoform used for design. They were purified on a cobalt resin and exchanged into 20 mM HEPES and 150 mM NaCl, pH 7.4.

Hits and fits at Adaptyv Bio. Adaptyv Bio identified hits by their own analysis. A design was a hit against a toxin if it gave a quantifiable binding curve with a fitted K_D or, where no fit was reliable, if its association signal was at least 300% above the negative control's. In their data package, a pair is a hit if any of its replicates is. Fits were made with Adaptyv Bio's own fitting pipeline, using a global 1:1 model in which k_{off} and K_D were fitted and k_{on} was derived from them. They graded each fit Good, Medium, or Poor. Other models were chosen where fit quality called for them, but for no binder in the main screen (Section S3.2). We used the hits and K_D values in their data package, apart from the upper limit for Cn2-B1 (below).

K_D for a pair. Our 1:1 fit to a pair's Twist Bioscience series starts each injection from its measured initial response. Twist Bioscience also fitted 45 of the 48 hits, and our fits agreed with theirs within 2.3-fold (median ratio 0.98). We also fitted Adaptyv Bio's traces with a 1:1 model with a drift slope, and for the 105 main-screen replicates where their 1:1 K_D was 1 µM or tighter (85 by SPR and 20 by BLI), our K_D agreed with theirs within 2-fold for 104 and within 2.5-fold for all (median ratio 1.00). The symbol $\leq$ marks an upper limit where the off-rate was not resolved, as for Cn2-B1, whose limits at both labs come from a method that needs no rate constant (Section S3.9). The symbol $\approx$ marks Adaptyv Bio's values for the four AaH I binders, each the mean of two fitted replicates that differ 3.0- to 13-fold, and their values for other designs under the same rule (fitted replicates more than 3-fold apart, a fit below the lowest concentration of its series, or another model selected); Twist Bioscience's values for Cn2-B7, 9a8cdc7a, and e32555c5 against Cn2, all near the highest concentration tested; reverse BLI values that lie below the lowest concentration tested or come from replicates for which Adaptyv Bio selected another model; and the reverse BLI values of 9a8cdc7a against Cn2, which Adaptyv Bio's team does not consider reliable. Other values near or above the highest concentration tested are no more certain, although they carry no sign.

The theoretical R_{max}. The theoretical R_{max} depends on the amount of design on the surface and on the mass of the toxin construct. Criterion 3 was applied to Twist Bioscience's data alone and not at Adaptyv Bio, where antigen-binding fragments (Fabs) raised against other toxins gave signals (Section S3.4). With Adaptyv Bio's capture levels (Table S7) and the constructs of their toxins (Table S11), Section S3.3 reports a check of the criterion on their data.

$$R_{\text{max, theoretical}} = R_{\text{capture}} \times n \times \frac{M_{\text{toxin}}}{M_{\text{captured}}}$$

Here $R_{capture}$ is the capture level; $M_{captured}$ is the mass of the captured fusion; M_{toxin} is the mass of the toxin with its tags, derived from Twist Bioscience's theoretical R_{max} values; and n is the number of design copies per Fc dimer, two (Table S10). Assuming every captured design is active, a response above the theoretical R_{max} means that something besides one-to-one binding adds to the signal or that the toxin is heavier than assumed.

Criteria for labeling a design a binder. Criterion 1 (Figure 3D) requires a hit in every replicate. Of the 35 pairs that were hits at both labs, 34 met it; 9a8cdc7a was a hit against Cn2 in one of its three replicates at Adaptyv Bio. Criterion 4 treated a hit or a weak or near-background signal against CLEC12A in either run as off-target binding. None of the 17 binders exceeded the median against CLEC12A by more than 1.4 SD in either run, so the choice of labels does not affect them. The same 15 of the 35 pairs meet all four criteria at any limit on the fitted R_{max} from 0.82 to 4.1 times the theoretical R_{max}. At 4.2 times, AaHII-B4 against Ts1 is added.

Controls. A positive control (Table S12) worked at Twist Bioscience if its fit passed their quality check, and at Adaptyv Bio if its response exceeded the designs' median by 5 robust SD or they selected a fit (Table S11). We say that a control bound or gave a signal against a toxin if Adaptyv Bio identified it as a hit, as for the designs. Serrumab's result against Ts1 at Adaptyv Bio is inconclusive (Section S3.4).

S2 Campaign details

S2.1 We set the rules and Claude executed the campaign

Table S4 lists what our prompt specified and what Claude decided. We wrote the prompt jointly with Citrini Research, adapting the binder design prompt of our first technical report [37]. The safety rules required Claude, among other things, to stop at once if Anthropic staff asked and to pause and report any genuine safety concern.

The prompt, about 19,000 words long, opened with this statement of the task, and the rest of it expanded each clause into the rules listed in Table S4:

> Task: Autonomously *de novo* design miniprotein binders against five scorpion-venom neurotoxins that act on voltage-gated sodium (Na_V) channels, drawn from the two medically most important structural families (β-toxins Cn2, UniProt P01495, and Ts1, P15226; α-toxins AaH II, P01484, and AaH I, P01479, and the α-like toxin LqhIII, P56678), using a structure-based design-and-filter approach, targeting each toxin's Na_V-receptor-binding surface so that a bound binder is expected to block channel engagement, with four goals: 1) high-affinity binders zero-shot, 2) high overall hit rate, 3) cross-reactive binders that cover a toxin family with one molecule, and 4) multivalent molecules that cover all five toxins.

Table S4. What the prompt specified, and what Claude decided. The columns are separate lists, aligned by topic. Four decisions on the right followed instructions in our follow-ups: the extra predictor, the fusions' design, the changes to the process in the second window, and the third window's schedule and budget split.

Specified by the prompt	Decided by Claude
The targets, eight design sets, and how the 180 designs are split among them (in our first follow-up we halved Set 8 and added Set 9); for AaH I, at least two predicted models	Which toxin structures to use (Table S9)
Epitope: each toxin's channel-binding site	The exact hotspot residues within each site (Table S5)
Rules for each molecule: one domain of 50 to 160 residues, no cysteine, no known binder as a starting point or graft (our first follow-up allowed fusions of two domains)	The domain order and linker of the fusions, in the third window
Three predictors for ranking, their settings, one seed to screen and five to rank (ten in the third window); score thresholds set from controls and decoys and fixed before production	The thresholds' values; an extra predictor as a check in the third window
Scores and ranking formulas; novelty thresholds and diversity caps; several strategies for binding multiple toxins, each written down beforehand with a success criterion	The strategies themselves, and changes to the design process in the second window
A default team, which Claude could refine but had to keep hierarchical: one orchestrator, with separate agents for each design set and each shared service; compute allocated by evidence, not in equal shares	How compute was divided; how much to use each design tool, above a minimum the prompt set; the schedule and budget split in the third window
A license rule (open-source tools whose licenses permit commercial use), with 13 example design tools, eight of them required to supply at least 50 backbones per design set; rules on how results may be described; and safety rules	Which other tools to run: four of the five examples that were not required

Table S5. Claude chose five hotspots on each toxin, 25 in all. Hotspots are the residues that Claude chose for designs to contact, numbered from the first residue of each toxin's sequence. The 243 design models are in the data release; for 25 designs of Set 10, they are files built or rebuilt in their place. In parentheses, the models with a design Cβ atom within 8 Å of the hotspot's Cβ; the 14 models without a design Cβ atom count as no contact. Some designs of Sets 6 to 8 were generated with other hotspot lists.

Toxin	Design models	Hotspots (design models in contact)
Cn2	83	E15 (79), Y24 (13), R27 (3), F44 (79), W58 (46)
Ts1	47	H8 (6), K12 (39), R25 (3), I29 (3), W39 (42)
AaH II	60	F15 (50), W38 (51), Y42 (17), R62 (32), H64 (29)
AaH I	35	Y8 (5), Y14 (4), F36 (34), V38 (34), L42 (32)
LqhIII	18	H15 (7), F39 (17), V41 (17), H43 (10), L45 (17)

Claude worked in three windows, each ending with a release of frozen designs, and, as instructed, asked us no questions. The items below follow the markers in Figure 2B.

P. Our prompt opened the original window we had planned, which released no design that passed the score thresholds against all five toxins, but showed promising results, motivating us to extend the design campaign.

1. The first follow-up extended the campaign, froze Sets 1 to 7, cut Set 8 from 60 to 30 designs, and added the 30 fusions for Set 9. This follow-up was meant to encourage Claude to continue pushing for a five-toxin binder.
2. The second follow-up asked Claude to send no designs to Twist Bioscience before our review.
3. The third follow-up extended the campaign again, allowed swaps in the frozen sets, and changed the selection keys. Most of the 180 designs entered the final list only then, among them Cn2-B1.

4. Because of our fourth follow-up, Set 8 holds the 30 single domains that passed the score thresholds against the most toxins. In place of the diversity rules (Section S1.1), places were filled in order of the number of toxins passed, with at least 10 substitutions between designs and no cap per backbone. All 30 passed against all five toxins; 28 are variants of one binding mode on one backbone, and 2 use a second backbone (Table S6).

The campaign's records show one other human input that Claude acted on: Twist Bioscience's reply to the format questions that Claude sent at our first follow-up's request. At the third follow-up's request, Claude sent Twist Bioscience the 180 sequences. We drafted the prompt and the four follow-ups, and followed Claude's work between windows, in a separate session with Claude, which informed the campaign only through those drafts.

The campaign ran on Modal, using H100 GPUs in up to about 350 containers at a time, and used on the order of 10,000 GPU hours. Claude gave scoring to one service and selection to one coordinating agent, so no agent that generated a design also scored or selected it. Claude fixed its score thresholds before any design was scored. The third follow-up barred N-X-S/T sequons, histidine runs, and an asparagine in the last two positions, and no selected design carries one.

Table S6. The design sets, from generation to selection. Counts are cumulative over the three windows. 'Screened' and 'Five-seed' count design-toxin pairs screened with one seed and co-folded again with five, so for Sets 6 to 9 they can exceed the designs generated. 'Passed' counts designs that passed the score thresholds against every intended toxin in the final release. 'Backbones' counts distinct starting backbones among the selected designs; the N-terminal domains of the Set 9 fusions come from four backbones. Set 10, the 31 designs added later: 29 single domains on the two backbones of Set 8, and 2 fusions. [a]By the same thresholds; the scores of 28 of the 31 come from a re-run after the campaign. By the campaign's own scores, 20 had passed (Section S1.1).

Set	Intended toxins	Selected	Generated	Screened	Five-seed	Passed	Backbones
1	Cn2	12	20,189	12,301	1,158	12	12
2	Ts1	12	10,573	8,648	925	12	12
3	AaH II	12	18,964	15,259	746	12	12
4	AaH I	12	18,871	15,414	1,643	12	12
5	LqhIII	12	17,547	13,976	824	12	12
6	Cn2, Ts1	30	40,208	50,139	4,164	30	14
7	AaH II, AaH I, LqhIII	30	75,861	152,724	3,108	29	10
8	All five, single domain	30	115,430	280,406	15,733	30	2
9	All five, fusion	30	841	4,185	150	17	n/a
1 to 9		180	318,484	553,052	28,451		
10	All five, added later	31	n/a	n/a	n/a	19[a]	2
All tested		211					

S2.2 Nearly every selected pair passed the thresholds, and no design resembles a toxin

Of the 510 selected design-toxin pairs, all but 14 passed the score thresholds (Figure S1A; Figure S2 defines the terms), as selection did not require a pass against every intended toxin. The 14 belong to one design in Set 7 and 13 fusions in Set 9 (Table S6). Of the 15 ipSAE thresholds, 7 stood at the floor of 0.15. With ESMFold2, the median scDockQ over intended pairs was 0.97 for single-toxin designs but 0.74 in Set 6, 0.57 in Set 7, 0.37 in Set 8, and 0.59 in Set 9. Part of this shortfall is a result of the reference, because a toxin missing from the design model was superposed onto the toxin in it (Section S1.1). Section S5 compares these scores with the measurements.

The highest TM-score between any aligned model of the 211 designs and any of the five toxins was 0.42 (110493ed against Cn2), below the prompt's limit of 0.5, whereas the toxins reach 0.65 to 0.89 with one another. With the campaign's alignment setting, no local alignment with a toxin reaches the prompt's limit of 40 aligned residues. We aligned each of the designs' 3,897 models to each toxin structure with TM-align [86], normalizing by either length. The models are the ESMFold2 and Protenix v2 monomers of Section S5.1 (five seeds each; 2,110), the design chains of the campaign's best-seed complexes (1,560), and the design models (227). These comparisons rest on predicted models and design models, the AaH I structure was itself predicted, and 16 of the 31 added designs were represented by monomer models only. A distinct fold does not prove that a design is safe.

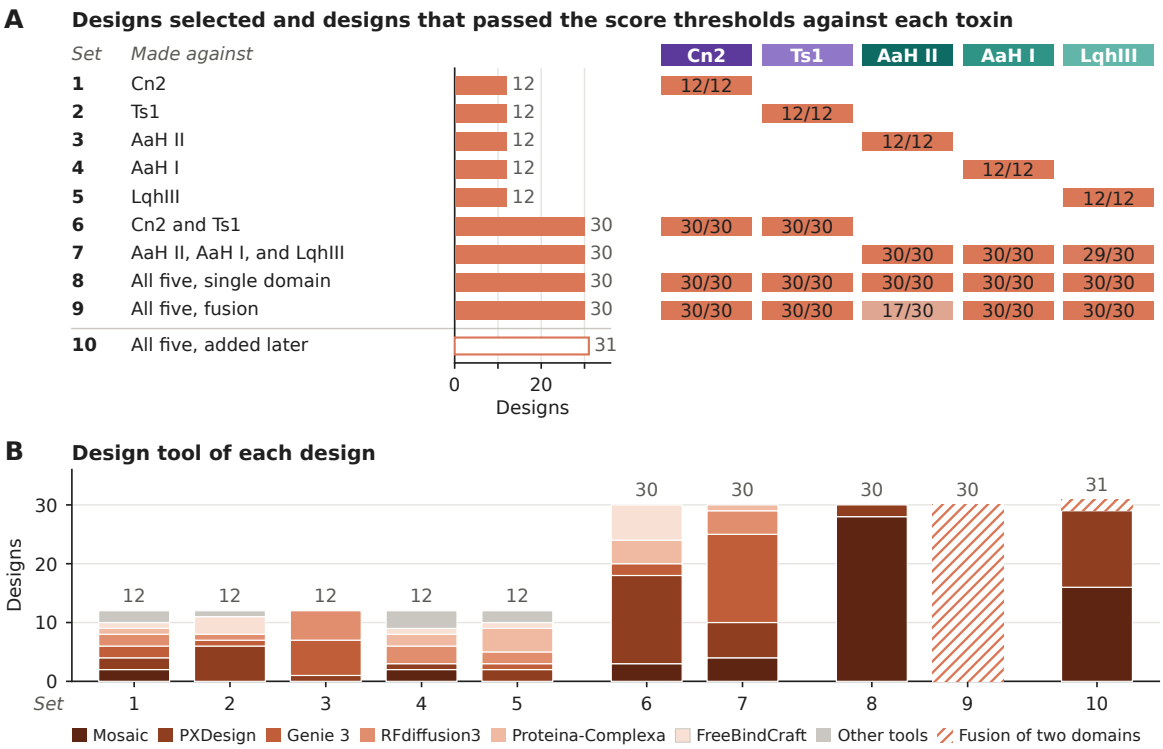


Figure S1. The 211 tested designs and the tools that generated them. A) Designs selected per set (filled, 180; sets as in Table S6) and added later (open, 31); cells, designs that passed the score thresholds against each toxin, out of those selected (496 of 510 pairs). **B)** Design tools (other tools: BoltzGen, Protein Hunter, BoltzDesign1, and RFdiffusion, 8 designs in all); added single domains are assigned the tool of the nearest Set 8 design.

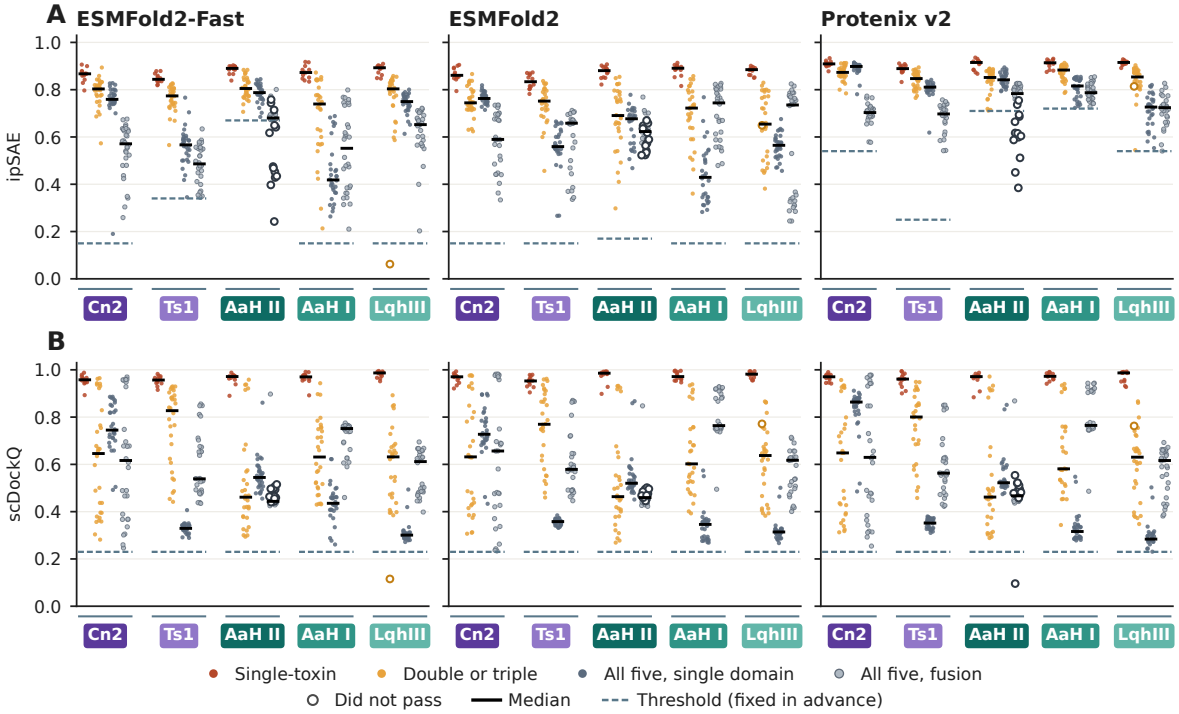


Figure S2. Single-toxin designs scored higher than multi-toxin designs on every term. The six terms of the final score for the 510 selected design-toxin pairs, by predictor and toxin. **A)** ipSAE (lower directional value, best of five seeds). **B)** scDockQ, the DockQ against the design model. Colors, design groups (double β-toxin designs for Cn2 and Ts1, triple α-toxin designs for the other three); bars, medians; open dots, the 14 pairs that did not pass; dashed lines, thresholds (ipSAE, Table S26; scDockQ, 0.23, Table S1).

S3 Experimental details and full results

S3.1 Production and coverage of the designs

Twist Bioscience produced and could measure nearly all of the designs. They reported on 210 of the 211, all but 6a7a1c19, and measured each assessable design once against four toxins. Of those 210, 203 gave product, at a median of 1.6 mg/mL and a median main peak in size-exclusion chromatography (SEC) of 98.3%; 25 peaks were under 90%, none from a binder (Table S7). The other seven gave 0.06 mg/mL or less. Five of them were captured at less than 50 RU and were not assessable (Table S8), and two, 6da28c32 and a8a3f1f9, were captured just above 50 RU and were not hits.

Adaptyv Bio expressed all 211 designs cell-free and measured each against the toxins it was made against, two to four times per pair (Table S8). This covers all 665 such pairs and 679 of the 1,055 possible pairs, with 14 pairs of a design and a toxin it was not made against. Of those 14, only AaHII-B4 against Ts1 was a hit, at both labs (Section S4.3).

Table S7. The binders were produced at least as well as the other designs. Twist Bioscience expressed each design as a human Fc fusion and reported its concentration (absorbance at 280 nm) and the main peak's share of the signal in analytical size-exclusion chromatography (SEC). Purity by CE-SDS was 100% for all 203 designs with product. Adaptyv Bio measured the crude yield, the soluble design in the clarified cell-free reaction, with a split-GFP assay, quantifiable from 2.5 to 25 µM, and, separately, the concentration after purification on Strep-Tactin XT beads (n.d., no value reported). Mass is that of the construct they expressed (design with GFP11 and twin-Strep tags). Capture is the level of design on the SPR chip in each main-screen replicate against the binder's toxin, in the order of the replicates; values in nm give the range over the probes of a replicate run by BLI. Residues count the design without tags or Fc.

		Twist Bioscience		Adaptyv Bio			
Design	Residues	Concentration (mg/mL)	SEC main peak (%)	Crude yield (µM)	Purified (µM)	Mass (kDa)	Capture, by replicate (RU; BLI, nm)
Cn2-B1	104	1.52	100	>25	44.2	17.9	303; 1,099
Cn2-B2	81	1.87	99.3	19.0	26.3	14.8	690; 475
Cn2-B3	103	2.40	100	21.7	25.9	17.5	592; 935; 247
Cn2-B4	118	2.67	100	12.5	18.0	19.3	617; 513
Cn2-B5	110	2.00	98.8	14.8	9.9	18.1	188; 792; 612
Cn2-B6	97	1.86	91.5	>25	19.9	17.7	999; 807
Cn2-B7	74	2.01	98.9	25.0	18.8	14.5	274; 1,003
AaHII-B1	95	2.05	95.7	3.2	18.8	16.5	626; 590
AaHII-B2	74	2.50	99.4	>25	29.2	14.5	693; 724
AaHII-B3	74	2.60	99.2	>25	24.4	14.6	743; 827
AaHII-B4	109	1.25	100	8.3	31.9	18.6	703; 714; 466; 493
AaHI-B1	75	2.32	99.4	22.2	6.2	14.2	915; 3.1 to 3.2 nm
AaHI-B2	93	0.74	99.3	>25	12.8	17.0	869; 535
AaHI-B3	234	0.88	100	7.3	n.d.	33.2	4.5 to 4.7 nm; 651
AaHI-B4	107	1.88	90.8	<2.5	2.9	18.7	3.9 to 4.1 nm; 480
LqhIII-B1	99	1.47	100	15.7	11.4	16.7	602; 725; 645
LqhIII-B2	78	1.59	100	19.9	5.1	15.0	644; 838; 695

At Twist Bioscience, median (range) over designs with product: binders (17), 1.9 (0.74 to 2.7) mg/mL and 99.4% (90.8 to 100%); single domains (171), 1.7 (0.10 to 2.7) mg/mL and 98.6% (33.7 to 100%); fusions (32), 1.0 (0.41 to 2.4) mg/mL and 94.9% (61.5 to 100%). At Adaptyv Bio, the median crude yield over all 211 designs was 6.7 µM (51 below 2.5 µM and 23 above 25 µM), and the binders' median was 20 µM (1 below and 5 above the range).

S3.2 Adaptyv Bio's fits and replicates

We fitted each pair's single Twist Bioscience series with a 1:1 model (Section S1.2). Adaptyv Bio fitted all injections of each replicate together with their own pipeline, sharing one on-rate, off-rate, R_{max}, and drift slope, with a bulk refractive-index offset for each phase (Table S8). A pair's K_D is the arithmetic mean over its fitted replicates, by SPR or BLI. For every binder, Adaptyv Bio selected the 1:1 model in the main screen. For five main-screen pairs, they selected a conformational-change model: a981802c and 6d0b4f3f against AaH I, and 33b2180d, 6d0b4f3f, and f8242edc against LqhIII, all fusions and none a binder. They did the same for 19 of the 29 reverse BLI replicates against AaH I. In these cases, we report their own 1:1 fits of the same measurements, marked approximate, so that all K_D values come from one model. In reverse BLI against Cn2, Adaptyv Bio selected the 1:1 model for every design so that the values can be compared across designs, although, by their team's account, some designs fit other models better.

Against AaH I, all 19 of Adaptyv Bio's hits were hits, with a fit, in every replicate. The replicates of a pair differ by a median of 3.3-fold (1.1- to 16-fold), and those of the four binders differ 3.0- to 13-fold. The widest of these, for AaHI-B2, lies between two SPR replicates, so it is not a difference between SPR and BLI. For AaHI-B2, AaHI-B3, and AaHI-B4, the difference lies mainly in the on-rates (11- to 30-fold) and exceeds the gap between the labs' values (1.1- to 2.0-fold), so these means are approximate (Table S18). The replicates of the other 12 binders with two or more fits differ at most 2.8-fold.

Table S8. Designs measured, replicates per pair, and the fit behind each K_D, by lab and toxin. Designs: those measured against the toxin; at Twist Bioscience, those assessable (at least 50 RU captured) of the 210 they reported. Replicates: replicates per pair, with the number of pairs in parentheses. Model: the fit model of each reported K_D, with the number of pairs (at Twist Bioscience, the 48 that were hits and 22 weak or ambiguous ones). Adaptyv Bio's 1,713 replicates include 39 run by BLI (5 against Cn2, 10 against Ts1, 6 against AaH II, and 18 against AaH I), and the mean K_D values of AaHII-B4 and f8242edc against Ts1 and of 18 pairs against AaH I each include one BLI fit. Adaptyv Bio's pairs include 14 with toxins the design was not made against: Cn2-B1, Cn2-B2, and Cn2-B4 against AaH II and Ts1, and AaHII-B1 to AaHII-B4 against Cn2 and Ts1.

Toxin	Designs	Replicates per pair (pairs)	Model (pairs)	K_D of a pair
Twist Bioscience, fitted by us				
Cn2	205	1 (all)	1:1 (16)	Our single fit; for Cn2-B1, an upper limit
Ts1	205	1 (all)	1:1 (10)	Our single fit
AaH II	205	1 (all)	1:1 (21)	Our single fit
AaH I	205	1 (all)	1:1 (23)	Our single fit
LqhIII	not measured			
Adaptyv Bio, fitted by their team				
Cn2	137	2 (122), 3 (14), 4 (1)	1:1 (8)	Arithmetic mean over replicates; for Cn2-B1, our upper limit
Ts1	140	2 (111), 3 (27), 4 (2)	1:1 (7)	Arithmetic mean over replicates
AaH II	136	2 (70), 3 (65), 4 (1)	1:1 (7)	Arithmetic mean over replicates
AaH I	133	2 (18), 3 (115)	1:1 (19)	Arithmetic mean over replicates, SPR and BLI alike
LqhIII	133	2 (7), 3 (126)	1:1 (7)	Arithmetic mean over replicates

S3.3 The toxins and the theoretical R_{max}

Table S9 lists the five toxins and the structures used for design. Twist Bioscience's three commercial toxins carry tags (Table S11), and their product sheets report purity by reducing SDS-PAGE and no test of activity. No control excludes binding to the His-SUMO tag of AaH II. The position of each tag matters, because many predicted interfaces reach the toxin's C-terminus. In the predicted complexes (ESMFold2), 12 of the 17 binders, Cn2-B1 among them, have a heavy atom within 5 Å of the toxin's C-terminal residue, and only Cn2-B2 has one within 5 Å of the N-terminal residue. The last residue of AaH II, H64, is a hotspot (Table S5), and the last four residues of Cn2 belong to its channel-binding site (defined in Section S4.1). Of the documented constructs, Ts1 carries a C-terminal tag at both labs, and AaH I and LqhIII carry one at Adaptyv Bio (Table S11).

For Cn2, 9 of the 12 single-toxin designs contact the C-terminal residue, and 4 of the 9 are binders. For Ts1, 11 of the 12 contact it, as do 28 of the 30 double β-toxin designs. The C-terminal tag could therefore have covered the designed site on Ts1 at both labs. Of the 6 binders against AaH I and LqhIII, however, 4 contact the toxin's C-terminal residue in their predicted complexes, and all 6 were hits against these tagged constructs, so a C-terminal tag does not by itself prevent binding. Misfolding is a second possible explanation, since recombinant Ts1 refolded from *E. coli* inclusion bodies was reported to form a stable, disordered disulfide isomer [72]. Ts1 without a C-terminal tag, or purified from venom, would distinguish both explanations from designs that do not bind. Until then, the lack of Ts1 binders cannot be attributed to the designs, and so the result for the 30 double β-toxin designs cannot be either. Native Cn2, Ts1, AaH II, and LqhIII also end in a C-terminal amide, which recombinant proteins from yeast or *E. coli* are not expected to carry.

Whether a pair stays within the theoretical R_{max} depends on the toxin's mass. We used the masses recovered from the theoretical R_{max} values that Twist Bioscience reported (Table S10). These equal the datasheet masses of Cn2 and Ts1; the mass of AaH II is above its datasheet mass and equals the mass on two distributors' product sheets; AaH I has no datasheet. At the mature mass of AaH I, its four binders would exceed the theoretical R_{max}, as would 9C2 (R_{max} ratio of 1.43 instead of 0.70). If 9C2 binds one to one, AaH I as used must weigh at least 9.7 kDa per binding site. At 9.7 kDa, AaHI-B2 is at 1.13 times the theoretical R_{max}, and from 11.0 kDa all four binders stay within it. The 15 binders measured at Twist Bioscience reached 39 to 82% of their theoretical R_{max}, assuming saturation within the injection, so such binders were detectable against Cn2, AaH II, and AaH I. Against Ts1, the limit for a hit left a binder a far narrower range, and all five hits exceeded the theoretical R_{max}.

We also checked criterion 3 on Adaptyv Bio's data, which the labels do not use. With the capture levels and the construct masses (Tables S7 and S11), we compared the response at the end of the 1 µM injection of each SPR replicate with the theoretical R_{max} for one toxin per captured design. The 11 binders against Cn2 and AaH II reached 15 to 70% of it, and the six against AaH I and LqhIII reached at most 39%. By BLI, AaHI-B1, AaHI-B3, and AaHI-B4 reached 0.70 to 0.92 times the theoretical R_{max}, and 9C2 reached 0.86 and 0.96. AaHII-B4 against Ts1 is the one pair that criterion 3 alone excludes at Twist Bioscience, and at Adaptyv Bio this single domain reached 2.4 times the theoretical R_{max} on both SPR spots measured, which were in one run. This check was less strict than the one at Twist Bioscience. The capture level was recorded when the design was captured, so any design that left the surface afterward lowered the true limit below the one computed. We therefore did not use this check to label binders.

Table S9. The five toxins span both families of sodium-channel toxins. Family gives the toxin family and its receptor site on the sodium channel. Mature chain gives the residues of the UniProt entry that form the mature toxin, with their number in parentheses; for AaH I, the 63-residue isoform that the prompt specified, not the 64-residue sequence of the UniProt entry. Claude chose the structures from those named in the prompt's reading list. Experimental structures of four of the toxins have been published [87–90]; we found none of AaH I, for which the prompt required at least two predicted models. Cryo-electron microscopy (cryo-EM) has shown AaH II, LqhIII, and Cn2 in complex with Na_V channel proteins [44, 47, 48].

Toxin	Species	Family, site	UniProt	Mature chain (aa)	Structure used for design
Cn2	*Centruroides noxius*	β, site 4	P01495	17 to 82 (66)	PDB 25IH, chain D; with $Na_V1.6$ (cryo-EM) [44]
Ts1	*Tityus serrulatus*	β, site 4	P15226	21 to 81 (61)	PDB 1NPI, chain A; free toxin (X-ray) [45]
AaH II	*Androctonus australis*	α, site 3	P01484	20 to 83 (64)	PDB 1AHO, chain A; free toxin (X-ray) [46]
AaH I	*Androctonus australis*	α, site 3	P01479	20 to 82 (63)	AlphaFold DB AF-P01479-F1 (v6) [85], trimmed; Claude's predictions with ESMFold2 and Protenix v2
LqhIII	*Leiurus hebraeus*	α-like, site 3	P56678	1 to 67 (67)	PDB 7K18, chain B; with $Na_V1.5$ (cryo-EM) [47]

Table S10. Toxin masses used to compute the theoretical R_{max}. Left, the constant K (toxin mass times binding sites) recovered from Twist Bioscience's values, the mass per site at two sites per captured Fc dimer, the supplier's predicted mass, and the mature-toxin mass (kDa). Right, for the eight AaH II and AaH I pairs among the 15 that met all criteria, the ratio of our fitted R_{max} to the theoretical R_{max}, as used in this report and rescaled to 20.2 kDa (AaH II) or to the mature 6.80 kDa (AaH I). Above 1, a pair would not meet criterion 3.

Protein	K (Da)	Per site	Datasheet	Mature	Pair	Toxin	As used	Rescaled
Cn2	19,200	9.6	9.6	7.59	AaHII-B1	AaH II	0.56	0.65
Ts1	28,600	14.3	14.3	6.88	AaHII-B2	AaH II	0.46	0.53
AaH II	46,600	23.3	20.2	7.24	AaHII-B3	AaH II	0.57	0.66
AaH I	27,800	13.9	none	6.80	AaHII-B4	AaH II	0.54	0.62
CLEC12A	51,200	25.6	none		AaHI-B1	AaH I	0.62	1.26
					AaHI-B2	AaH I	0.79	1.61
					AaHI-B3	AaH I	0.59	1.21
					AaHI-B4	AaH I	0.57	1.17

S3.4 Positive controls

Adaptyv Bio ran all six positive controls against Cn2, Ts1, and AaH II as single-chain variable fragments (scFvs) and all but serrumab as Fabs (Table S12). The scFv of 9C2 did not express in the Ts1 screen. Against AaH I they ran 13 constructs of the six antibodies, and against LqhIII nine constructs of five, each once and all but one by BLI (Tables S11 and S12). Controls run against their own toxin in both formats gave the same result in each (Figures S3 and S4A; Table S11): 9004G, LR, and RU1 bound Cn2 in both, 9C2 bound AaH I in both, and 4C1 bound AaH II in neither.

Cn2. At Adaptyv Bio, no binding against Cn2 was detected for any antibody raised against another toxin, in either format. 9004G and LR also bound Twist Bioscience's Cn2, which came from another supplier, consistent with folded toxin at both labs. RU1 had a steady-state K_D of 13 nM in Twist Bioscience's SPR and no kinetic fit. In Twist Bioscience's competition BLI, LR and RU1 bound the captured Cn2 (Section S4.1).

Ts1. No control shows that Ts1 was folded at either lab. Twist Bioscience's anti-Ts1 antibody did not pass the quality check (Table S11). At Adaptyv Bio, serrumab gave a K_D of 15 nM (12 and 17 nM). Fabs raised against other toxins gave similar values: LR, 18 nM; 9004G, 19 nM; 9C2, 31 nM; and RU1, 66 nM. The scFvs of 9004G, LR, and RU1 and both formats of 4C1 gave none (Figure S4). By BLI, with the antibody on the sensor and Ts1 at 1 µM, the Fabs of 9004G, LR, and 9C2 again gave signals (0.6 to 4.1 nm). The three scFvs gave at most 0.09 nm. Adaptyv Bio attributes these signals to the reformatting as Fabs with human constant domains, which are mispaired with the mouse variable domains of 9C2 and 4C1, and not to recognition of Ts1. The Ts1 preparation passed their incoming quality control, and the same Fabs were not hits against AaH II in the main screen. Whatever the cause, antibodies raised against other toxins gave values like serrumab's, so we treat serrumab's result as inconclusive.

Negative responses were common against Ts1 at Adaptyv Bio. The response at 1 µM lay below −5 RU in 154 of 301 SPR replicates of designs against Ts1 (99 of 140 designs), compared with 48 of 1,373 replicates against the other four toxins. Negative responses arise when a toxin gives more signal on the reference surface than on a design's spot, or when the baseline drifts downward, and either could hide weak binding.

AaH II. 4C1 worked at neither lab, so no antibody of known epitope shows that the toxin was folded. At Adaptyv Bio, no binding was detected for it in either format (0 of 5 replicates) or against any other toxin. At Twist Bioscience, its fit did not pass the quality check (Table S11). None of the other nine control constructs was a hit against AaH II in Adaptyv Bio's main screen (Figure S4A).

AaH I and LqhIII. 9C2 bound AaH I at both labs (Table S11): Twist Bioscience's custom-made preparation and the one that Adaptyv Bio expressed in-house (Fab and scFv, K_D of 3.4 and 1.4 nM by BLI, below the lowest concentration tested). None of the other 11 constructs was a hit against Adaptyv Bio's AaH I. At Twist Bioscience, the two toxins whose controls did not work, Ts1 and AaH II, were expressed in *E. coli*, whereas Cn2 was expressed in yeast. None of the nine constructs run against LqhIII was

Table S11. Each toxin preparation, its positive control's result, and the designs measured against it, by lab. Preparation: supplier and catalog number as given by the teams; host and tags, with the tags' positions, from the suppliers' and distributors' product sheets for these catalog numbers or, for Adaptyv Bio's preparations, as given by their team, with residues numbered as in the UniProt entries (Table S9). Affinities and responses are from SPR unless BLI is named. A control works if its fit passes Twist Bioscience's quality check or, at Adaptyv Bio, if its response exceeds the limit (designs' median plus 5 robust SD) or they selected a fit. Its result is inconclusive if antibodies raised against other toxins gave similar fits. Designs measured: at Twist Bioscience, the 205 assessable of the 210 run. Hits: all, with those of designs made against that toxin in parentheses; Adaptyv Bio's counts include 14 pairs of designs with toxins they were not made against. Binders: criteria in Figure 3D.

Toxin	Preparation	Positive control and result	Designs measured	Hits	Binders
Twist Bioscience					
Cn2	Cusabio CSB-YP355662DRB; yeast; N-terminal His tag	9004G and LR: picomolar fits, below the lowest concentration; 0.97 and 0.94 of the theoretical R_{max}; both work	205	12 (12 of 128)	7
Ts1	Cusabio CSB-EP323347TON; *E. coli*; N-terminal His tag and C-terminal Myc tag	Anti-Ts1 antibody, two spots: 196 and 184 nM; 11.0 and 11.3 times the theoretical R_{max}; fails the quality check	205	5 (4 of 128)	0
AaH II	Cusabio CSB-EP355659AJZ; *E. coli*; N-terminal His-SUMO tag; datasheet mass 20.2 kDa	4C1: 49.8 nM; 19.7 times the theoretical R_{max}; fails the quality check; no signal in BLI	205	12 (12 of 132)	4
AaH I	Cusabio CSB-YP365567AJZ, made to order; construct not documented	9C2: 2.2 nM, below the lowest concentration tested; 0.70 of the theoretical R_{max}; works	205	19 (18 of 131)	4
LqhIII	Cusabio CSB-EP348725LDS, made to order; not available in time	Not measured			
Adaptyv Bio, main screen					
Cn2	Biomatik RPC25213; yeast; N-terminal His tag; residues 17 to 82; 9.6 kDa	9004G, LR, and RU1, each as Fab and as scFv: 14 of 14 replicates fitted; all work	137	10 (10 of 133)	7
Ts1	Biomatik RPC28240; *E. coli*; N-terminal 10×His tag and C-terminal Myc tag; residues 21 to 81; 14.3 kDa	Serrumab: 2 of 2 replicates fitted, 12 and 17 nM. Fabs of four antibodies raised against other toxins also fitted, 18 to 66 nM; inconclusive	140	7 (6 of 133)	0
AaH II	Biomatik RPC22662; *E. coli*; N-terminal His-SUMO tag; residues 20 to 83; 20.2 kDa	4C1 as Fab and as scFv: no fit; Fab 11, −11, and −14 RU (limits 90 and 93); did not bind. No other control was a hit	136	7 (7 of 133)	4
AaH I	Expressed in-house, cell-free in an *E. coli* system; N-terminal methionine and C-terminal GFP11 and 8×His tags; residues 20 to 83, one more than the isoform used for design; 11.2 kDa	9C2 as Fab and as scFv, by BLI, one replicate each: 3.4 and 1.4 nM, below the lowest concentration tested (31.6 nM); works	133	19 (19 of 133)	4
LqhIII	Expressed in-house, cell-free in an *E. coli* system; N-terminal methionine and C-terminal GFP11 and 8×His tags; residues 1 to 67; 11.3 kDa	Nine constructs of five antibodies by BLI, none raised against LqhIII: no hit	133	8 (8 of 133)	2
Adaptyv Bio, second supplier (Section S3.7)					
Cn2	Cusabio CSB-YP355662DRB, as at Twist Bioscience	9004G, LR, and RU1, in both formats: 6 of 6 replicates fitted; all work	137	11 (11 of 133)	
AaH II	Cusabio CSB-EP355659AJZ, as at Twist Bioscience	4C1, in both formats: no hit. Fab of 9004G: a hit, no fit	132	4 (4 of 129)	

raised against it, and none was a hit. Adaptyv Bio's K_D values against LqhIII therefore come without a working control (Table S11). Serrumab, reported to reduce the effect of LqhIII on Na_V channels [40], is a candidate control. It was not among the nine, but as an scFv on the probe in Adaptyv Bio's competition assay it captured no LqhIII (at most 0.02 nm in eight reads at 200 nM).

S3.5 Hits across all 1,055 design-toxin pairs

Of the 1,055 design-toxin pairs, 64 were hits at one lab or both, and 17 of these met the criteria for binding (Figure S5; Table S13). Another 902 were not hits, and 89 had no assessable measurement. Of the 35 pairs that were hits at both labs, 34 were hits in every replicate, and 15 of these met the other three criteria (Figure 3D). The other 2 binders, both against LqhIII, are among the 29 hits at one lab only and met the single-lab rule (Table S14). The 17 pairs belong to 17 of the 211 designs (8.1%, Wilson interval 5.1

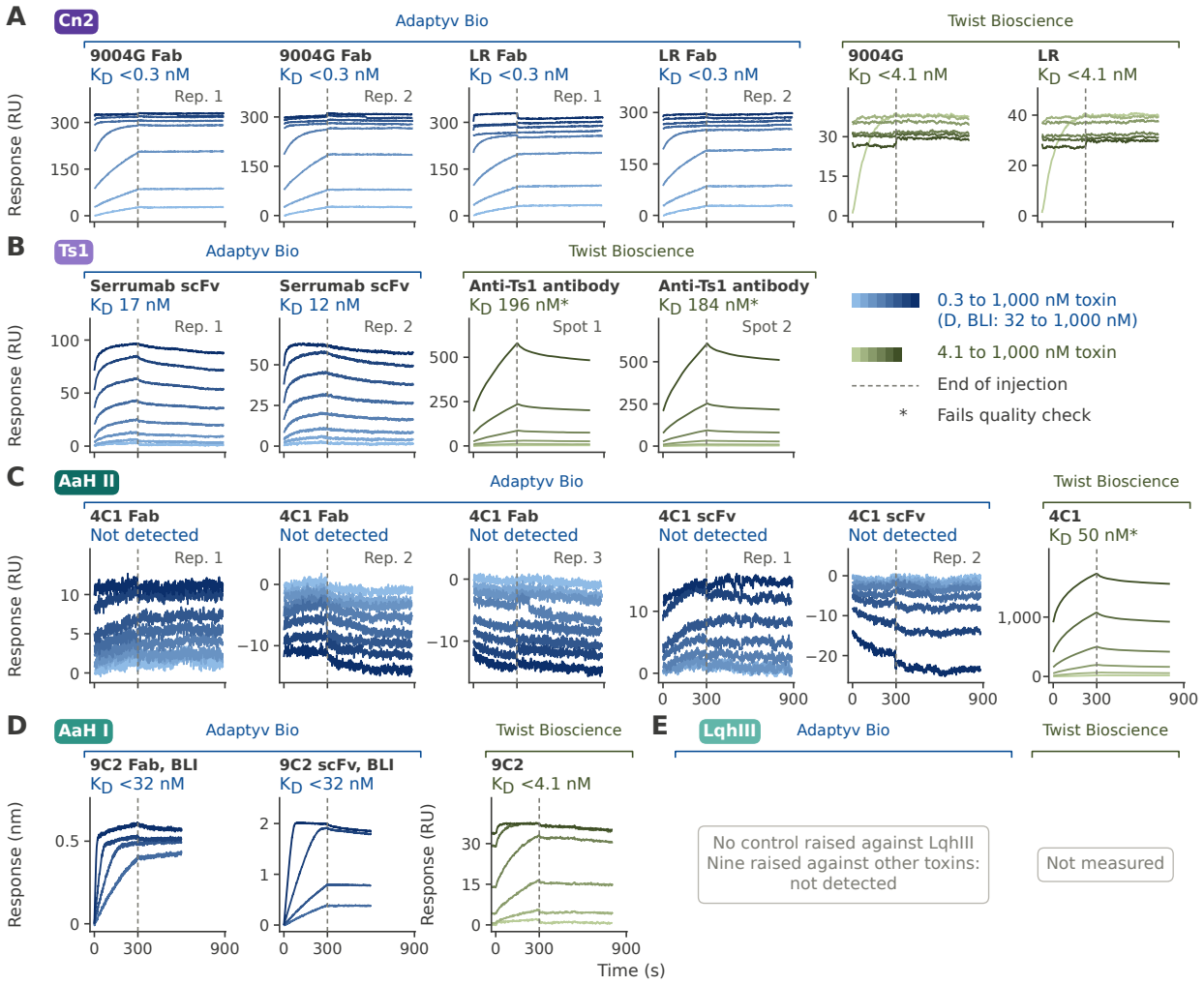


Figure S3. Positive controls bound Cn2 and AaH I at both labs; the control against Ts1 at Adaptyv Bio is inconclusive. Main-screen sensorgrams of the controls (SPR; BLI at Adaptyv Bio in D) against **A)** Cn2, **B)** Ts1, **C)** AaH II, **D)** AaH I, and **E)** LqhIII. Every replicate is drawn, with two exceptions against Cn2. At Adaptyv Bio, the Fabs of 9004G and LR stand for all 14 replicates (Figure S4A). At Twist Bioscience, RU1, which had no kinetic fit, is not drawn. Serrumab against Ts1 is inconclusive because Fabs raised against other toxins also gave signals against Ts1 (Figure S4). Headers give the control, its format where the sequence shows it, the method where it is BLI, and the fitted K_D (Adaptyv Bio's or Twist Bioscience's own 1:1 fit). In BLI, the dashed line marks the end of the association step. Not detected: no binding was detected. Fits below the lowest concentration tested are given as <0.3 nM (SPR) or <32 nM (BLI) at Adaptyv Bio, or <4.1 nM (Twist Bioscience). No control was raised against LqhIII; no binding was detected for the nine controls raised against other toxins that Adaptyv Bio ran against it (Figure S4A). Not measured: Twist Bioscience ran no control against LqhIII.

Table S12. The antibodies and other controls named in this report, with each team's names for them. Descriptions give the published form. Adaptyv Bio names each construct by antibody and format (Figure S4): Fab (builds v1, v2, or alt) or scFv (domain order VH-VL or VL-VH). After 'AaH I', the formats run against AaH I; before it, those run against the other toxins. Twist Bioscience's BLI names come from the competition assay (for 9C2, from reverse BLI). Outcomes: Table S11 and Section S3.4. scFv, single-chain variable fragment; n/a, not used at that lab.

Name	What it is	Twist Bioscience	Adaptyv Bio
Cn2			
9004G	Human scFv against Cn2, in PDB 2YC1 [66]	SPR 'anticn2_2yc1'; BLI 'anti-Cn2_2YC1'	Fab; scFv VH-VL. AaH I: Fab alt; scFv VL-VH
LR	Human scFv against Cn2, derived from 9004G; with RU1 in PDB 4V1D [18]	SPR 'anticn2_4v1d'; BLI 'anti-Cn2_4V1D'	Fab; scFv VH-VL. AaH I: Fab alt; scFv VL-VH
RU1	Human scFv against Cn2, from another parent; with LR in PDB 4V1D [18]	SPR 'anticn2_4v1d_ru1'; BLI 'anti-Cn2_4V1D_RU1'	Fab (v1); scFv VH-VL. AaH I: Fab (v2); scFv VL-VH
Ts1			
Anti-Ts1 antibody	Not stated	SPR 'anti-ts1toxin_1' and 'anti-ts1toxin_2'	n/a
Serrumab	Human scFv that partly neutralizes Ts1 [40]; sequence from patent application BR 10 2013 005043-1	BLI 'Serrumab'	scFv VH-VL. AaH I: Fab; scFv VH-VL alt; scFv VL-VH
AaH II			
4C1	Mouse monoclonal antibody against AaH II, in PDB 4AEI [67]	SPR 'antiaah2_4aei'; BLI 'anti-AaH2_4AEI'	Fab (v1); scFv VH-VL. AaH I: Fab (v2); scFv VL-VH
AaH I			
9C2	Mouse monoclonal antibody against AaH I, without toxin in PDB 4AEH [67]	SPR 'antiaah1_4aeh'; BLI 'antiAaH1_4AEH'	Fab (v1); scFv VH-VL. AaH I: Fab (v2); scFv VL-VH
Negative and off-target controls			
Trastuzumab	Antibody against HER2, unrelated to the toxins	SPR, two spots	n/a
Buffer	No antibody	SPR, two spots	n/a
Polyclonal human IgG	Unrelated human antibodies	BLI, one control sensor or series per run; none in the runs on AaH I	n/a
CLEC12A	Human protein unrelated to the toxins, the off-target analyte	SPR, two runs	n/a
Ovalbumin	Egg-white protein unrelated to the toxins, the analyte of the off-target control (Section S4.4)	n/a	SPR, two runs

Table S13. From measured pairs to binders, toxin by toxin. Counts are design-toxin pairs in the main screen. Measured: pairs with a result at Adaptyv Bio, or assessable at Twist Bioscience (at least 50 RU captured). Both labs: hits at both labs. Criterion 1: a hit at both labs in every replicate or, for LqhIII, at Adaptyv Bio in every replicate. Not met: pairs that met criterion 1 but not criterion 2 (every K_D at 1 μM or tighter; for LqhIII, a K_D reported), 3 (within the theoretical R_{max}), or 4 (no signal against CLEC12A); a pair can miss more than one. Criteria: Figure 3D. n/a, not applicable.

	Adaptyv Bio		Twist Bioscience				Not met			
Toxin	Measured	Hits	Measured	Hits	Both labs	Criterion 1	2	3	4	Binders
Cn2	137	10	205	12	8	7	0	0	0	7
Ts1	140	7	205	5	4	4	0	4	3	0
AaH II	136	7	205	12	7	7	3	0	0	4
AaH I	133	19	205	19	16	16	0	0	12	4
LqhIII	133	8	not measured		n/a	7	0	n/a	5	2
All	679	51	820	48	35	41	3	4	20	17

to 12.5%), 15 of them hits at both labs (7.1%, 4.4 to 11.4%). Among single-toxin designs, 11 of 60 were binders (18.3%, 10.6 to 29.9%; Figure 7). These intervals are too narrow, because designs that share a starting backbone are not independent. Figure 7 counts intended toxins only and so leaves out three hits against other toxins, one at Adaptyv Bio and two at Twist Bioscience (Section S4.3).

A **The control antibodies at Adaptyv Bio, as Fab and as scFv**

Main screen, K_D (nM)

			Cn2	Ts1	AaH II	AaH I	LqhIII
9004G	Cn2	Fab	<0.3	19	None	None	None
		scFv	<0.3	None	None	None	None
LR	Cn2	Fab	<0.3	18	None	None	None
		scFv	<0.3	None	None	None	None
RU1	Cn2	Fab	14	66	None	None	None
		scFv	3.6	None	None	None	None
4C1	AaH II	Fab	None	None	None	None	Not measured
		scFv	None	None	None	None	None
9C2	AaH I	Fab	None	31	None	<32	None
		scFv	None	Not measured	None	<32	None
Serrumab	Ts1	Fab	Not measured	Not measured	Not measured	None	Not measured
		scFv	None	15	None	None	Not measured

Binding detected against its own toxin / another toxin / No binding detected

B **Against Ts1, main screen: each Fab above its scFv**

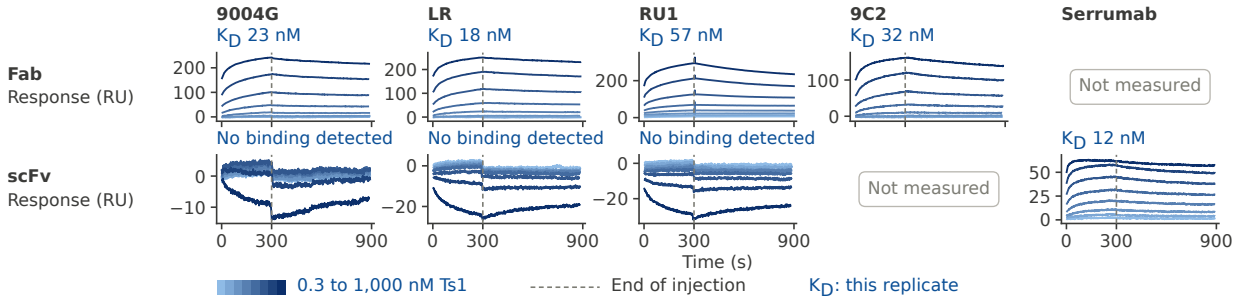


Figure S4. In Adaptyv Bio's main screen, Fabs raised against other toxins gave signals against Ts1 and against no other toxin. A) Main screen: mean K_D over the fitted replicates, by SPR against Cn2 and Ts1 and by BLI against AaH I. <0.3 (SPR) and <32 (BLI): below the lowest concentration tested. None: no binding detected. **B)** Ts1, main screen, highest-numbered SPR replicate with its own K_D; 4C1, with no binding detected in either format, is not drawn. Not measured: the scFv of 9C2 did not express in the Ts1 screen; in the other six cells, that control was not run in that format.

S3.6 Agreement between the two labs

The labs agreed more closely on hits than on affinities, with a Cohen's κ of 0.78 (interval 0.68 to 0.86) on the 533 pairs measured at Adaptyv Bio and assessable at Twist Bioscience. Of these, 53 were hits at either lab and 35 at both, and 10 of the other 18 were fusions. The affinities agreed within 10-fold for 28 and within 3-fold for 16 of the 32 pairs that were hits at both labs with a K_D at both and no upper limit (Figure S6). Spearman's ρ was 0.50 (95% interval 0.16 to 0.74, designs resampled).

Among the 15 binders measured at both labs, the gap between the labs mostly follows the toxin rather than the design (Figure S6). Adaptyv Bio's K_D was a median 1.7 times Twist Bioscience's against Cn2 (6 pairs, 1.2- to 4.0-fold apart), 46 times against AaH II, and 0.54 times against AaH I (4 pairs, 1.1- to 4.1-fold apart). The gap against AaH II lies mainly in the association rates (Figure S10). These rates agree within 2.3-fold against Cn2, and Twist Bioscience's are 10 to 19 times higher against AaH II. Against AaH I, Adaptyv Bio's are 0.7 to 7.0 times Twist Bioscience's. An apparent association rate scales with the share of toxin able to bind, so unequal active shares of toxin, or the designs' different formats and surfaces, could produce such gaps. Against AaH I, the labs' preparations differ, and for three of the four binders Adaptyv Bio's own replicates differ more than the labs do (Section S3.2).

Three surface effects could also distort the rate constants, and we did not vary the flow rate or the surface density to test for them. Transport of toxin to the surface may limit the fastest association rates: in Cn2-B1's two replicates at Adaptyv Bio, the fitted on-rate fell 3.6-fold as the top response rose 4.5-fold. Rebinding of released toxin may also slow the apparent off-rates, an effect that grows with surface density. If part of a tagged toxin is oligomeric, it could be held by avidity at either lab. The two copies on each Fc fusion could hold it at Twist Bioscience, and neighboring immobilized designs could do the same at Adaptyv Bio, more so at higher surface density. At Twist Bioscience, 4C1's signal of 19.7 times the theoretical R_{max} suggests such species in AaH II, and

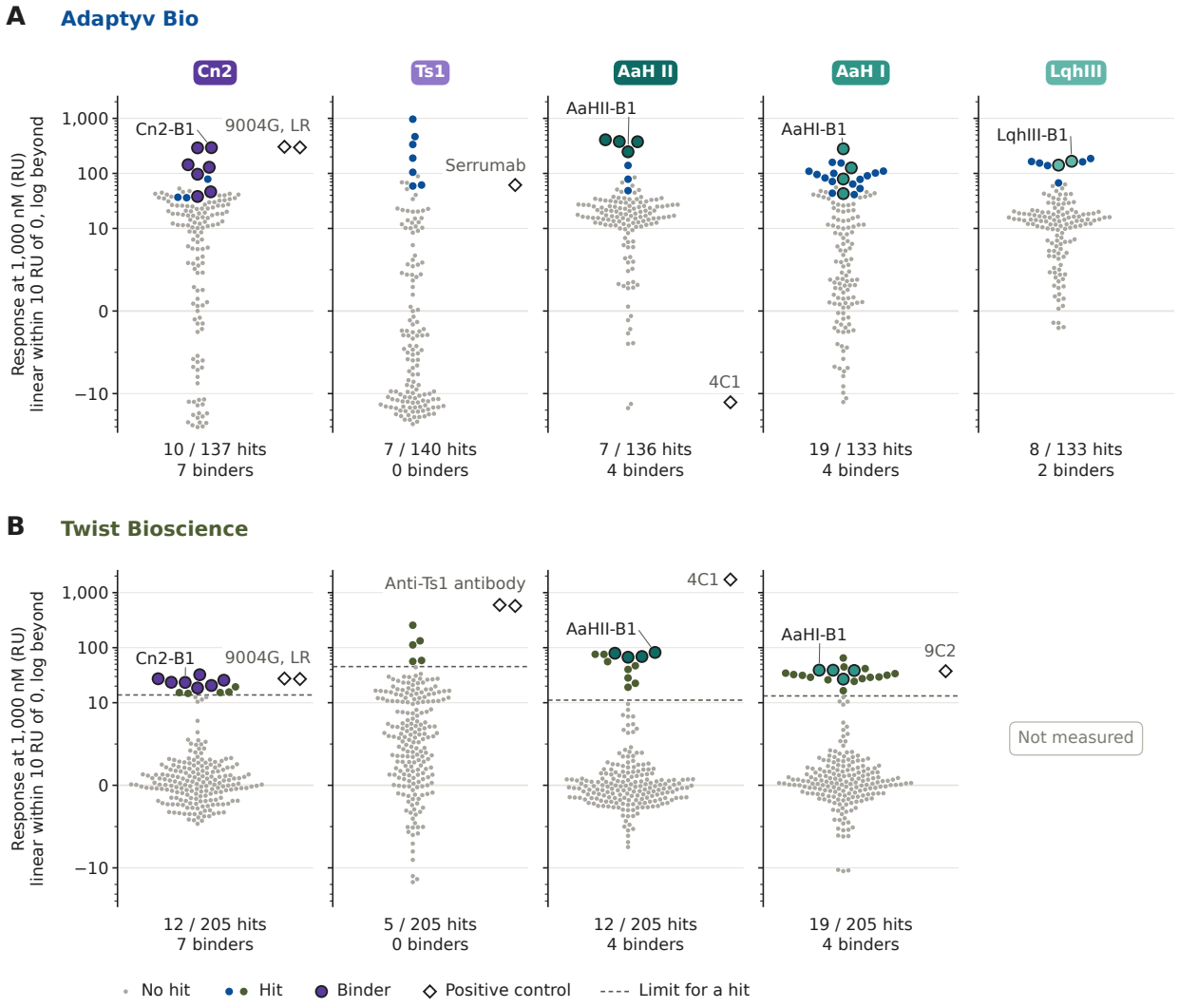


Figure S5. The main screen: 1,499 responses of 211 designs against five toxins at two labs. A) Adaptyv Bio. **B)** Twist Bioscience, who did not measure binding against LqhIII. Each dot is the SPR response of one design against one toxin at the end of the 1,000 nM injection (Adaptyv Bio, highest-numbered SPR replicate). Small gray, no hit; lab color, hit; large with a dark edge, in the toxin's color, binder. Open diamonds, positive controls by SPR; Adaptyv Bio measured 9C2 against AaH I by BLI (Figure S3). Dashed lines, the limit for a hit on Twist Bioscience's data (13.8, 45.3, 11.1, and 13.3 RU), set on all 210 designs of the run, of which the 205 assessable are drawn. Under each strip, hits among the designs measured, and binders.

AaHII-B1 to AaHII-B3 dissociated 2.5 to 4.4 times more slowly than at Adaptyv Bio, although AaHII-B4 dissociated 3.1 times faster. The rate constants may therefore depend on the surface. The upper limits on Cn2-B1's K_D use no rate constant.

S3.7 SPR against toxins from a second supplier

Adaptyv Bio ran the designs by SPR against Cn2 and AaH II from Cusabio, the supplier of Twist Bioscience's toxins (Table S11). Both products carry the same catalog numbers as the Cn2 and AaH II that Twist Bioscience used, but we do not know whether the lots were the same. Each pair had one or two replicates at five concentrations, 10 to 1,000 nM. Against Cn2, 11 of 137 designs were hits: the seven binders, 9a8cdc7a, 3f0f990a, and two designs that were hits at neither lab in the main screen, 4efbc623 and eade0bd3. Against AaH II, 4 of 132 were hits: the four binders (Figure S7). Of the ten binders with a K_D in all three measurements, six lay farther from Twist Bioscience's value than in the main screen (Table S15). Cn2-B3 to Cn2-B7 gave 3.2 to 5.9 times their main-screen values. The AaH II binders stayed within 1.8-fold of theirs, at 2.1 to 50 times Twist Bioscience's. The toxin product thus does not explain the gap between the labs. We did not use these measurements to label binders. Adaptyv Bio's preparation of Ts1 from this supplier (CSB-EP323347TON, the catalog number Twist Bioscience also used) failed their quality check for aggregation, and their team found an assay problem in a second run against AaH II, so we report neither.

Table S14. Four classes of hit, and why pairs that were hits at both labs did not meet the criteria for binding. Pairs are design-toxin pairs that were hits at one lab or both (64). Indented rows split the class above. UUIDs and names are those of the data release, which also holds the K_D values and rate constants of all 35 pairs that were hits at both labs. CLEC12A hit, a hit against CLEC12A, an unrelated protein, which suggests off-target binding. Criterion 3 (theoretical R_{max}) does not apply to LqhIII, which only Adaptyv Bio measured.

Class	Pairs	Criterion, or pairs and reason
Both labs, all criteria met	15	Hits at both labs in every replicate and the other three criteria met (Figure 3D)
Both labs, not all criteria met	20	Hits at both labs and at least one criterion not met
Not a hit in every replicate	1	9a8cdc7a against Cn2, a hit in one of three replicates at Adaptyv Bio
Hit against CLEC12A	14	11 against AaH I and 3 against Ts1; the 3 against Ts1 also exceed the theoretical R_{max}
K_D above 1 μM	3	bbe37cf8, 47dd0a8a, and 1296b21d against AaH II (1.3, 3.4, and 3.4 μM by SPR at Adaptyv Bio)
Near-background signal against CLEC12A	1	cf8652bc against AaH I
Above the theoretical R_{max}	1	AaHII-B4 against Ts1, at 4.2 times the theoretical R_{max}, with hits across toxin families at both labs
One lab only	4	Measured and a hit at one lab only, with a fitted K_D of at most 1 μM and no CLEC12A hit
Labeled binders	2	LqhIII-B1 and LqhIII-B2, single-toxin designs that were hits against LqhIII in three of three replicates
Not labeled binders	2	3f0f990a against Cn2 (CLEC12A not assessable) and 4fcc0f88 against AaH I (irregular traces); the single-lab rule covers only LqhIII
Other hit	25	A hit at one lab only, where the labs disagree, no K_D was reported, or the design was a hit against CLEC12A
Against LqhIII, CLEC12A hit	5	All fusions
No K_D	2	a1f798ca against LqhIII and 925ee271 against Cn2 (not assessable at Twist Bioscience); Adaptyv Bio reported no K_D
Measured at the other lab, no hit there	18	Against Cn2, Ts1, AaH II, or AaH I; 2 of the 18 were weak or ambiguous at the other lab (Figure S9A)

Table S15. The 11 binders against Cn2 and AaH II were hits against both suppliers' toxins at Adaptyv Bio. K_D (nM) by SPR. Second supplier, Adaptyv Bio's 1:1 fit of one replicate. Main screen and Twist Bioscience, as in Table 1 but to two digits. ≤ and <, upper limits (<10, a fit below 1 nM, given as the lowest concentration tested); ≈, approximate (for the second supplier, outside the concentrations tested).

Design	Second supplier	Main screen	Twist Bioscience	Design	Second supplier	Main screen	Twist Bioscience
Cn2-B1	<10	≤3.1	≤1.7	AaHII-B1	85	78	1.7
Cn2-B2	≈2.5	5.6	3.4	AaHII-B2	100	180	3.8
Cn2-B3	592	160	40	AaHII-B3	158	200	3.9
Cn2-B4	383	110	48	AaHII-B4	13	21	6.3
Cn2-B5	286	89	50				
Cn2-B6	371	63	54				
Cn2-B7	≈1,650	350	≈580				

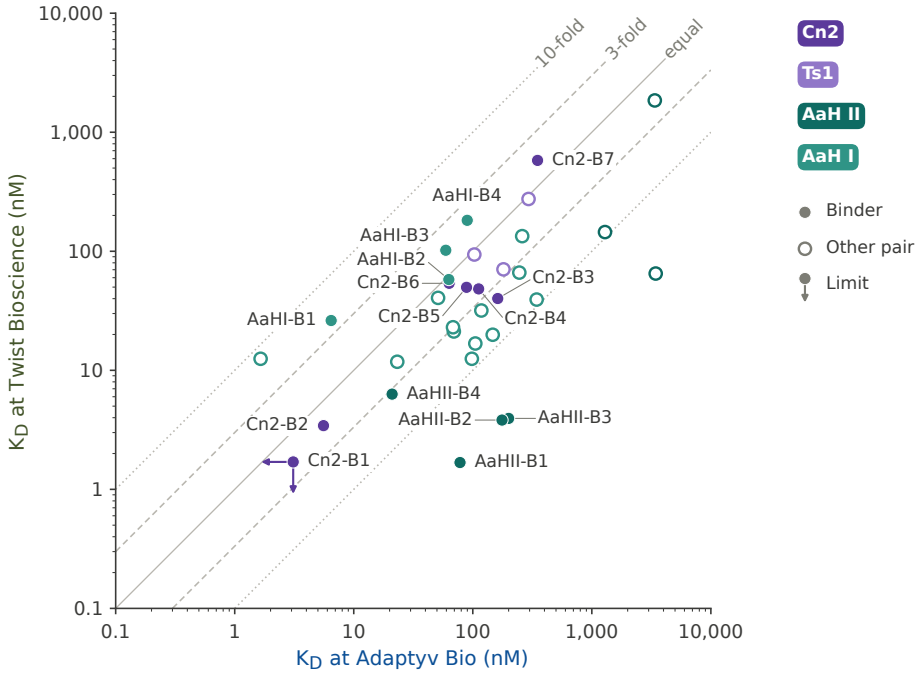


Figure S6. The two labs' K_D values agreed within 10-fold for 28 of 32 pairs, and all four larger gaps were against AaH II. Each mark is one design-toxin pair that was a hit and had a K_D at both labs. Filled, binders; open, other pairs. Lines, equal values and 3-fold and 10-fold differences. Arrows, limits. Not among the 32: Cn2-B1, whose values are upper limits at both labs; and AaHII-B4 against Ts1, whose value at Adaptyv Bio is an upper limit and which is not drawn. 9a8cdc7a against Cn2 had no K_D at Adaptyv Bio and is not drawn. Adaptyv Bio's values are means of the fitted replicates, which include one BLI replicate for 15 of the 16 pairs against AaH I and for f8242edc against Ts1. Means of replicates more than 3-fold apart are approximate (Table 1), and three BLI fits against AaH I lie below the lowest concentration of their series (31.6 nM). All values come from 1:1 fits (Section S3.2).

S3.8 The number of binders under other definitions

Table S16 gives the number of pairs that meet all criteria when the definition of a hit or of a criterion changes. Criterion 1 asks for a hit in every replicate. Counting a hit in any replicate would add 9a8cdc7a against Cn2, a hit in one of three replicates at Adaptyv Bio. Against LqhIII, the triple α-toxin design a1f798ca was also a hit in one of three replicates and has no reported K_D, so it misses two criteria.

Claude also identified hits in Adaptyv Bio's traces by a fixed rule like its rule for Twist Bioscience's data. In at least half of a pair's main-screen SPR replicates, the response over the last 10 s of association at 1,000 nM had to exceed the toxin's median in that run by 5 SD (for the seven runs with fewer than 30 designs, the median and SD of a larger run against the same toxin) and to rise with concentration (Spearman's ρ at least 0.8). The rule kept 43 of Adaptyv Bio's 51 hits, including 14 of the 15 pairs that met all criteria, and added none. Its hits agreed with those from Twist Bioscience's data (κ 0.74, 95% interval 0.62 to 0.83) about as well as Adaptyv Bio's own did (0.78).

S3.9 The 17 binders in detail

Figure S8 shows the sensorgrams of the 17 binders, Table S17 describes the molecules, and Table S18 and Figure S10 give the rate constants and fits behind their K_D values. Among the 120 designs of Sets 1 to 7, the binders by design tool were: RFdiffusion3, 5 of 17; Genie 3, 5 of 27; Proteina-Complexa, 2 of 12; Mosaic, 2 of 11; PXDesign, 2 of 33; FreeBindCraft, 0 of 12; and the other four tools, 0 of 8. All three binders of Set 7 came from multistate sequence design (3 of 16 designs, compared with 0 of 14 without it; Section S1.1).

One pair that was a hit at both labs, 9a8cdc7a against Cn2, misses criterion 1 alone, because criterion 2 tests only the K_D values that were fitted. Adaptyv Bio identified it as a hit in one of three SPR replicates and fitted no K_D, which the rule for LqhIII would count as a second miss. At Twist Bioscience, its 15.7 RU at 1 µM lay just above the limit of 13.8 RU for a hit (Figure S5B). Its K_D there was about 770 nM. It was also a hit against the second supplier's Cn2 (Section S3.7), in reverse BLI at both labs (Section S3.10), and against ovalbumin (Section S4.4). Cn2-B5 comes from a backbone designed against Ts1 and was not a hit against Ts1. It had scored higher against Ts1 than against Cn2 (final scores 0.86 versus 0.59). Likewise, AaHII-B2 and AaHII-B3 share a backbone designed against AaH I and were not hits against AaH I at either lab.

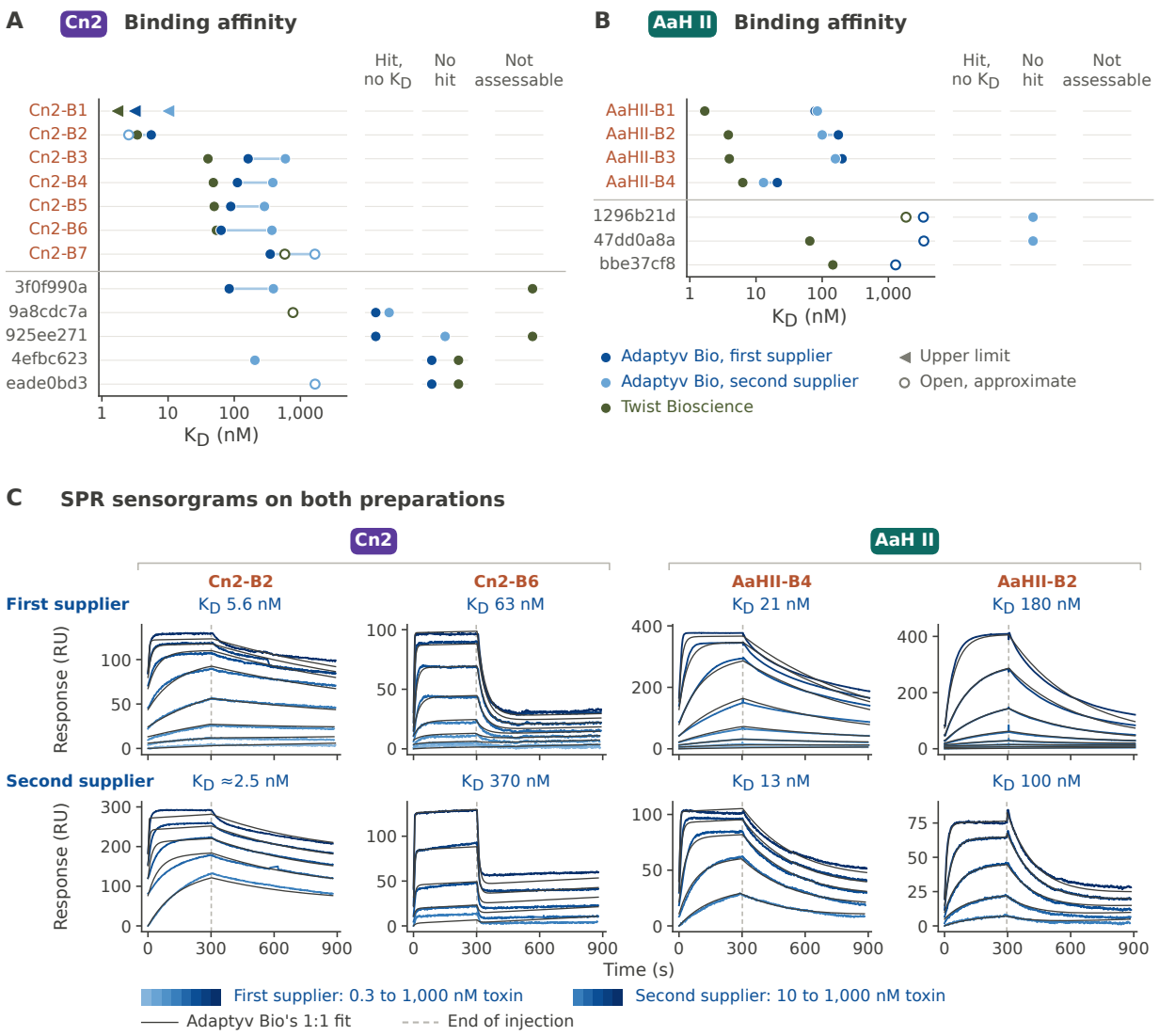


Figure S7. All 11 Cn2 and AaH II binders were hits at Adaptyv Bio against a second supplier's toxin, and 6 of 10 K_D values moved away from Twist Bioscience's. A and B) Hits against either preparation (mean of each pair's replicates, or its single replicate). Cn2-B1's upper limit against the second preparation is the lowest concentration tested (10 nM). bbe37cf8 was not measured against the second preparation. Open marks: at Adaptyv Bio, replicates more than 3-fold apart or K_D outside the concentrations tested; at Twist Bioscience, K_D above half the highest concentration tested. **C)** Highest-numbered SPR replicate that was a hit, with 1:1 fits, under the pair's K_D: per toxin, the tightest binder fitted on both and, among the rest, the largest change.

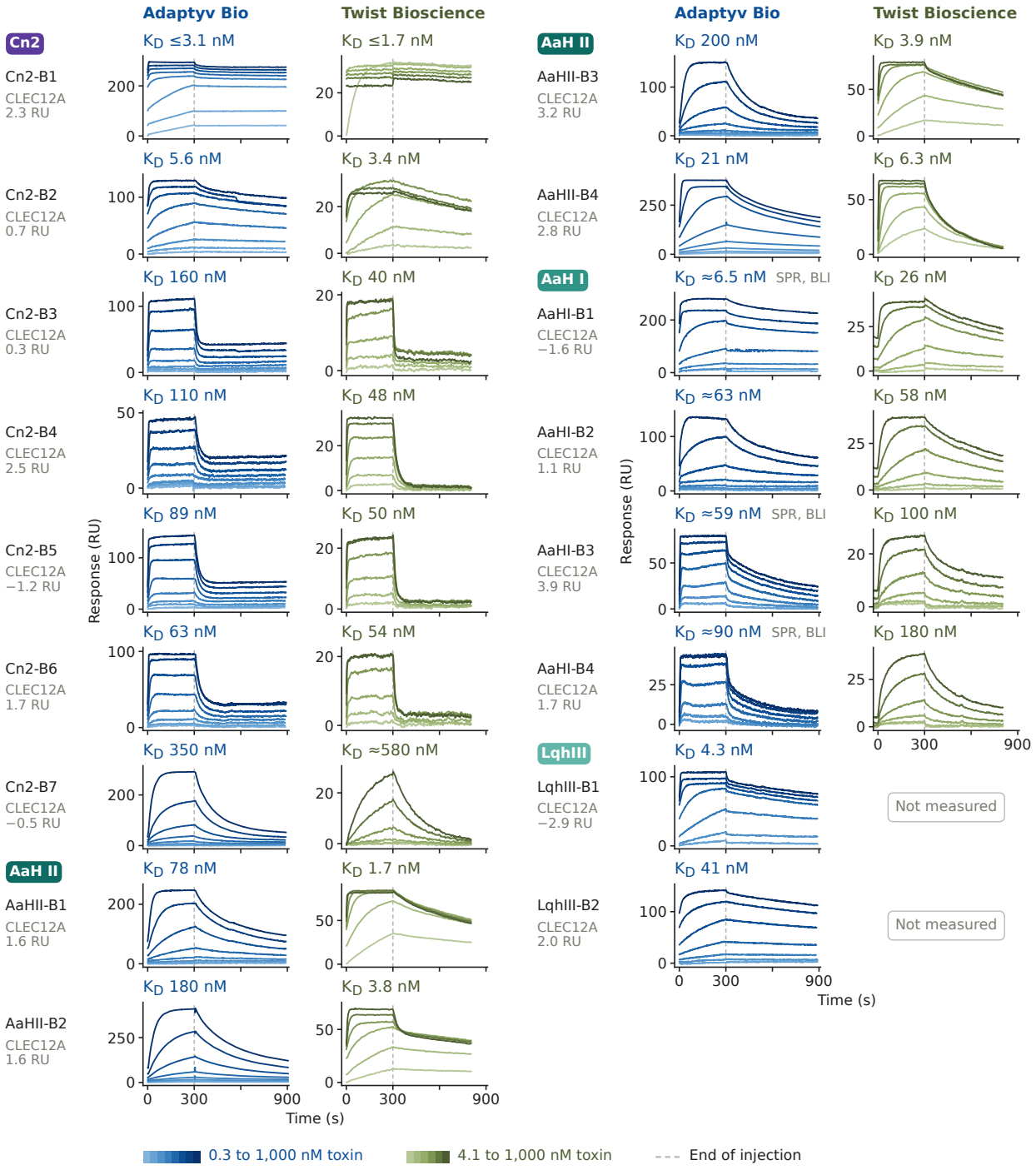


Figure S8. Sensorgrams of the 17 binders. Each row shows one binder against its toxin (colored label), in SPR at Adaptyv Bio (blue, the highest-numbered SPR replicate in the data release, except replicate 1 for Cn2-B3, AaHII-B3, AaHI-B2, and LqhIII-B1) and Twist Bioscience (green, single measurement); traces darken with rising toxin concentration (against AaH I and LqhIII, the series shown from Adaptyv Bio began at 1 nM). Headers give each lab's K_D for the pair, as in Table 1 but to two digits: ≤, upper limit; ≈, approximate; gray tag, mean of the SPR replicate shown and one BLI replicate. CLEC12A, response to 1 μM of this unrelated protein at Twist Bioscience (second of two runs); none of the binders was a hit against it. Five values lie below the lowest concentration tested (4.1 nM at Twist Bioscience).

Table S16. Pairs meeting all criteria under other definitions. The first row is the definition used throughout this report for pairs measured at both labs (Methods; Figure 3D; Table 1); the two LqhIII pairs, which only Adaptyv Bio measured, are not counted. Each later row changes or adds one element, and indented rows add a condition to the row above. 'Three changes together' combines no hit against a non-intended toxin (by Adaptyv Bio's own analysis); the fixed rule applied to Adaptyv Bio's traces; and the theoretical R_{max} at the datasheet's mass or, for AaH I, the mature mass. Standard deviations are estimated from the median absolute deviation.

Definition	Pairs added or lost	Both labs, all criteria met
As used in this report, with the limit for a hit at 5 standard deviations		15
Limit for a hit at 3, 4, 6, or 7 standard deviations	Cn2-B3 lost at 7	15, 15, 15, 14
Also no hit against a non-intended toxin	AaHII-B4 lost (hits against Ts1)	14
Also no hit against ovalbumin (Section S4.4)	Cn2-B1 and AaHI-B4 lost	13
A hit at Adaptyv Bio in any replicate, not in every one	9a8cdc7a against Cn2 added (one of three replicates)	16
AaH II at the datasheet's 20.2 kDa		15
AaH I at its mature mass, 6.80 kDa (construct undocumented)	AaHI-B1 to AaHI-B4 lost	11
Twist Bioscience's own flag (fitted R_{max} 5-fold or more from the theoretical R_{max}) in place of criterion 3	AaHII-B4 against Ts1 added	16
No limit on K_D	bbe37cf8, 47dd0a8a, and 1296b21d against AaH II added	18
The fixed rule applied to Adaptyv Bio's traces (Section S3.8)	AaHI-B4 against AaH I lost	14
A K_D value at both labs	Cn2-B1 lost (upper limits)	14
and the two within 10-fold	AaHII-B1 to AaHII-B3 lost	11
Three changes together (see caption)	Cn2 7, AaH II 3	10
and a positive control that worked at both labs	All Cn2	7

Most pairs dissociated within minutes: of the 15 measured at both labs, the median off-rate was 4.7×10^{-3} s^{-1} among the 14 resolved at Adaptyv Bio and 3.2×10^{-3} s^{-1} among the 14 resolved at Twist Bioscience (Table S18). At Adaptyv Bio, part of the signal remained at the end of dissociation, so their off-rates describe the fast part of the decay and their drift term describes the rest. In the highest-numbered SPR replicate, 14 of the 15 pairs ended 9 to 46 percentage points above the level the mean off-rate predicts, and AaHI-B1 ended 10 points below it. The median of the 15 was 19 points, compared with +5 points (−8 to +18) at Twist Bioscience. A longer dissociation phase would show whether this remainder dissociates slowly or stays bound.

Cn2-B1 dissociated from Cn2 most slowly (Figure 6B): in each of its three SPR measurements, its signal fell by at most 17% over 8 to 10 min of dissociation. Its upper limit on K_D uses no rate constant. It comes from the fraction of the top level reached at the lowest concentrations, counts every loss between injections as lost surface, and allows for drift. We set each spot's zero at the start of the first injection and took R_i as the mean response over the last 10 s of injection i. With the top level R_{top} as the highest R_i, reached at C_{top}, and $r_i = R_i/R_{top}$, 1:1 binding gives $K_D \le C_i C_{top}(1 - r_i)/(r_i C_{top} - C_i)$. This is an upper limit because, as concentrations rise on an empty surface, occupancy nears equilibrium from below. Both allowances loosen the limit. First, we counted all loss between injection i and the top level as lost surface, not released toxin: in place of R_{top}, we used R_i plus the rise within each later injection, plus that injection's length times the faster rate of fall over the last 390 s of the two adjoining dissociation phases. Second, we subtracted from R_i the highest level on a blank spot of the same run at that injection (0.2 and 8.0 RU at the lowest concentration). At Adaptyv Bio, the four injections up to 10 nM each gave a valid limit, and we took the smallest. At Twist Bioscience, the first injection (4.1 nM) gave the top level, so we used what the later injections added. With ρ_k the level at injection k over that at the first, $K_D \le C_1 C_k(\rho_k - 1)/(C_k - \rho_k C_1)$, with both allowances (highest reference spot, 3.6 RU), and we again took the smallest.

Adaptyv Bio's two replicates give 1.8 and 3.1 nM, and we report the looser of these; Twist Bioscience's limit is 1.7 nM. Three other methods gave 0.064 to 2.2 nM at Adaptyv Bio: the largest off-rate the traces allow over the fitted on-rate, a stress test with the lowest on-rate they allow, and the limit above without its allowances. The first two gave 0.05 to 0.06 nM at Twist Bioscience, so the reported limits are the loosest at each lab. The limits assume that the toxin is at its stated concentration, and they hold for the sites that fill first, because a model with one class of sites does not fit the traces.

The two-lab criterion is conservative. It leaves out 13 pairs with a K_D of 1 µM or tighter at the lab where they were hits and no hit against CLEC12A (Figure S9B); for 9 of them, CLEC12A gave background signal in both runs, as criterion 4 requires. For some, the other lab's traces rise with concentration without reaching the level of a hit, most clearly for the fusion 40090fca against AaH I, which reached 14 to 56 RU in three replicates at Adaptyv Bio with slow dissociation. Measuring these 13 pairs again at both labs would show how many bind.

A **Hits at one lab only (not labeled binders, except *)**

	Design	Format	K_D	Other lab	CLEC12A
Cn2	3f0f990a	Single	84 nM	No usable measurement	
	925ee271	Single	no K_D	No usable measurement	
	700d103b	Fusion	580 nM	No hit	
	e32555c5	Single	≈750 nM	No hit	
	8382cd8f	Single	870 nM	No hit	
	7491eb22	Fusion	490 nM	No hit	●
Ts1	173c84c2	Single	17 nM	No hit	
	6a9bcd5b	Fusion	19 nM	No hit	
	6d0b4f3f	Fusion	810 nM	Weak or ambiguous	●
	b6e5be90	Fusion	250 nM	No hit	●
AaH II	4eae42fc	Single	960 nM	No hit	
	4fcc0f88	Single	2,200 nM	No hit	
	31723df8	Single	2,700 nM	No hit	
	443207a9	Fusion	220 nM	No hit	●
	0af4b327	Fusion	240 nM	No hit	●

	Design	Format	K_D	Other lab	CLEC12A
AaH I	98052822	Fusion	92 nM	Weak or ambiguous	
	7773743e	Single	≈340 nM	No hit	
	917b6d05	Single	≈360 nM	No hit	
	4fcc0f88	Single	6.6 nM†	Not measured	
	569c6be9	Fusion	65 nM	No hit	
	40090fca	Fusion	86 nM	No hit	
LqhIII	LqhIII-B1*	Single	4.3 nM	Not measured	
	LqhIII-B2*	Single	41 nM	Not measured	
	a1f798ca	Single	no K_D	Not measured	
	443207a9	Fusion	21 nM	Not measured	●
	0af4b327	Fusion	21 nM	Not measured	●
	f8242edc	Fusion	≈76 nM	Not measured	●
	6d0b4f3f	Fusion	≈130 nM	Not measured	●
	33b2180d	Fusion	≈200 nM	Not measured	●

At Adaptyv Bio At Twist Bioscience ● Hit against CLEC12A * Labeled a binder

B **Sensorgrams of the 13 pairs with K_D of 1 µM or tighter and no hit against CLEC12A**

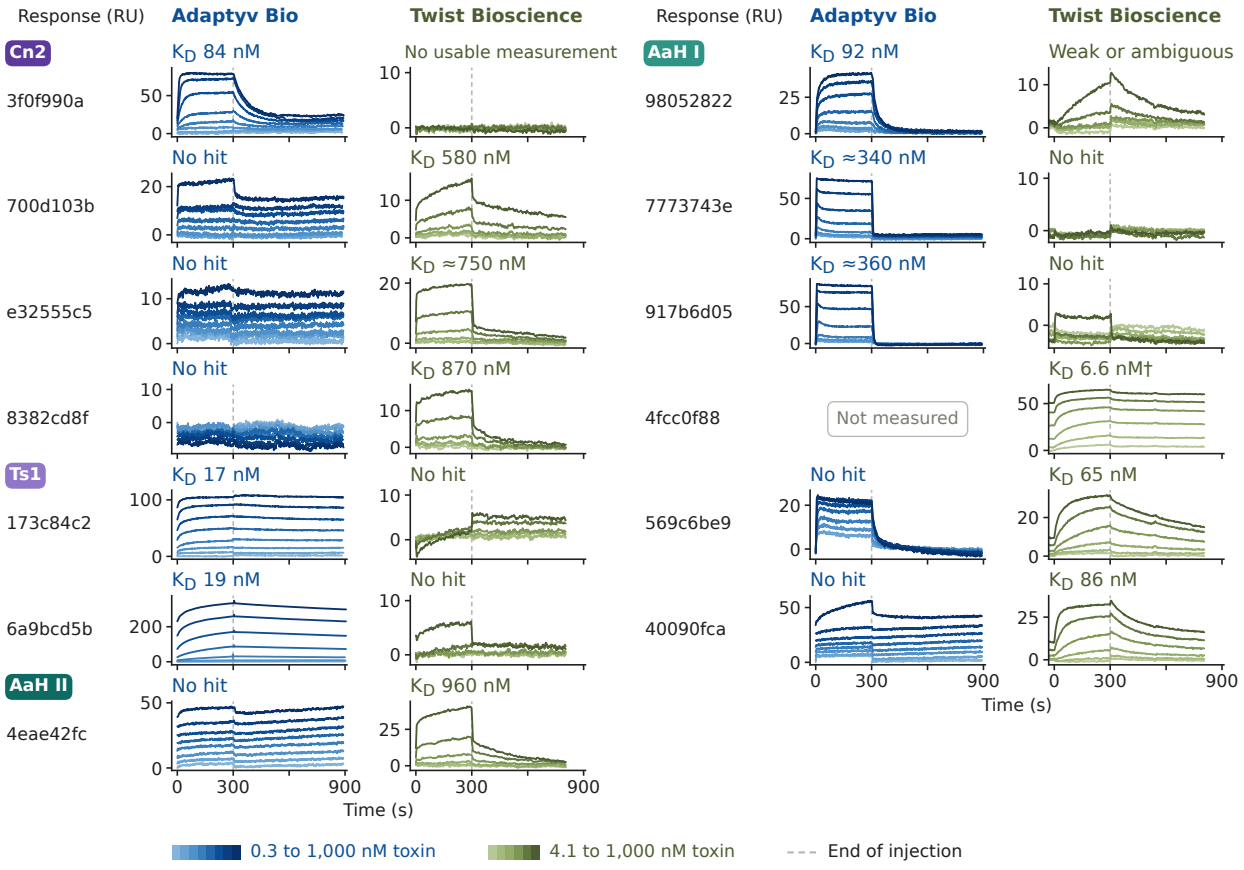


Figure S9. Of 29 design-toxin pairs that were hits at one lab only, the other team measured 18, and 8 were against LqhIII, which only Adaptyv Bio measured. These pairs are not among the 17 binders, except the two against LqhIII. **A)** All 29 pairs; K_D in the color of the lab where the pair was a hit (≈, approximate); Single, single domain; Fusion, fusion of two domains. Against AaH I, Adaptyv Bio's K_D is the mean of one SPR and one BLI replicate. **B)** SPR at both labs, drawn as in Figure S8, for the 13 pairs, against the four toxins that both teams measured, with a K_D of 1 µM or tighter at the lab where each was a hit and no hit against CLEC12A (3f0f990a, like 925ee271 in A, could not be assessed against it: under 50 RU captured, 'No usable measurement'); headers give the K_D or, without a hit, the result. ∗Labeled a binder (sensorgrams, Figure S8). †Made against AaH II, not AaH I, with irregular traces.

Table S17. The 17 binders are helical proteins that contact three to five hotspots in the predicted complex. Table 1 gives each design's intended toxins, its design tool, and its length. Fold, buried area, and contacts (Section S3.11) come from the ESMFold2 complex with the toxin it binds. Helices have eight or more residues; in a bundle, all helix axes lie within 40° of one another; buried area sums both sides (probe 1.4 Å); contact means heavy atoms within 5 Å. In AaHI-B3, only the second domain, 81% identical to AaHI-B4, contacts AaH I. pLDDT, mean pLDDT of the design predicted as a monomer by ESMFold2-Fast and ESMFold2 (mean of five seeds). Net charge at pH 7.4, from the sequence (Biopython, Bjellqvist pK_a set). BVP ELISA, Fc fusion at Twist Bioscience, mean of two plates, % of cixutumumab on the same plates; the binders of AaH I and LqhIII were read in the second run (Section S4.2). LR competition, share of the signal against Cn2 lost with LR bound first, as in Figure 6D; low signal, too weak to assess. n/a, not run. [a]Same starting backbone. [b]One domain of AaHI-B3 shares the starting backbone of AaHI-B4. [c]Fusion of two campaign domains, whose backbones came from PXDesign and Genie 3; no design tool was run on the fusion itself.

Design	UUID	Set	Fold	Buried area (Å^2)	Toxin residues in contact	Hotspots in contact (of 5)	pLDDT, monomer (two predictors)	Net charge, pH 7.4	BVP ELISA (%)	LR competition (%)
Cn2 binders										
Cn2-B1	888c8f04	1	four-helix bundle	2,129	23	5	85, 82	−0.6	28.8	94
Cn2-B2	66c2cc29	1	four helices	2,739	29	3	76, 79	−11.7	97.7	87
Cn2-B3	8ac202c8	1	four helices	2,353	24	5	83, 78	−6.8	108.1	95
Cn2-B4	02090d8c	1	six helices	2,606	27	5	79, 81	−10.8	28.5	93
Cn2-B5	9d008884	6	five helices	2,048	20	3	81, 78	−9.8	80.8	93
Cn2-B6	b6d46ba0	6	four helices	1,923	21	5	85, 86	−6.8	68.2	94
Cn2-B7	5f5b71ad	1	four helices	2,386	25	5	77, 75	−7.7	45.1	low signal
AaH II binders										
AaHII-B1	d2a254ba	3	five helices	2,113	28	5	86, 88	−7.4	47.8	n/a
AaHII-B2[a]	8f156315	7	four helices	1,665	23	4	77, 76	−5.8	13.3	n/a
AaHII-B3[a]	96ee6417	7	four helices	1,731	23	5	76, 74	−5.8	18.7	n/a
AaHII-B4	a199ee01	3	five helices	1,795	22	3	89, 91	−11.7	51.9	n/a
AaH I binders										
AaHI-B1	30fc4a66	4	four-helix bundle	1,788	27	4	81, 83	−3.7	71.6	n/a
AaHI-B2	03131527	4	four helices	1,656	20	5	86, 81	−10.7	42.5	n/a
AaHI-B3[b]	8ebe5db2	9	two domains[c]	2,440	25	5	85, 92	−17.7	94.8	n/a
AaHI-B4[b]	9cad7527	7	four helices	2,271	27	5	83, 82	−11.5	81.7	n/a
LqhIII binders										
LqhIII-B1	dceea244	5	seven helices	1,718	21	5	89, 89	−7.6	45.2	n/a
LqhIII-B2	dbb87a88	5	three helices	1,917	24	5	74, 72	−8.8	33.6	n/a

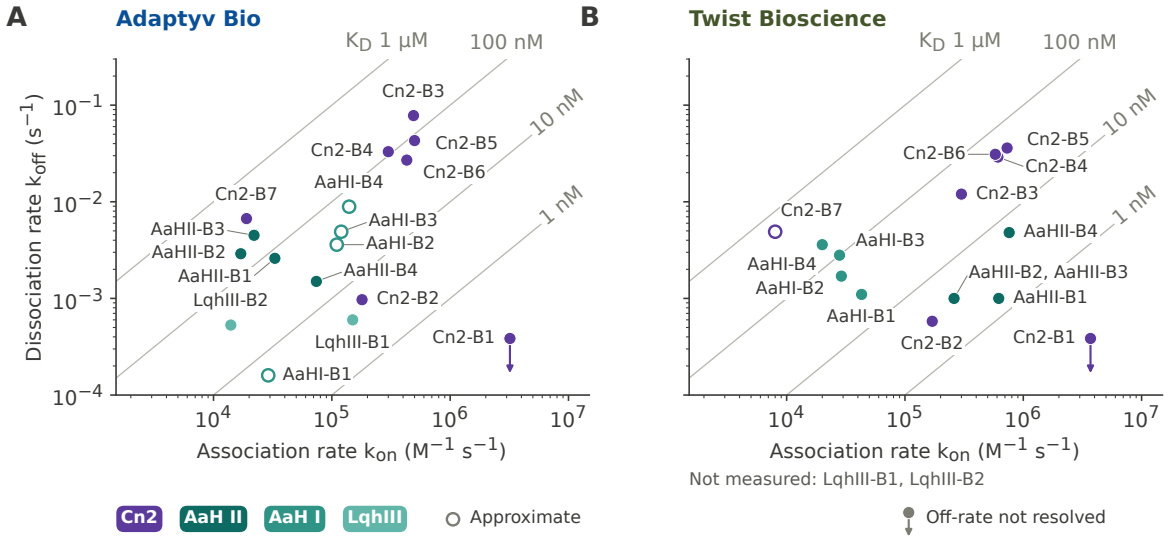


Figure S10. Cn2-B1 associated fastest at both labs and was among the slowest to dissociate. Rates from Table S18: **A)** Adaptyv Bio, **B)** Twist Bioscience; 17 binders fitted at Adaptyv Bio and 15 at Twist Bioscience. Diagonals: equal $K_D = k_{off}/k_{on}$. Open marks: approximate. At Adaptyv Bio, each rate is the geometric mean of the fitted replicates, one of them by BLI for AaHI-B1, AaHI-B3, and AaHI-B4. The replicates of the four AaH I binders differ more than 3-fold in K_D, so the ratio of their mean rates differs from their mean K_D by up to 1.9-fold. Arrows: off-rate not resolved; the mark stands at an upper limit of 3.9×10^{-4} s^{-1}, and Table 1 gives the looser upper limits of K_D that we report (3.1 and 1.7 nM). The mean off-rate of AaHI-B1 at Adaptyv Bio, 1.6×10^{-4} s^{-1}, lies below that limit.

Table S18. Most binders dissociated within minutes; the off-rate of Cn2-B1 was mostly not resolved. One row per pair, in SPR, except that one of the two fitted replicates of AaHI-B1, AaHI-B3, and AaHI-B4 at Adaptyv Bio was run by BLI. K_D at Twist Bioscience as in Table 1; at Adaptyv Bio, mean ± standard deviation of the fitted replicates (two; three for Cn2-B3, Cn2-B5, and the LqhIII binders; four for AaHII-B4), as in Table 1. Rates, Adaptyv Bio's 1:1 fit (geometric mean of replicates, so their ratio can differ from the mean K_D) and our 1:1 fit of Twist Bioscience's data: k_{on} in 10^5 M^{-1} s^{-1}, k_{off} in 10^{-3} s^{-1} (at Adaptyv Bio, fast decay only). Left, signal over the last 20 s of dissociation as a share of the signal at the end of injection (1 μM; at Adaptyv Bio, highest-numbered SPR replicate). R_{max} ratio, fitted over theoretical R_{max}. ≤, upper limit; ≈, approximate; †, below the lowest concentration tested.

	Adaptyv Bio				Twist Bioscience					
Design	K_D (nM)	k_{on}	k_{off}	Left (%)	K_D (nM)	k_{on}	k_{off}	Left (%)	R_{max} ratio	Note
Cn2 binders										
Cn2-B1	≤3.1	32	≤0.39	94	≤1.7†	37	≤0.39	94	0.62	*a*
Cn2-B2	5.6 ± 1.8	1.8	0.97	76	3.4†	1.7	0.58	71	0.57	*b*
Cn2-B3	162 ± 45	4.9	78	36	40	3.0	12	12	0.39	*b*, *c*, *d*
Cn2-B4	112 ± 13	3.0	33	46	48	6.1	29	4	0.72	*b*, *c*
Cn2-B5	89 ± 28	5.0	43	37	50	7.3	36	10	0.44	*b*, *c*
Cn2-B6	63 ± 17	4.3	27	34	54	5.8	31	11	0.41	*b*, *c*
Cn2-B7	350 ± 76	0.19	6.7	18	≈580	0.08	4.9	7	0.82	*b*, *e*
AaH II binders; the labs' K_D values differ 3- to 51-fold										
AaHII-B1	78 ± 1	0.33	2.6	39	1.7†	6.2	1.0	58	0.56	*b*
AaHII-B2	176 ± 9	0.17	2.9	30	3.8†	2.6	1.0	53	0.46	
AaHII-B3	200 ± 8	0.22	4.5	29	3.9†	2.6	1.0	56	0.57	*b*
AaHII-B4	21 ± 3	0.74	1.5	50	6.3	7.6	4.8	8	0.54	
AaH I binders; the labs' K_D values differ 1.1- to 4.1-fold										
AaHI-B1	≈6.5 ± 4.6†	0.29	0.16	81	26	0.43	1.1	61	0.62	*f*
AaHI-B2	≈63 ± 76	1.1	3.6	30	58	0.29	1.7	47	0.79	*b*, *f*
AaHI-B3	≈59 ± 59	1.2	4.9	31	100	0.28	2.8	42	0.59	*b*, *f*
AaHI-B4	≈90 ± 87	1.4	8.9	19	180	0.20	3.6	26	0.57	*b*, *f*
LqhIII binders; measured at Adaptyv Bio only										
LqhIII-B1	4.3 ± 1.8	1.5	0.60	85			not measured			*b*, *g*
LqhIII-B2	41 ± 19	0.14	0.53	79			not measured			*g*

† The lowest concentration tested was 4.1 nM at Twist Bioscience, so five K_D values lie below it. At Adaptyv Bio, the BLI replicate of AaHI-B1 (3.2 nM) lies below the lowest concentration of its series (31.6 nM). *a* Upper limits (Section S3.9); the off-rate was resolved in one of three measurements, so the upper limit drawn in Figure S10 is given (fitted: 0.28 at Adaptyv Bio; 0.01, the slowest our fit allows, at Twist Bioscience); Left at Twist Bioscience is measured from 10 s after the end of injection (107% if measured from its end). *b* The fitted off-rate overpredicts a replicate's loss at 1 μM by 15 percentage points or more (at Adaptyv Bio; for AaHI-B3, at both labs). *c* Equilibrium value, to about 2-fold (Adaptyv Bio; for Cn2-B3, both labs). *d* Twist Bioscience's 1:1 off-rate for Cn2-B3 is about 14 times slower than the fast phase of a two-phase fit to the single trace at 1 μM. *e* K_D near the highest concentration tested. *f* Adaptyv Bio's two fitted replicates differ 3.0-fold for AaHI-B1 (SPR 9.7 and BLI 3.2 nM), 13-fold for AaHI-B2 (SPR 117 and SPR 9.0 nM), 5.8-fold for AaHI-B3 (BLI 101 and SPR 18 nM), and 5.4-fold for AaHI-B4 (BLI 151 and SPR 28 nM). *g* Measured at one lab. The 4.3 nM of LqhIII-B1 is the mean of three replicates (2.3, 5.3, and 5.4 nM); the rates of LqhIII-B2 are similar to those of the signals given against Ts1 by Fabs raised against other toxins.

S3.10 Reverse BLI

In reverse BLI, the design is free in solution, so the result does not depend on how the design was immobilized. Adaptyv Bio measured 49 designs this way in 69 design-toxin pairs (Tables S19 and S20; Figures S11, S12, S13, and S14). All 17 binders were hits against their own toxin. Against Cn2, 12 of the 13 designs were hits, including 05efdf6e and ca06e722, which were hits at neither lab in the main screen. The exception, e32555c5, gave a weak signal that Adaptyv Bio's team does not count as binding (0.30 nm at 1 μM; the three weakest hits, 700d103b, 05efdf6e, and 9a8cdc7a, gave 0.13 to 0.34 nm). Against AaH I, 19 of 21 were hits: the four binders and 15 other designs, 11 of them hits at both labs in the main screen. Against these two toxins, the assay thus did not separate binders from the other designs. Against AaH II, 5 of 15 were hits: the four binders and bbe37cf8, whose weak signal was not fitted. The other ten, Cn2-B1 among them, gave shifts of −0.06 to 0.04 nm at up to 1 μM. Seven of them had been hits against AaH II at one lab or both in the main screen. Here the assay agreed with our labels for 14 of the 15 designs, so the toxin on the sensor did not bind designs indiscriminately. Against LqhIII, 3 of 10 were hits (both binders and 443207a9), and four were unresolved. The sensors carried less LqhIII the higher the design concentration (1.07 to 1.28 nm at 31.6 nM but 0.05 to 0.29 nm at 1 μM, compared with 1.4 to 1.8 nm of AaH I throughout). At 1 μM, all ten designs gave 0.12 to 0.37 nm, so hits and other designs differed little in signal there. Against Ts1, all ten designs were unresolved. All 22 pairs with a result in two runs gave the same result in both.

The K_D values are Adaptyv Bio's 1:1 fits (Section S3.2). Cn2-B1 gave the tightest K_D against Cn2, about 6.3 nM, and lost 3 to 8% of its signal over the 300 s wash. Ten binders had two fitted replicates, whose values differed 1.1- to 2.0-fold. For Cn2-B2 to Cn2-B6, the K_D was 6 to 15 times higher than by SPR at Adaptyv Bio and 12 to 36 times higher than at Twist Bioscience. For Cn2-B7, it was 2 to 4 times lower. For AaHII-B1 to AaHII-B3, it lay between the two labs' SPR values, and for AaHII-B4 above both. For the AaH I

binders, it lay within 3.2-fold of Adaptyv Bio's main-screen value and 1.9 to 5.5 times below Twist Bioscience's. Several points call for caution. Four values against Cn2 (Cn2-B3, Cn2-B5, Cn2-B6, and 05efdf6e) lie near or above the highest concentration tested, 1 μM. The values of eight binders lie below the lowest, 31.6 nM. Against AaH II, the binders' largest shifts at the end of association were only 0.13 to 0.25 nm, while blank sensors drifted by up to 0.06 nm, and Adaptyv Bio graded the five fits as Poor or Medium. Against AaH I, they selected another model for 15 of the 19 fitted pairs. We therefore read reverse BLI as support for binding, not as a measure of affinity.

Reverse BLI at Twist Bioscience. The Twist Bioscience team ran the same orientation, with the His-tagged toxin captured on the sensor and the design, a bivalent Fc fusion, in solution at 0.55 to 400 nM (Section S1.2). The runs covered 48 designs in 69 pairs, with 16 against Cn2, 14 against Ts1, 17 against AaH II, and 22 against AaH I (Figure S15). We use them for hit or no hit only, by Twist Bioscience's rule of a response of at least 0.05 nm, and report no K_D from them. Of the 15 binders tested against their own toxin, 12 were hits: 6 of 7 against Cn2, all 4 against AaH II, and 2 of 4 against AaH I. Cn2-B7 gave no signal, and AaHI-B3 (0.028 nm) and AaHI-B4 (0.048 nm) stayed under Twist Bioscience's limit. All three were hits at Adaptyv Bio, so all 15 were hits at one lab or both. Among the controls, LR and RU1 bound Cn2 and 9C2 bound AaH I, whereas 9004G, 4C1, and serrumab did not bind their toxins. The two labs gave the same result for 14 of the 15 designs that both teams ran against AaH II, and the exception, 47dd0a8a, was a hit at Twist Bioscience only. Against Cn2 they agreed on 9 of the 13: Cn2-B7, 05efdf6e, and 3f0f990a were hits at Adaptyv Bio only, and e32555c5 was a hit at Twist Bioscience only. Against AaH I they agreed on 10 of the 19 designs with a result at both labs, and the other nine were hits at Adaptyv Bio only. The sample of 3f0f990a at Twist Bioscience held no protein.

Against Ts1, all 14 designs were hits by Twist Bioscience's rule, including six that were not hits against Ts1 in SPR at either lab, whereas serrumab gave no signal. In 12 of the 14, the signal appeared only at the highest concentrations, and 3 to 50% of it was lost in the 600 s wash, read at the second-highest concentration. Three of these series began at 56 to 115 nM instead of 400 nM and still gave 0.08 to 0.11 nm at their highest concentration, against 0.10 to 0.15 nm for the other nine. The signal thus followed the dilution step rather than the molar concentration. In e32555c5, it did not fall in order with dilution. AaHII-B4 stood out only in degree: it gave the largest signal at each of the three highest concentrations (0.17 nm at 133 nM, against 0.05 to 0.11 nm for the other 13 at their second-highest concentration), and 74% of its signal was lost in the wash. Without a working control, we do not read this run as evidence of binding against Ts1.

Ts1 aside, 28 design-toxin pairs that we do not label binders were also hits with the toxin on the sensor at one lab or both. They comprise 16 against AaH I, 9 against Cn2, 2 against AaH II, and 1 against LqhIII. Of the 16, 12 were hits at both labs in the main screen, and 11 also against CLEC12A. Seven of the nine against Cn2 were the designs with which LR competed, with 58 to 88% of the signal lost (Section S4.1). All but 9a8cdc7a had been hits at one lab at most in the main screen. Of the 25 designs behind these pairs, 15 were also hits against CLEC12A or ovalbumin (Section S4.4). The label is thus a definition, and some of these pairs may bind weakly.

S3.11 Predicted interfaces and binding

We tested whether predicted interfaces separate binders from the other designs, using the ESMFold2 model of each design with each intended toxin (665 pairs). We fixed six comparisons before running them, with a threshold of $p = 0.008$: 1) toxin residues in contact, 2) buried area, 3) apolar share of the interface, 4) hotspots in contact, 5) design length, and 6) design tool. All 60 single-toxin designs had passed the same score thresholds, with ESMFold2 among the three predictors that selected them. Table S5 counts contacts differently (Cβ atoms within 8 Å in the design models), so the two are not comparable. These analyses rest on predicted structures and on a small number of binders (11 and 6), and contacts and buried area are strongly correlated (Spearman's $\rho = 0.86$ within toxins).

Among the single-toxin designs, binders had larger predicted interfaces (60 designs, only 11 binders; Figure S16). Medians were 24 versus 19 toxin residues in contact (rank-biserial $r = 0.78$, pooled within toxins; permutation $p < 0.001$) and 2,113 versus 1,666 Å^2 of buried area ($r = 0.71$, $p < 0.001$). Protenix v2 models show both differences ($r = 0.83$ and 0.62). Without the Cn2 designs, however, neither difference passed the threshold (contacts, $r = 0.65$, $p = 0.013$; buried area, $r = 0.57$, $p = 0.038$). The other four comparisons did not separate the groups. Apolar share was lower in binders with both predictors, but the difference passed the threshold only with Protenix v2, and hotspots in contact, length, and design tool showed no difference.

Smaller predicted interfaces do not explain the lack of Ts1 binders. Of each toxin's 12 designs, the binders and mean buried area were: Cn2, 5 and 1,999 Å^2; Ts1, 0 and 1,890 Å^2; AaH II, 2 and 1,792 Å^2; AaH I, 2 and 1,624 Å^2; and LqhIII, 2 and 1,492 Å^2. No site property tracked these yields, whether net charge, site area and apolar share, aromatic hotspots, or mean buried area ($n = 5$). In the predicted complexes, multi-toxin designs contact their intended toxins through overlapping surfaces whose toxin residues mostly differ, yet bury about as much area per toxin as single-toxin designs, or more (Table S21). Among the six binders designed against multiple toxins, the toxin they bound ranked first by final score in three and by buried area in one, compared with 2.2 expected by chance.

Table S19. In reverse BLI at Adaptyv Bio, 12 of the 13 designs measured against Cn2 were hits, as were 5 of the 15 measured against AaH II; the main screen's SPR results are shown alongside. Rows are grouped by the toxin on the sensor. Design, name for the 17 binders, UUID for the others. Strength and fit grade (Good, Medium, or Poor) are Adaptyv Bio's; two strengths or grades, two replicates; $n = 2$, the same result in two replicates. K_D, Adaptyv Bio's 1:1 fit: for one fitted replicate, with its interval; for two, mean ± standard deviation; not graded, a 1:1 fit that they did not select (Section S3.2). Main screen, each lab's SPR K_D against the same toxin (Twist Bioscience, our 1:1 fit). ≤, upper limit; ≈, approximate (in reverse BLI, below the lowest concentration tested, 31.6 nM, not graded, or, for 9a8cdc7a, not considered reliable by Adaptyv Bio's team).

	Reverse BLI			Main screen (SPR), K_D (nM)	
Design	Strength	K_D (nM)	Fit	Adaptyv Bio	Twist Bioscience
Cn2 on the sensor					
Cn2-B1	strong	≈6.3 ± 2.0	Medium, Good	≤3.1	≤1.7
Cn2-B2	strong	40 ± 4	Good, Good	5.6	3.4
Cn2-B3	weak	1,460 ± 90	Medium, Medium	160	40
Cn2-B4	medium	642 (632 to 653)	Good	110	48
Cn2-B5	medium, weak	1,040 ± 360	Medium, Good	89	50
Cn2-B6	weak, medium	977 ± 474	Medium, Medium	63	54
Cn2-B7	medium	145 ± 41	Good, Good	350	≈580
05efdf6e	weak	2,120 (1,880 to 2,400)	Medium	no hit	2,700, weak or ambiguous
3f0f990a	medium	570 (536 to 607)	Poor	84	not assessable
700d103b	weak	no fit	no fit	no hit	580
9a8cdc7a	medium	≈125 ± 80	Poor, Medium	hit, no fit	≈770
ca06e722	medium	93 (90 to 96)	Medium	no hit	no hit
e32555c5	no hit			no hit	≈750
AaH II on the sensor					
AaHII-B1	strong	≈8.3 (7.8 to 8.9)	Poor	78	1.7
AaHII-B2	strong	≈13 (12 to 14)	Poor	180	3.8
AaHII-B3	medium	103 (99 to 106)	Medium, no fit	200	3.9
AaHII-B4	strong, medium	60 ± 22	Poor, Medium	21	6.3
Cn2-B1	no hit ($n = 2$)			no hit	no hit
0510b42f	no hit			no hit	24, weak or ambiguous
0af4b327	no hit			no hit	240
1296b21d	no hit ($n = 2$)			3,400	1,800
31723df8	no hit			no hit	2,700
443207a9	no hit			no hit	220
47dd0a8a	no hit			≈3,400	65
4eae42fc	no hit ($n = 2$)			no hit	960
4fcc0f88	no hit ($n = 2$)			no hit	2,200
6a9bcd5b	no hit			no hit	no hit
bbe37cf8	weak ($n = 2$)	no fit	no fit	1,300	150

Table S20. In reverse BLI at Adaptyv Bio, 19 of the 21 designs measured against AaH I were hits, as were 3 of the 10 measured against LqhIII. Columns and signs as in Table S19. Unresolved, Adaptyv Bio gave no result. <32, a fit below 1 nM, given as the lowest concentration tested. Main screen: at Adaptyv Bio against AaH I, the mean of SPR and BLI replicates (≈, replicates more than 3-fold apart, a 1:1 fit that they did not select, or a fit below the lowest concentration of its series). All 10 designs measured against Ts1 were unresolved: Cn2-B1, Cn2-B5, AaHII-B4, 0510b42f, 0af4b327, 173c84c2, 443207a9, 6a9bcd5b, 6d0b4f3f, and e32555c5.

	Reverse BLI			Main screen, K_D (nM)	
Design	Strength	K_D (nM)	Fit	Adaptyv Bio	Twist Bioscience
AaH I on the sensor					
AaHI-B1	strong	≈4.8 ± 1.5	Good, Good	≈6.5	26
AaHI-B2	strong, medium	≈25 ± 10	not graded	≈63	58
AaHI-B3	medium	≈19 (18 to 20)	not graded	≈59	100
AaHI-B4	medium	≈97 ± 45	not graded	≈90	180
0af4b327	strong	≈15 (15 to 16)	not graded	≈1.7	13
0c86c52b	weak	3,850 (3,180 to 4,840)	Medium	no hit	not assessable
33b2180d	medium	≈52 (51 to 54)	not graded	120	32
39f80384	strong	≈16 (16 to 17)	not graded	150	20
40090fca	strong	≈17 ± 1	not graded	no hit	86
443207a9	strong	≈15 ± 7	not graded	≈23	12
4fcc0f88	no hit (n = 2)			not measured	6.6
6d0b4f3f	strong	≈17 (17 to 18)	not graded	≈68	23
7491eb22	strong	≈16 (15 to 17)	not graded	≈340	39
7773743e	unresolved			≈340	no hit
917b6d05	medium	648 ± 19	Medium, Medium	≈360	no hit
98052822	strong	≈17 (16 to 18)	not graded	92	1,400, weak or ambiguous
b6e5be90	medium	≈92 (91 to 94)	not graded	69	21
cac1ca4c	medium	≈123 (118 to 128)	not graded	≈250	66
cf8652bc	medium	161 ± 57	Medium, Medium	≈260	130
e80c46be	strong	≈15 (13 to 16)	not graded	51	41
f8242edc	medium	≈49 (48 to 51)	not graded	≈100	17
LqhIII on the sensor					
LqhIII-B1	strong	<32	Good	4.3	not measured
LqhIII-B2	strong	≈5.9 (5.8 to 5.9)	Good	41	not measured
AaHI-B3	unresolved			no hit	not measured
33b2180d	unresolved			≈200	not measured
39f80384	no hit			no hit	not measured
443207a9	hit	no fit	no fit	21	not measured
6d0b4f3f	unresolved			≈130	not measured
7773743e	unresolved			no hit	not measured
a1f798ca	no hit			hit, no fit	not measured
f8242edc	no hit			≈76	not measured

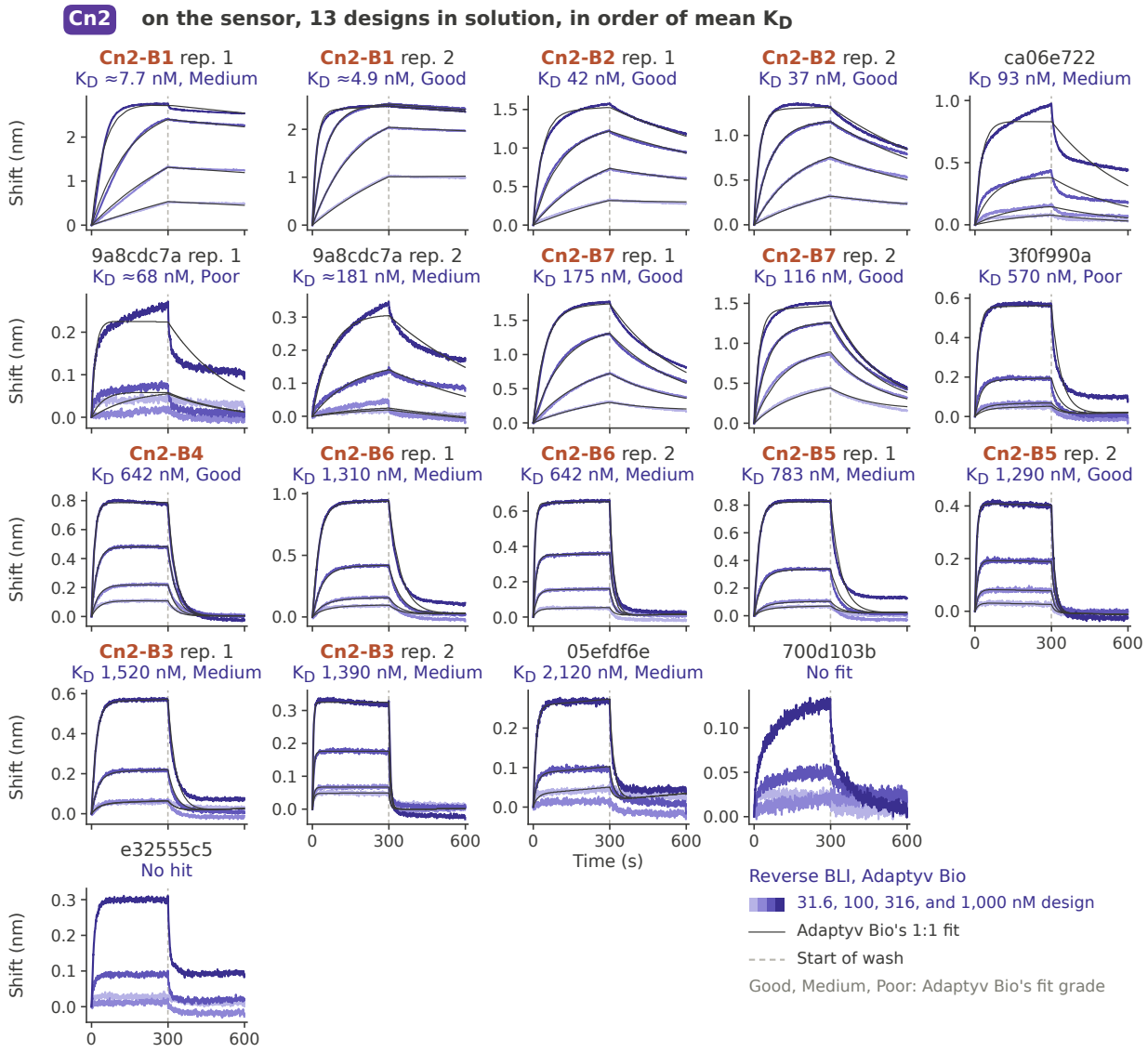


Figure S11. With Cn2 on the sensor, 12 of the 13 designs were hits, and Cn2-B1 gave the tightest K_D. Reverse BLI at Adaptyv Bio with their 1:1 fits; design in solution at 31.6 to 1,000 nM. Designs are in order of mean K_D, with both replicates of the seven measured twice. Headers give Adaptyv Bio's K_D for that replicate (≈, approximate) and the grade of its fit; no fit, a hit that Adaptyv Bio did not fit; unresolved, Adaptyv Bio gave no result; names in color, binders; UUIDs, other designs. Each plot has its own scale. Two fits (replicate 1 of 9a8cdc7a and 05efdf6e) omit the 100 nM trace. Values in Table S19.

Table S21. In the predicted complexes, multi-toxin designs contact their toxins through overlapping surfaces at which few toxin positions are identical. Medians over the designs of each group, from the ESMFold2 complexes. Toxin positions are matched by pairwise sequence alignment, not by the alignment of Figure 1B. Shared design residues, the share of a design's contact residues that touch every intended toxin. RMSD, the largest Cα RMSD between two intended toxins after the design chains are superposed. Buried area, the group's mean area per pair relative to that of the single-toxin designs against the same toxin, averaged over the group's toxins (in parentheses, the range over toxins). Blank, not computed, since a fusion's two domains need not share one surface.

Design group	Designs	Shared design residues (%)	Toxin positions in contact in all	Identical (without cysteines)	RMSD (Å)	Buried area (%)
Double β-toxin	30	70	14	6 (5)	4.9	−2 (−9 to +5)
Triple α-toxin	30	64	11.5	3 (1)	7.5	−2 (−15 to +13)
All five, single domain	59	41	7	2 (0)	8.7	+32 (+7 to +46)
All five, fusion	32					+12 (−1 to +31)

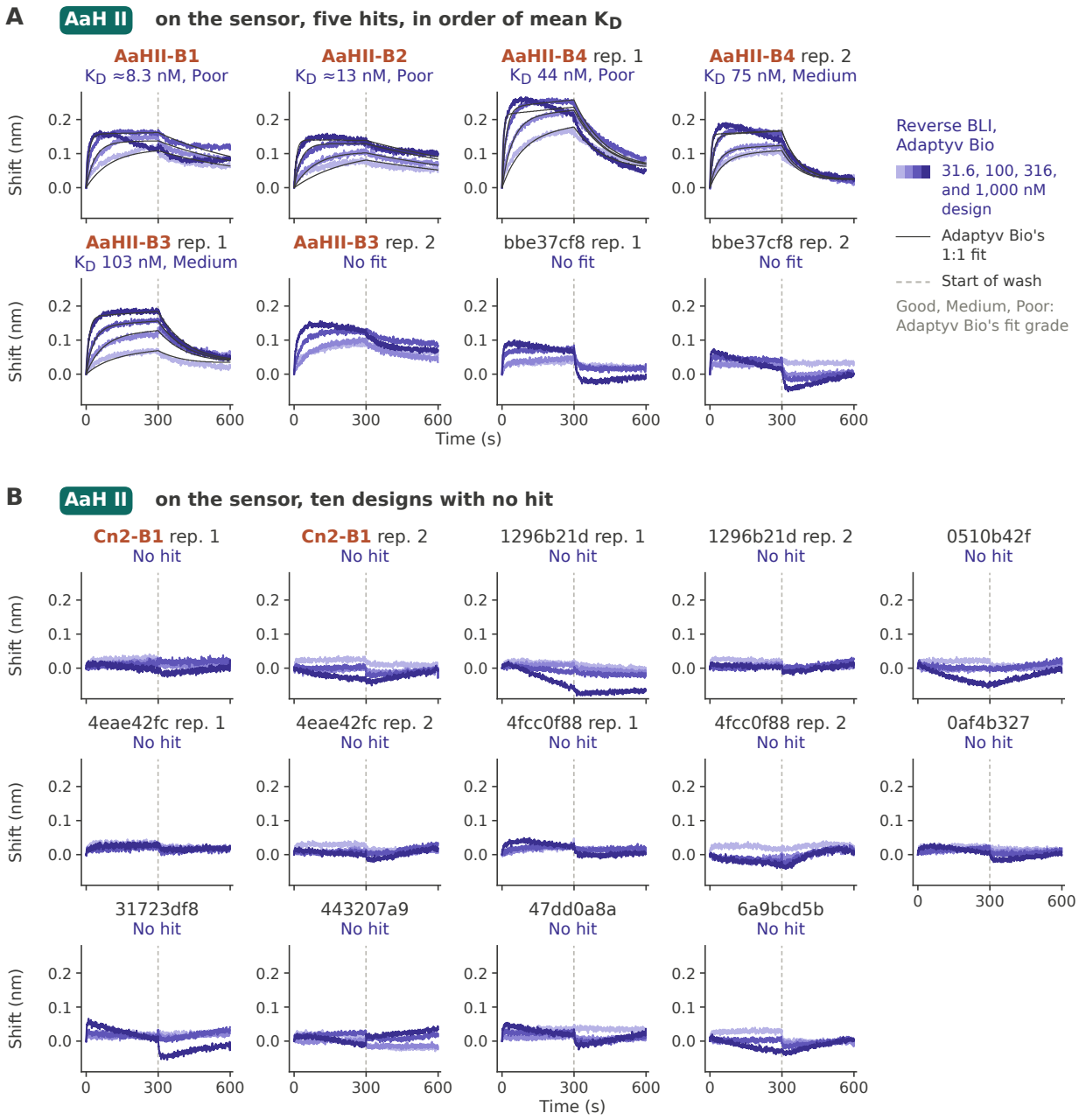


Figure S12. With AaH II on the sensor, all four AaH II binders were hits, and 10 of the other 11 designs were not. Reverse BLI at Adaptyv Bio with their 1:1 fits; design in solution at 31.6 to 1,000 nM; both replicates of the seven designs measured twice. **A)** The five hits, in order of mean K_D. **B)** The ten designs that were not hits, among them Cn2-B1. Headers as in Figure S11. All plots share one scale. The fit of AaHII-B1 omits the 1,000 nM trace. Values in Table S19.

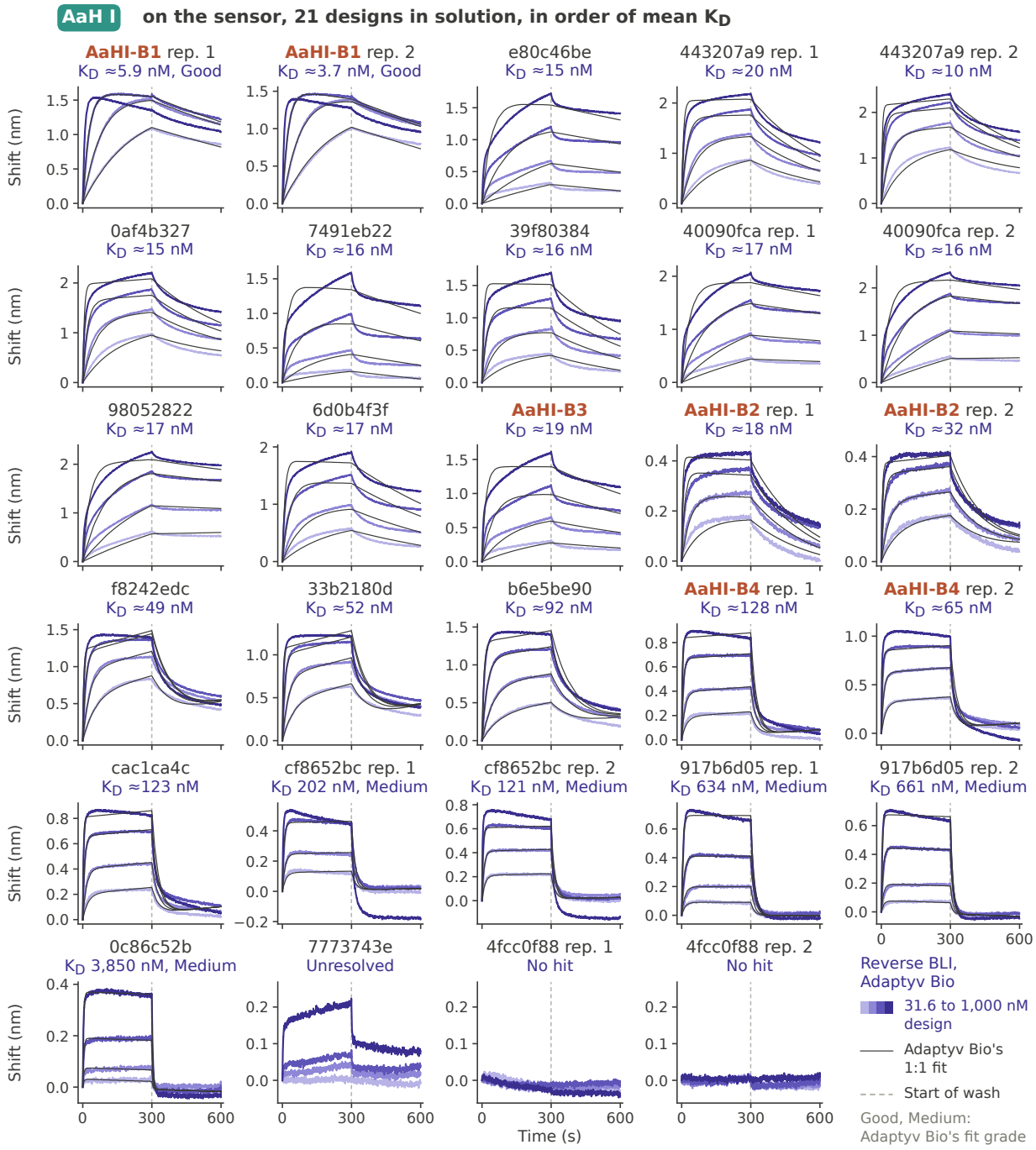


Figure S13. With AaH I on the sensor, 19 of the 21 designs were hits, including all four AaH I binders. Reverse BLI at Adaptyv Bio with their 1:1 fits; design at 31.6 to 1,000 nM. Designs are in order of mean K_D, with both replicates of the eight measured twice; 7773743e was unresolved, and 4fcc0f88 was not a hit. Headers as in Figure S11; ≈ also marks the 19 of 26 fits for which Adaptyv Bio selected a conformational-change model (no grade). Each plot has its own scale. Five fits (both replicates of AaHI-B1, replicate 2 of AaHI-B4, and both replicates of cf8652bc) omit the 1,000 nM trace. Values in Table S20.

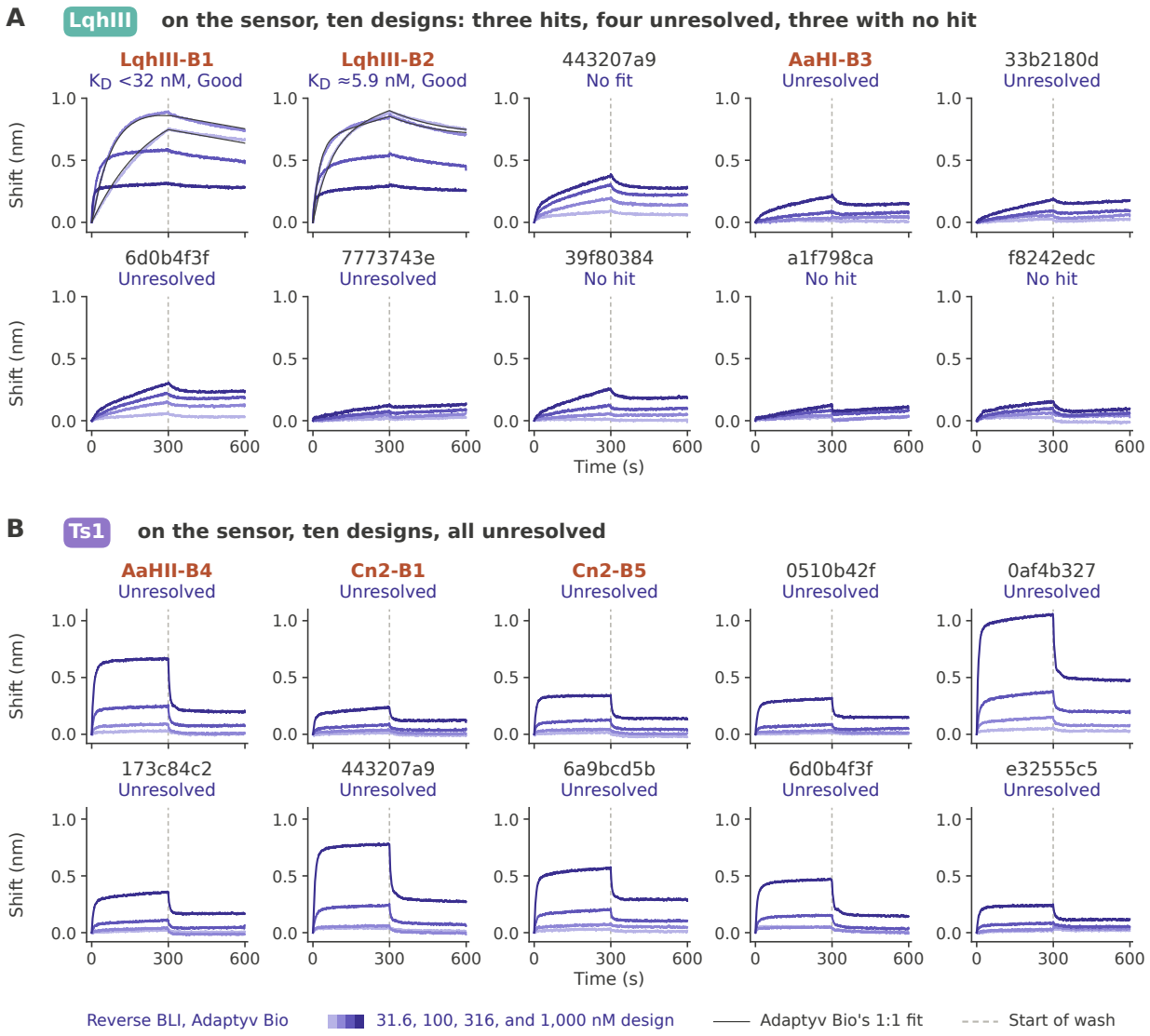


Figure S14. With LqhIII on the sensor, both LqhIII binders were hits; with Ts1, all ten designs were unresolved. Reverse BLI at Adaptyv Bio, 1:1 fits; headers as in Figure S11; one scale per panel. **A)** One replicate each. Sensors carried less LqhIII at higher design concentrations (1.07 to 1.28 nm at 31.6 nM, 0.05 to 0.29 nm at 1,000 nM), so signals are not comparable across them; both fits use only the two lowest. The off-rate of LqhIII-B1 lay at the fit's lower bound (10^{-5} s^{-1}), so its header gives the lowest concentration as a limit (<). **B)** Highest-numbered of up to three replicates per design; all 21 were unresolved. Values for LqhIII in Table S20.

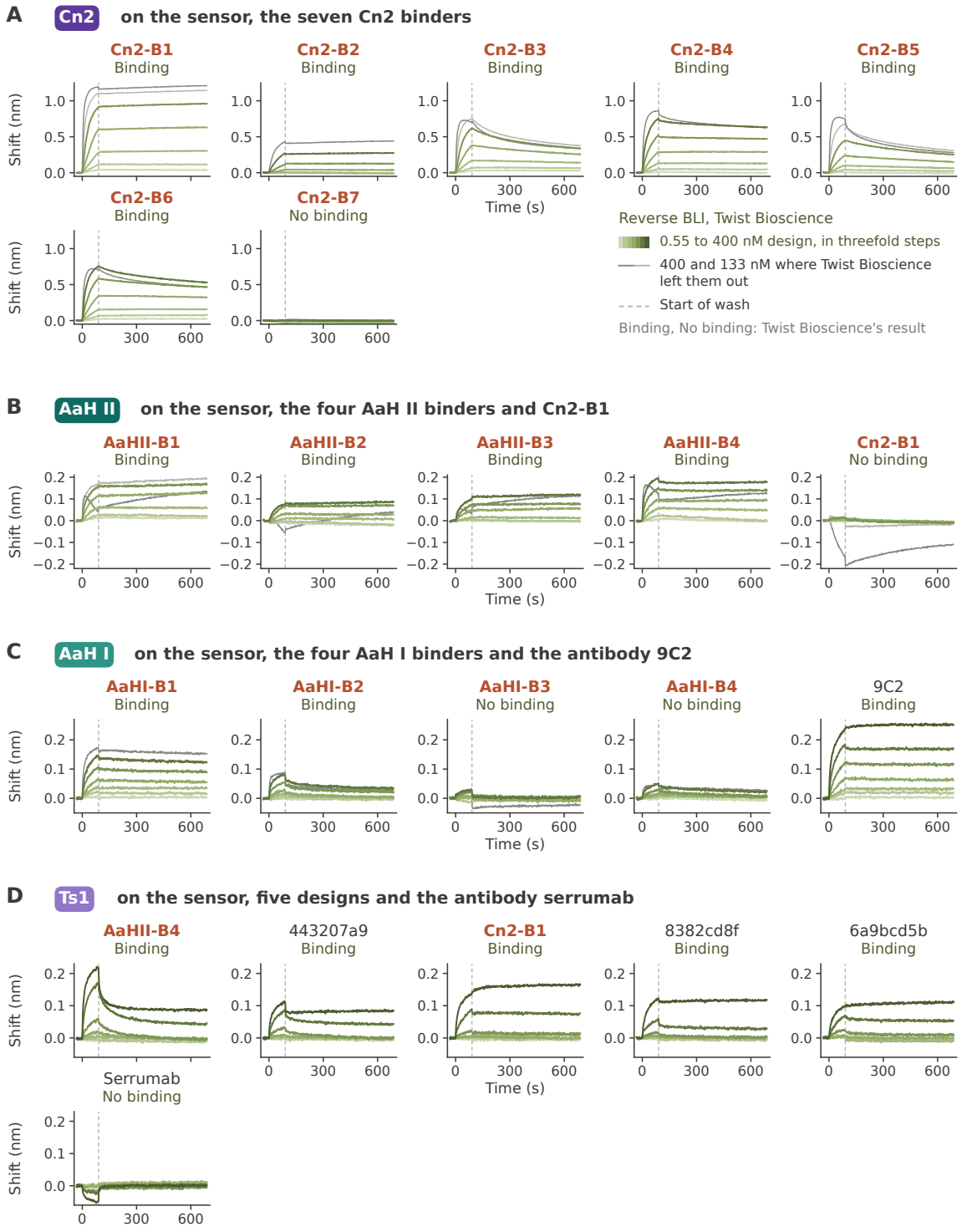


Figure S15. In reverse BLI at Twist Bioscience, 12 of the 15 binders tested were hits against their own toxin, and every design tested gave a signal against Ts1. His-tagged toxin captured on the sensor; design, a bivalent Fc fusion, in solution at 0.55 to 400 nM in threefold steps (443207a9 in D, 0.12 to 84 nM); 90 s association and 600 s wash; one series per pair. Gray, concentrations that were run but that Twist Bioscience left out. Headers give Twist Bioscience's result; no binding, under 0.05 nm at 80 to 82 s on one sensor (mostly 133 nM in A and C, 400 nM in B and D). **A)** The seven Cn2 binders. **B)** The four AaH II binders and Cn2-B1. **C)** The four AaH I binders and 9C2. **D)** Five designs and serrumab on Ts1; all but AaHII-B4 are among the 12 designs with a signal at the highest concentrations only. One scale per panel. The falls at 400 nM and the flat washes in A and B reflect the sample solution and the loss of toxin from the sensor, and serrumab's dip comes from the reference. 4C1 (not drawn) and serrumab did not bind.

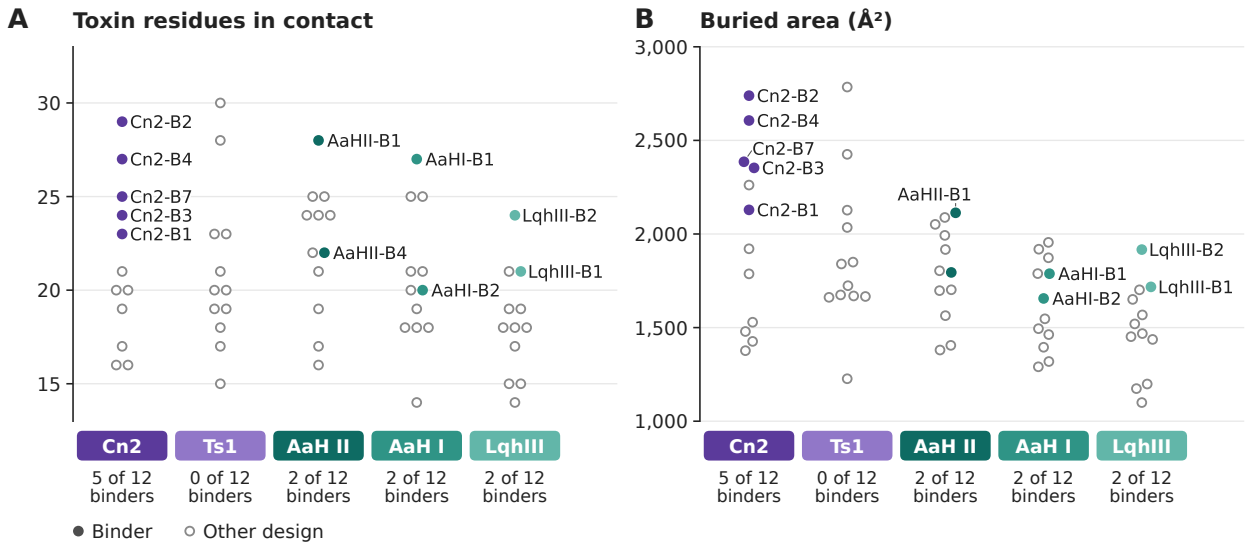


Figure S16. In predicted complexes, the 11 binders among the 60 single-toxin designs had larger interfaces, at the median, than the other 49. Each dot, one design with its intended toxin (ESMFold2). **A)** Toxin residues with a heavy atom within 5 Å of the design; medians, 24 versus 19. **B)** Area buried on both sides (probe 1.4 Å); medians, 2,113 versus 1,666 Å^2; the unnamed filled dot is AaHII-B4.

S4 Epitope and specificity assays

Two additional assays at Twist Bioscience tested where the designs bind Cn2 and whether they are polyreactive. Neither assay tests neutralization. Competition with reference antibodies, measured by BLI, covered 33 designs in 47 pairs with Cn2, Ts1, and AaH II, including 33 of the 34 pairs that were hits against these toxins at either lab (all but 925ee271 against Cn2). A separate run with 9C2 added 22 pairs with AaH I. Adaptyv Bio measured competition with LR and 9C2 in a different assay. Baculovirus particle (BVP) ELISA, for assessing polyreactivity, was run on the same 33 designs and, in a second run, on 18 more, the other designs that were hits against AaH I or LqhIII at either lab (Section S4.2). A control at Adaptyv Bio tested 52 designs in SPR against ovalbumin, an unrelated protein (Section S4.4).

S4.1 Competition with reference antibodies

Competition could be assessed for Cn2 and, in part, for AaH I. In tandem BLI (Table S3; Figure 3B), the toxin on the sensor was exposed to a reference antibody or to buffer and then to a design. The reference antibodies were 9004G, LR, and RU1 against Cn2; 4C1 against AaH II; serrumab against Ts1; and 9C2 against AaH I (Table S12). We assessed competition only where the antibody bound its toxin in the same assay. 4C1 did not bind AaH II, and serrumab did not bind Ts1. For LR and RU1, the assay worked as intended. Both bound the captured Cn2 (1.24 to 1.82 nm) and stayed bound in plain buffer (Figure S17A). When run as samples, each competed with itself but hardly with the other (Figure S17B), as expected for antibodies with separate epitopes. LR lost 96% of its signal after LR, and RU1 lost 87% after RU1, whereas RU1 lost 13% after LR, and LR lost 14% after RU1. The run with 9004G could not be assessed, because about half the captured Cn2 left the sensor in the 9004G wells before samples were added (Figure S18).

LR competed with all six Cn2 binders that could be assessed (87 to 95% of the signal lost), and five of the six lost little after RU1 (15 to 22%; Table S22; Figure S17C). RU1 binds a separate epitope, and in the crystal structure (PDB 4V1D) [18], LR and RU1 bind the same Cn2 molecule at once without touching each other (closest atoms 6.8 Å apart). RU1 therefore shows how much signal a design loses when an antibody of the same kind occupies another part of the toxin.

Cn2-B2 lost the most signal after RU1 (66%), and its predicted complex suggests why (Figure S19). We superposed the predicted complexes of the 16 designs tested against Cn2 on the crystal structure by the toxin and counted design atoms within 2.5 Å of RU1. With all three predictors, five residues of Cn2-B2 at a loop between two helices (D34 to P38; 32 atoms) overlap RU1's light chain, and its footprint holds 6 or 7 of RU1's 17 epitope residues (0 to 4 for the other designs; Table S22). In this pose, Cn2-B2 and RU1 cannot bind Cn2 at the same time, which suggests the assay told its pose apart from those of the other binders. Cn2-B1 overlaps RU1 only at its C-terminal end (R100 to A104; 14 atoms) and lost 16%, like designs without overlap; a free chain end may move aside more easily than a loop held between two helices. Cn2-B7 (signal too low to be assessed) has one such atom, and the other 13 designs have none. The models do not explain every loss, since e32555c5 (36%) and 8382cd8f (30%) lost signal after RU1 without predicted overlap.

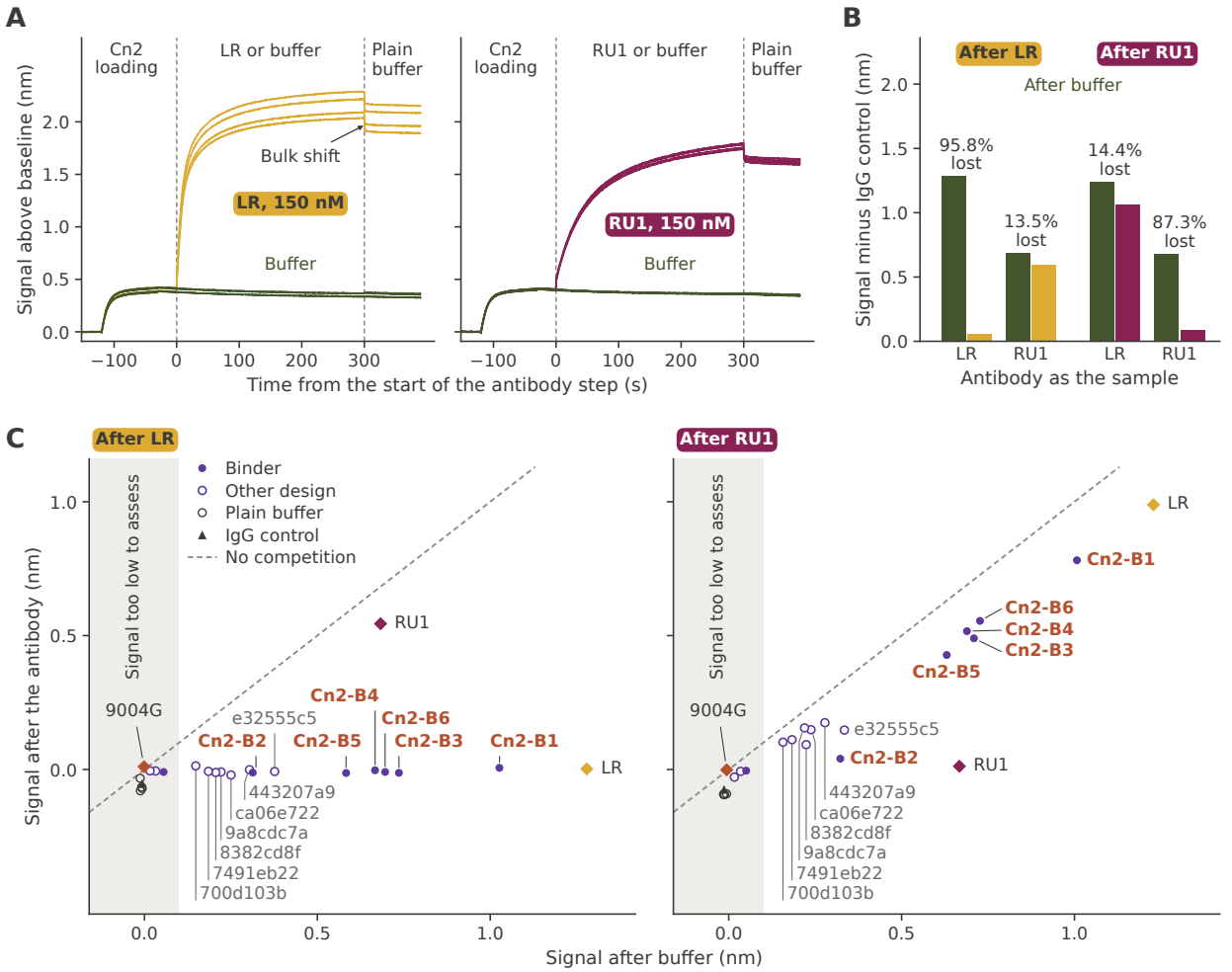


Figure S17. LR and RU1 bound the captured Cn2 and stayed bound, and each competed with itself but hardly with the other. Competition BLI at Twist Bioscience. **A)** Four sensors, two successive cycles each: Cn2 captured, then antibody or buffer for 300 s, then plain buffer. **B)** The two antibodies run as samples. **C)** Each sample's signal after the antibody against its signal after buffer, before the IgG control is subtracted; UUIDs name the seven other designs discussed in the text. With the control subtracted (Table S22), binders and other designs alike lost most of their signal after LR, and RU1 lost little; after RU1, all kept most of theirs except Cn2-B2 and RU1 itself.

LR also competed with seven designs that we did not label binders (Figure S20; Table S22). In the main screen at Adaptyv Bio, six were not hits against Cn2 and 9a8cdc7a was a hit in one of its three replicates, so none met criterion 1. At Twist Bioscience, 9a8cdc7a, e32555c5, 700d103b, 8382cd8f, and 7491eb22 were hits (K_D 490 to 870 nM), 443207a9 gave a weak or ambiguous signal, and ca06e722 was not a hit. The responses of all seven there rose with concentration and had not leveled off at 1 µM (8 to 19 RU). At Adaptyv Bio, the six that were not hits gave −7 to 32 RU at 1 µM, within the range of the other designs that were not hits (−41 to 53 RU), and 9a8cdc7a gave 26 to 77 RU. The three fusions among them (700d103b, 7491eb22, and 443207a9) carry 9a8cdc7a as their first domain. Designs 8382cd8f, 7491eb22, and 443207a9 also gave a signal against CLEC12A (criterion 4). All seven lost most of their signal after LR (58 to 88%) and far less after RU1 (−3 to 36%), as most of the binders did. Sticking to the sensor or to protein in general would not depend on which antibody occupied the toxin. In their predicted complexes, all seven sit on the surface that LR covers: their footprints hold 11 to 15 of LR's 21 epitope residues and none of RU1's. Four of them gave no signal against CLEC12A: the single domains ca06e722, e32555c5, and 9a8cdc7a and the fusion 700d103b. Three of the four, all but e32555c5, were also hits in reverse BLI at Adaptyv Bio, and all seven were hits in reverse BLI at Twist Bioscience (Section S3.10). Some of the seven may therefore bind the intended site very weakly. Four of the seven (9a8cdc7a, 700d103b, 7491eb22, and 443207a9) were hits in the control against ovalbumin (Section S4.4). Competition shows where a signal arises, not how strong or selective the binding is. Each value is one measurement on a signal of 0.15 to 0.38 nm (cutoff 0.10 nm), with the same preparations as Twist Bioscience's SPR. The 17 binders are thus a strict count, and designs near the detection limit are hard to classify.

The competition with LR matched what the predicted complexes led us to expect. The footprints of the double β-toxin designs lay a median of 71% inside LR's epitope and 0% inside RU1's (three predictors pooled), so we expected competition with LR only.

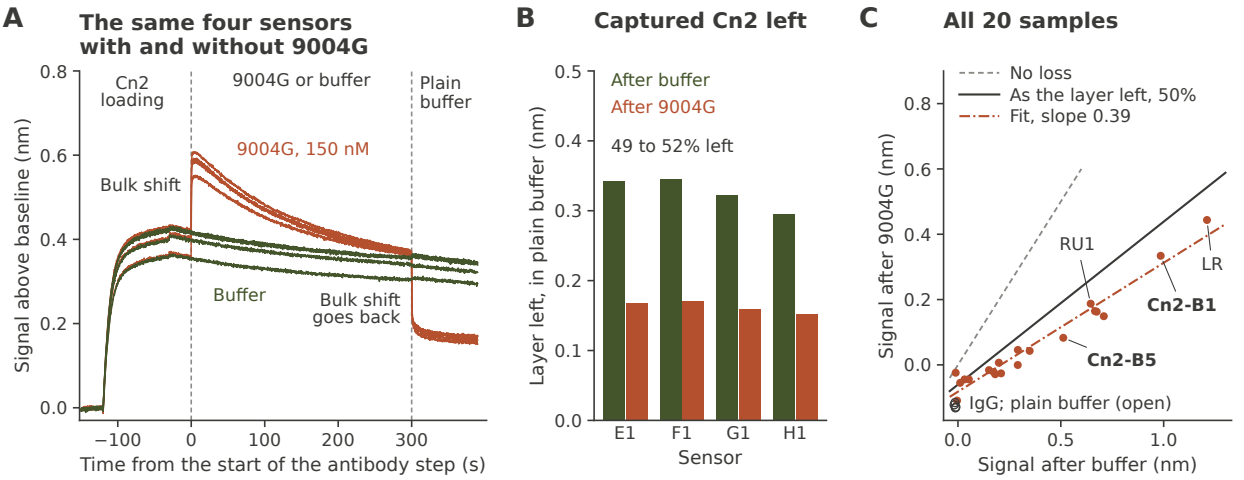


Figure S18. About half the captured Cn2 left the sensor in the 9004G wells before any sample was added. Raw biolayer interferometry, Twist Bioscience. **A)** Sensors E1 to H1 in two successive cycles: Cn2 captured, then 9004G (rust) or buffer (green) for 300 s, then plain buffer. On these four sensors, the signal with 9004G jumps 0.15 to 0.17 nm immediately (bulk shift) and then falls 0.19 to 0.23 nm; in buffer it falls 0.05 to 0.06 nm. **B)** Signal left at the end of A, measured from the baseline before loading: after 9004G, 49, 49, 50, and 52% of that after buffer. **C)** Each sample's response (last point minus first) after 9004G, plotted against its response after buffer. Solid line, 50% of toxin left; fit to the 20 samples, $r = 0.98$; open circles, plain buffer.

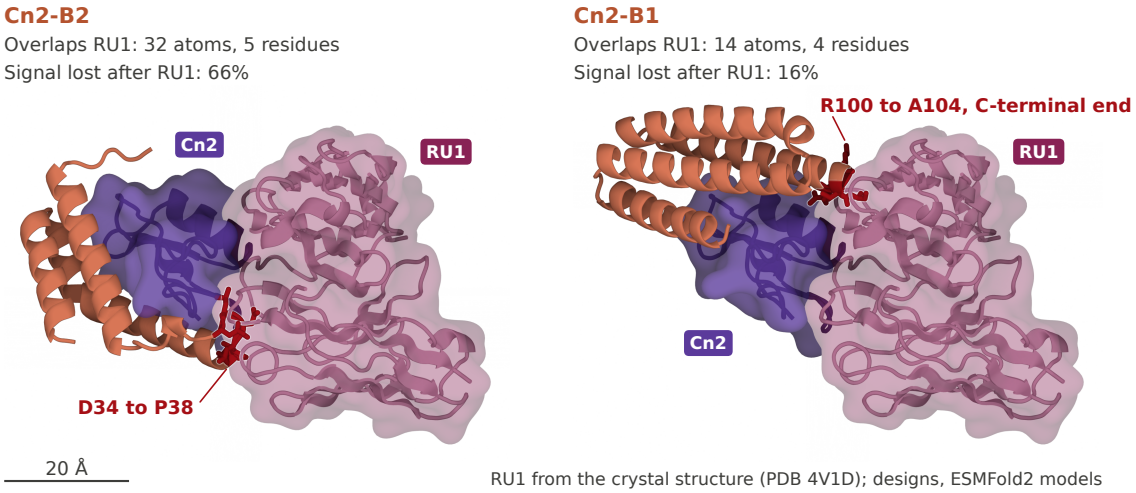


Figure S19. Cn2-B2, the design that lost the most signal after RU1, overlaps RU1 in its predicted complex. ESMFold2 models of Cn2-B2 and Cn2-B1 with Cn2; RU1 is placed from the crystal structure of Cn2 with LR and RU1 (PDB 4V1D) by superposing the toxins (Cα RMSD 0.5 Å), and nothing was relaxed. Red, design residues with an atom within 2.5 Å of RU1: in Cn2-B2, a loop between two helices; in Cn2-B1, the end of the last helix and of the chain (V102 does not overlap). RU1's surface is drawn thinner over the red residues.

With all three predictors (Table S22), all 16 footprints hold 10 to 21 of LR's 21 epitope residues. Epitopes are the Cn2 residues with a heavy atom within 4.5 Å of the antibody in its crystal structure, and footprints are defined likewise from each design's predicted complex (best seed). LR's epitope is pooled over LR (PDB 4V1D) [18] and its parent 9004G (PDB 2YC1) [66]. Competition shows that a design and the antibody cannot bind Cn2 at the same time. Because an antibody is larger than its epitope, competition does not determine the design's binding mode.

LR's epitope overlaps Cn2's channel-binding site. It holds 13 of the site's 20 residues (those within 5 Å of the channel), twice the 6 expected at random over Cn2's 66 residues. The ESMFold2 footprint of Cn2-B1 holds 16 of them, that of Cn2-B2 holds 9, and RU1's epitope holds 6. LR's epitope also holds four of the five hotspots (hotspots, Table S5; epitope, Figure 6A). The overlap with the hotspots follows from the design, because Claude chose hotspot E15 for its place in the epitopes of LR and 9004G.

AaH I. In the same assay, captured AaH I was exposed to 9C2 at 150 nM or to buffer and then to one of 22 designs, all those that were hits against AaH I at one lab or both (Figure S21). 9C2 bound the toxin (0.30 to 0.51 nm). The raw signals, however, cannot be read as they stand. Four designs gave no signal against AaH I at any concentration in Twist Bioscience's reverse BLI (Section S3.10):

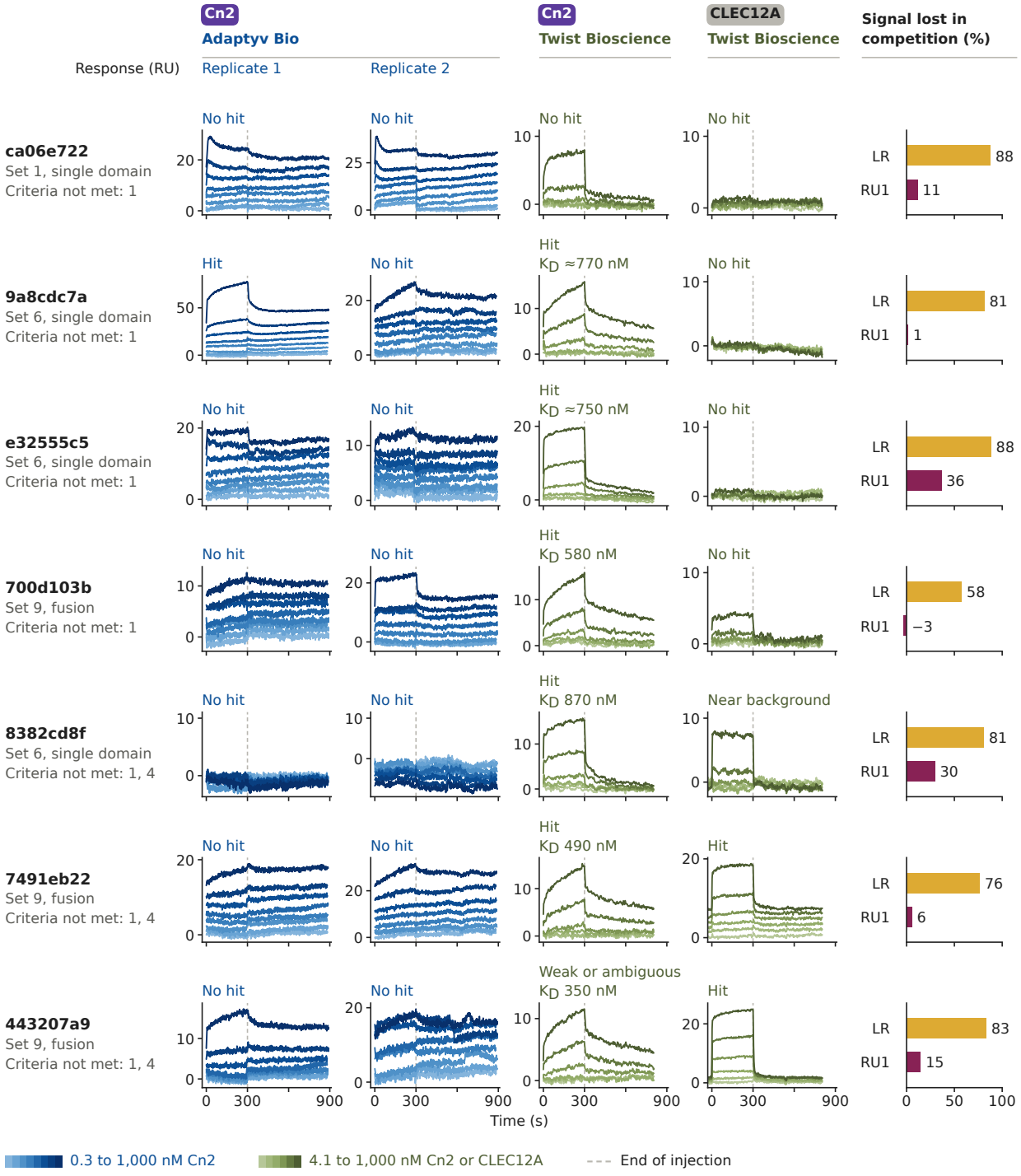


Figure S20. Seven designs with weak responses against Cn2 lost most of their signal after LR and far less after RU1. One row per design, with its set, format, and the criteria for binding that it did not meet (Methods). Main-screen SPR against Cn2 at Adaptyv Bio (two replicates; the third replicates of ca06e722, 9a8cdc7a, 7491eb22, and 443207a9, none a hit, are not drawn) and at Twist Bioscience, and against CLEC12A at Twist Bioscience (the run with the larger response). Headers give the result (at Adaptyv Bio, that of the replicate drawn) and, where fitted, K_D (≈, approximate); each panel has its own scale. At Adaptyv Bio, each curve starts where that of the next lower concentration ended. Bars, signal lost in competition BLI (Table S22).

Table S22. Competition for Cn2 at Twist Bioscience: signal lost when LR or RU1 bound the toxin first. One measurement per design and antibody. Signal without antibody, the same sensor's response after buffer, LR run. Lost after LR or RU1, the share of that signal lost when the antibody bound Cn2 first, after subtracting that run's IgG control from both responses (Figure 6D). Assessed, at least 0.10 nm without antibody, about the IgG control's largest drop, a threshold set after the data were seen. In LR's or RU1's epitope, how many of the epitope's 21 or 17 residues lie in the design's footprint with ESMFold2-Fast, ESMFold2, and Protenix v2. Bottom rows, LR and RU1 run as samples. n/a, not computed or not applicable.

Design	Signal without antibody (nm)	Lost after LR (%)	Lost after RU1 (%)	Assessed	In LR's epitope (21)	In RU1's epitope (17)
Cn2 binders						
Cn2-B1	1.027	94	16	yes	16, 18, 16	3, 4, 3
Cn2-B2	0.313	87	66	yes	10, 12, 12	6, 6, 7
Cn2-B3	0.736	95	22	yes	16, 15, 15	0, 0, 0
Cn2-B4	0.667	93	16	yes	19, 19, 18	2, 2, 2
Cn2-B5	0.583	93	22	yes	13, 12, 11	0, 0, 0
Cn2-B6	0.696	94	15	yes	17, 19, 20	0, 1, 1
Cn2-B7	0.055	n/a	n/a	no	21, 20, 19	2, 2, 0
Other designs						
700d103b	0.148	58	−3	yes	15, 15, 14	0, 0, 0
9a8cdc7a	0.221	81	1	yes	15, 15, 14	0, 0, 0
7491eb22	0.185	76	6	yes	15, 14, 14	0, 0, 0
ca06e722	0.250	88	11	yes	11, 11, 11	0, 0, 0
443207a9	0.304	83	15	yes	15, 14, 14	0, 0, 0
8382cd8f	0.206	81	30	yes	15, 14, 14	0, 0, 0
e32555c5	0.376	88	36	yes	12, 11, 12	0, 0, 0
05efdf6e	0.032	n/a	n/a	no	17, 15, 16	0, 0, 0
3f0f990a	0.016	n/a	n/a	no	18, 16, 13	0, 0, 0
Antibodies run as samples						
LR	1.279	96	14	yes	n/a	n/a
RU1	0.683	13	87	yes	n/a	n/a

7773743e, 4fcc0f88, 98052822, and cf8652bc. Yet they still rose by 0.06 to 0.09 nm after buffer and by no more than 0.005 nm after 9C2, so a design that binds nothing reads as almost fully blocked. The 22 samples also rose by 0.03 to 0.10 nm on the toxin-free sensors of the reverse BLI, which points to the sample solution. We therefore took the mean of the four as the zero after buffer and after 9C2. We give each loss net of it, with the range taken over the extremes of the four. By this measure 9C2, run as a sample, lost 99% of its signal after itself (94 to 102%). We assessed competition only for designs that were also hits against AaH I in Twist Bioscience's reverse BLI. Under this rule, 10 of the 22 designs could be assessed (AaHI-B1, AaHI-B2, and eight designs not labeled binders), and 12 could not, among them AaHI-B3 and AaHI-B4. Ten seconds into the wash of that assay, the ten still held 0.04 to 0.15 nm, and the four designs of the zero held −0.01 to 0.01 nm.

9C2 competed with AaHI-B1, which lost 80% of its signal (74 to 83%), and with AaHI-B2, which lost 89% (80 to 93%). The other eight lost 66 to 90%. For the ten designs, however, 27 to 67% of the signal after buffer lies at the level of the zero, and with the least favorable zero the eight lost only 26 to 70%. For five of them, that value falls within what RU1 cost the Cn2 binders (15 to 66%); RU1 binds a separate epitope. We therefore do not say that 9C2 competed with any of the eight.

This result is weaker than the one for Cn2. It rests on one antibody. Because no sample kept its signal after 9C2, the run cannot show that a bound antibody leaves any design unblocked, as RU1 did on Cn2. No structure of 9C2 on AaH I has been solved [67], so competition places no design on a known site. On every sensor, the cycle with 9C2 came first and the cycle with buffer second. The run included no IgG control and no toxin-free sensor, the designs' concentration was not recorded, and each value is a single measurement.

Competition at Adaptyv Bio. At Adaptyv Bio (Table S3), the antibody (LR or 9C2, as an scFv) sat on the probe, and 200 nM toxin premixed with 1 µM His-tagged design was in solution. Competition thus shows as less toxin captured. Only Cn2 and AaH I could be assessed. Adaptyv Bio's coefficient allows for the design's mass, so a design that changes nothing scores 0.56 to 0.77. We therefore report the plain ratio to toxin alone and count only their call Potentially competing as competition (Table S23; Figure S22).

With LR, Cn2-B2 gave 0.05 and 0.18 (toxin alone, 0.90 to 1.13). Cn2-B1 showed no capture of toxin, but its traces fell below the baseline, so its ratio is not quantitative. With 9C2, toxin alone scattered more (0.33 to 1.06 in the first run and 0.43 to 1.04 in the second): AaHI-B1 lay clearly below it (0.06 and 0.11), AaHI-B2 narrowly (0.25 and 0.37), and AaHI-B4 in one of two runs (0.18 and 0.45). The two labs agree firmly on Cn2-B2 and AaHI-B1 and less firmly on Cn2-B1 and AaHI-B2. Only Adaptyv Bio has a result for AaHI-B4, and neither lab for AaHI-B3. This assay cannot detect the five weaker Cn2 binders, so only Twist Bioscience has a result for four of them, and neither lab has one for Cn2-B7. At equilibrium, 1 µM design with a K_D of 40 to 580 nM leaves 9 to 79 nM toxin free, far above LR's K_D.

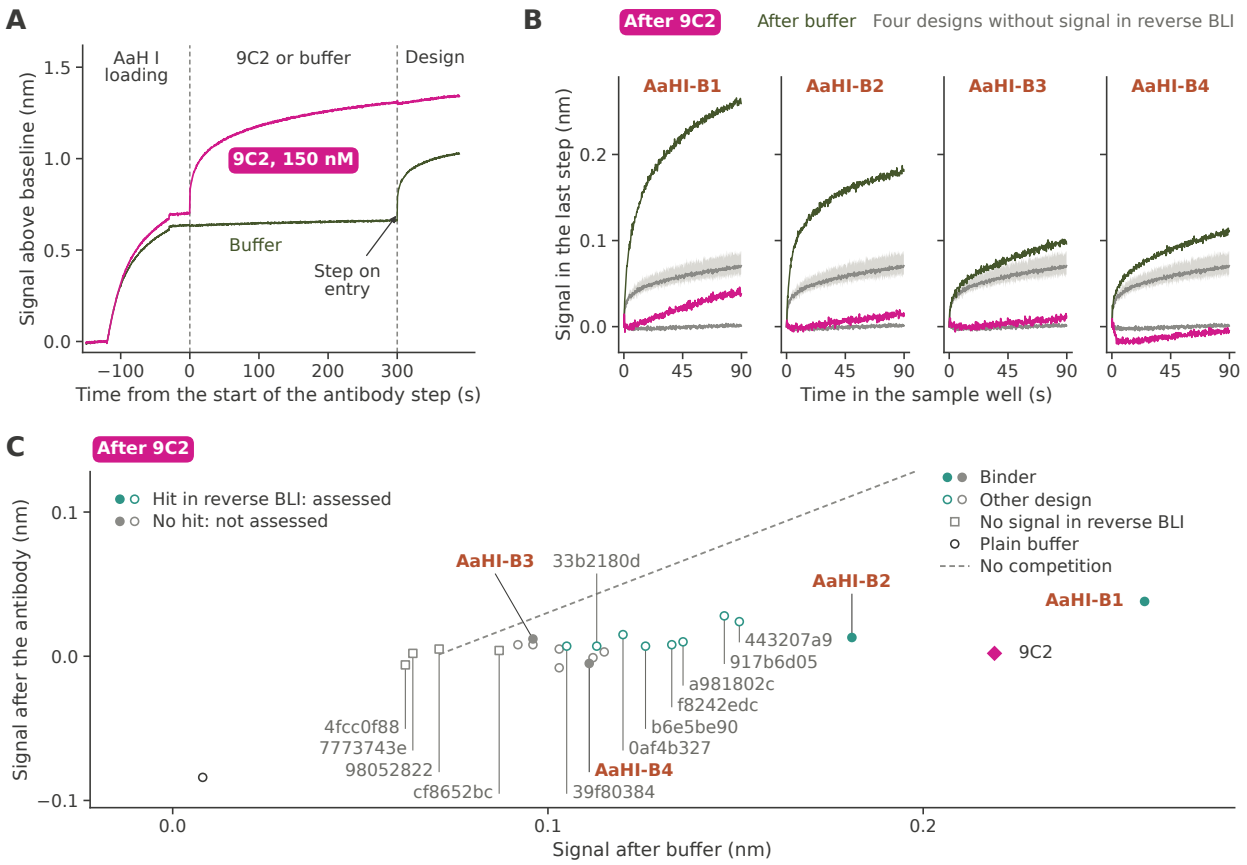


Figure S21. The neutralizing antibody 9C2 competed with AaHI-B1 and AaHI-B2. Competition BLI at Twist Bioscience, one measurement per sample. **A)** One sensor in two successive cycles: AaH I captured, then 9C2 at 150 nM or buffer for 300 s, then AaHI-B1; raw signal from the baseline before loading. **B)** The four AaH I binders in the last step, from the step's first point; gray, the four designs without signal against AaH I in reverse BLI: mean and range after buffer (upper) and mean after 9C2 (lower). **C)** Each sample's signal in the last step (last point minus first, nothing subtracted), after 9C2 against after buffer. Teal, a hit against AaH I in Twist Bioscience's reverse BLI (at least 0.05 nm); competition is assessed for these 10 of the 22 designs. Gray, no hit (12), not assessed, including AaHI-B3 (0.028 nm) and AaHI-B4 (0.048 nm). Filled, binders; open, other designs. Squares, the four gray designs without signal at any concentration; their mean (0.071 nm after buffer, 0.001 nm after 9C2) is the zero, and the dashed line of slope 1, no competition, starts there. Plain buffer fell by 0.084 nm after 9C2.

S4.2 BVP ELISA

BVP ELISA tests for polyreactivity more broadly than SPR against CLEC12A, because baculovirus particles present many kinds of molecules at once. Twist Bioscience ran the assay twice, on 33 designs and on 18 more, with no design in both runs. Of the 51 designs, we excluded 3f0f990a, whose signal matched the wells without sample. The other 50 gave 4 to 108% (median 75%) of the signal of cixutumumab, the highest of Twist Bioscience's three control antibodies on the same plates (palivizumab, 12 and 13% in the two runs; atezolizumab, 27 and 20%). The 32 designs of the first run gave 4 to 108% (median 51%), and the 18 of the second gave 34 to 106% (median 90%). Figure 6E shows the seven Cn2 binders, and Section S4.3 all 50. Twist Bioscience's 12 reference nanobody-Fc fusions (VHH-Fc), run on another plate on the day of each run and scaled to that plate's cixutumumab, gave 28 to 94% in the first run and 16 to 102% in the second. We compared each design with the range of its own run: 7 of the 50 designs fell below it, 38 within it, and 5 above it (second run: none, 17, and 1). Each design was read at one concentration: 0.07 to 0.13 mg/mL for 45 of the 50 and 0.02 to 0.05 mg/mL for the other five. We did not apply the assay's published cutoff, a normalized score above 5 set on IgG [68], because Twist Bioscience's readings are on a different scale.

BVP ELISA and SPR against CLEC12A flagged partly different designs. Of the 12 designs measured here that were hits against CLEC12A, 10 fusions and the single domain cac1ca4c gave 73 to 100%, whereas 05efdf6e, the strongest against CLEC12A, gave 8%. Of the 38 that were not hits against CLEC12A, 17 gave 68 to 108%, and seven of the 17 are binders, four of Cn2 and three of AaH I. Net charge explains the signals only in part. More negative designs tended to give higher signals (Spearman's $\rho = -0.57$, 95% interval -0.75 to -0.33), but the fusions are the largest designs, and 14 of the 16 are among the 16 most negative. Three 74-residue designs of nearly equal net charge gave 13 to 95%. We also asked whether the signal follows from the designed interface, with every definition fixed before the comparison (Figure S23). In the ESMFold2 complex of each design with the toxin of its design model,



Figure S22. At Adaptyv Bio, competition was clearest for Cn2-B2 with LR and for AaHI-B1 with 9C2. A) Antibody fragment on the sensor, then 200 nM toxin, alone or mixed with 1 μM design. **B)** Association on LR, six designs. **C)** Highest signal per antibody loaded, relative to toxin alone; order, K_D at Adaptyv Bio (≤, upper limit; ≈, approximate). The five weaker binders (0.22 to 4.5) spread as widely as designs that are not hits (0.01 to 6.5). Run 1 loaded too little LR. Adaptyv Bio's coefficient is 0 where toxin and design are captured together; a design that changes nothing scores 0.56 to 0.77. **D and E)** The same on 9C2; AaHI-B3 was not measured. Where a trace falls below zero, as for Cn2-B1, the ratio shows only that no signal rose.

Table S23. Competition at Adaptyv Bio: toxin captured by the antibody in the presence of a design, relative to toxin alone. BLI with the antibody on the probe and 200 nM toxin premixed with 1 μM His-tagged design. Hit, against that toxin in Adaptyv Bio's main screen (n.m., not measured). Ratio, the 99th percentile of the association signal per nm of antibody loaded, over the median of the run's toxin-alone probes (16 for Cn2, 14 for AaH I); 1 means no change. For Cn2, the two runs with 3.5 to 7.8 nm of antibody are shown (runs 2 and 3 of Figure S22), and one with 0.03 to 3.2 nm is left out (run 1); Adaptyv Bio's calls rest on all three. The traces of Cn2-B1 ended the association below the baseline (−0.61 and −1.27 nm), as did at least one trace of each measured design called Failed. Nine other designs, AaHI-B3 among them, have no measurement against AaH I. No design was run alone on an antibody-loaded probe.

Design	Hit	Ratio, first run	Ratio, second run	Adaptyv Bio's call
Cn2, LR on the probe				
Cn2-B1	yes	0.00	−0.08	Potentially competing
Cn2-B2	yes	0.05	0.18	Potentially competing
Cn2-B3	yes	0.74	1.31	Potentially not competing
Cn2-B4	yes	0.24	0.40	Failed
Cn2-B5	yes	0.75	1.49	Potentially not competing
Cn2-B6	yes	1.35	4.46	Potentially not competing
Cn2-B7	yes	0.22	1.41	Partially competing
443207a9	no	0.07	0.94	Failed
7491eb22	no	0.01	1.01	Failed
700d103b	no	0.58	1.04	Failed
8382cd8f	no	0.68	0.98	Partially competing
05efdf6e	no	0.69	1.24	Partially competing
3f0f990a	yes	0.89	1.99	Potentially not competing
e32555c5	no	0.65	2.42	Potentially not competing
9a8cdc7a	yes	0.90	2.89	Potentially not competing
ca06e722	no	1.65	6.45	Potentially not competing
AaH I, 9C2 on the probe				
AaHI-B1	yes	0.06	0.11	Potentially competing
AaHI-B2	yes	0.25	0.37	Potentially competing
AaHI-B4	yes	0.18	0.45	Potentially competing
0af4b327	yes	−0.01	−0.07	Failed
7491eb22	yes	−0.01	−0.06	Failed
6d0b4f3f	yes	−0.01	−0.04	Failed
b6e5be90	yes	0.11	0.28	Potentially competing
f8242edc	yes	0.15	0.47	Potentially competing
cf8652bc	yes	0.30	0.63	Potentially competing
0c86c52b	no	0.65	1.27	Partially competing
443207a9	yes	1.08	1.01	Partially competing
4fcc0f88	n.m.	1.33	1.58	Partially competing
917b6d05	yes	1.68	2.61	Potentially not competing
40090fca	no	1.81	3.79	Potentially not competing

the apolar share of the design's buried area did not track the signal (Spearman's $\rho = 0.04$, 95% interval −0.25 to 0.32). The largest hydrophobic patch on the free design did ($\rho = 0.42$, 0.16 to 0.62), but not within either format: among the 34 single domains both values were near zero ($\rho = -0.01$ and 0.10), and among the 16 fusions the patch gave $\rho = -0.11$. The fusions have larger patches (median 654 versus 445 Å^2) and gave higher signals. The 50 designs form only 26 independent groups, and each well held an Fc fusion measured in one run at one concentration, so intervals this wide do not exclude a moderate correlation.

S4.3 Unintended binding

Signs of unintended binding were rare among single domains and more common among fusions (Figures S24 and S25). In the main screen, 301 pairs of a single domain with a toxin it was not made against could be assessed, all at Twist Bioscience. Of these, 2 were hits (0.7%) and 2 gave a weak or ambiguous signal (Figure S25C). With an intended toxin, 46 of 519 pairs were hits (8.9%), or 25 of 391 (6.4%) without the fusions. The first of the two hits, AaHII-B4 against Ts1, was found at both labs, but the signal exceeded the theoretical R_{max} at Twist Bioscience, so it did not meet criterion 3 (discussed below). The other, 4fcc0f88 against AaH I, was found at Twist Bioscience from irregular traces, and the pair was not run at Adaptyv Bio, so criterion 1 could not be tested. Only 14 of the 301 pairs were also measured at Adaptyv Bio, so the criteria for binding could not be applied to most, and we report hits rather than labels. Cn2-B1 gave signals in Adaptyv Bio's main screen that they did not classify as binding (69 RU against Ts1 and 85 RU against AaH II at 1 μM, one SPR replicate each), whereas Twist Bioscience recorded −2 and 6 RU against these toxins. At pH 7.4, 209 of the 211 designs carry a negative charge (median −8.5; computed as in Table S17), and four of the five toxins are basic. If opposite charges were enough for a hit, unintended hits would be common, but among single domains they were rare.

Against CLEC12A, 13 designs were hits in either run: 11 of the 32 fusions and 2 of the 173 single domains with an assessable run (Figure S25D). These 13 were more negative than the rest, largely because 11 of them are fusions. In the BVP ELISA, 1 of the 16

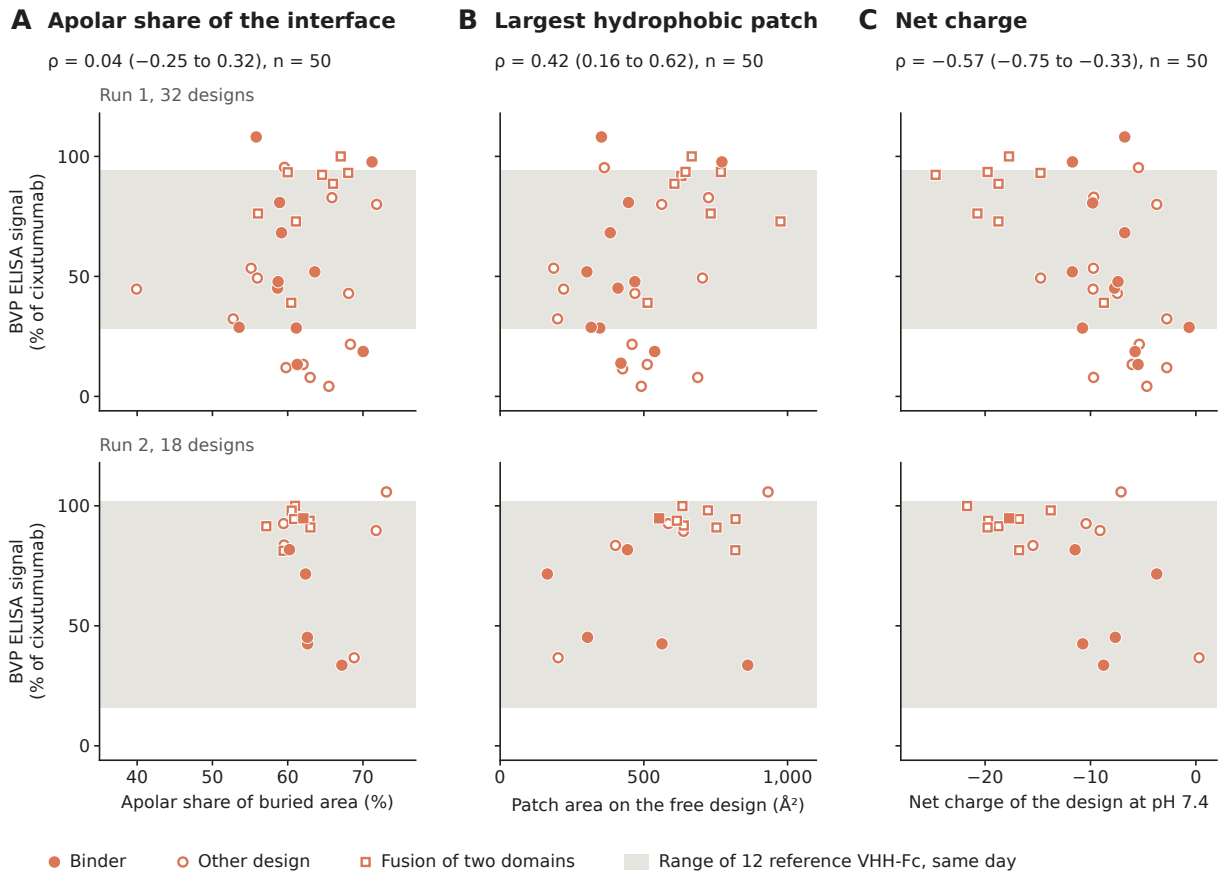


Figure S23. The apolar share of the predicted interface did not track the BVP ELISA signal, and the largest hydrophobic patch tracked it only across the two formats. Signal of the 50 designs, tested as Fc fusions at Twist Bioscience (mean of two plates), against three properties computed for the design alone. Top row, first run; bottom row, second run. **A)** Share of carbon and sulfur atoms in the design's area buried by the toxin, in the ESMFold2 complex with the toxin of the design model; the two interfaces of a fusion are pooled. **B)** Largest contiguous apolar area on the design taken alone from the same complex. **C)** Net charge at pH 7.4, from the sequence. Above each column, Spearman's ρ with its 95% bootstrap interval for all 50 designs. Gray band, range of Twist Bioscience's 12 reference VHH-Fc fusions on the day of that run (28 to 94% and 16 to 102%). Filled, binder (17); open, other design (33); circle, single domain (34); square, fusion of two domains (16, one of them a binder).

fusions and 4 of the 34 single domains read above the reference range of their run (Figure S25B). Against ovalbumin, tested on 52 designs, 5 of the 16 fusions and 6 of the 36 single domains were hits (Section S4.4), and these hits are not counted in Figure S25. Of the nine designs that were hits against multiple toxins (Figure S24A), four were hits at both labs against two toxins: AaHII-B4 and three fusions, which were hits against AaH I, Ts1, and CLEC12A.

AaHII-B4 was designed against AaH II only but was also a hit against Ts1 at both labs (Figure S24). Its K_D against AaH II was 21 nM by SPR at Adaptyv Bio and 6.3 nM at Twist Bioscience (Table S18), and against Ts1 ≤9.1 nM at Adaptyv Bio (the mean of one BLI and two SPR replicates) and 28 nM at Twist Bioscience. The signals against Ts1 need not be specific. At Adaptyv Bio, Fabs of four antibodies raised against other toxins (9004G, LR, RU1, and 9C2) gave signals against Ts1, with mean apparent K_D values of 18 to 66 nM, at or above that of the anti-Ts1 antibody serrumab (15 nM, two SPR replicates). The scFvs of 9004G, LR, and RU1 and both formats of 4C1 gave none (Section S3.4). At Twist Bioscience, every hit against Ts1 carried more signal than one-to-one binding allows. For AaHII-B4, the fitted R_{max} was 4.2 times the theoretical R_{max} against Ts1, compared with 0.54 against AaH II (Figure S24D; criterion 3, Figure 3D). The ratio was 1.5 to 4.2 for the five designs that were hits and 11 for the anti-Ts1 control antibody (Table S11). The predictors scored AaHII-B4 high against AaH II and low against Ts1 (final score 0.92 versus 0.23; ipSAE 0.85 to 0.87 versus 0.17 to 0.52; scDockQ 0.97 to 0.99 versus 0.13 to 0.14; Figure S24B). We therefore label AaHII-B4 a binder of AaH II only. Whether it also binds Ts1 specifically remains open. In a reverse BLI run at Twist Bioscience in which serrumab did not bind, it gave the largest response against Ts1 of the 14 designs tested, all of which gave a signal (Section S3.10). Reverse BLI at Adaptyv Bio gave 60 nM against AaH II, the mean of two replicates whose fits they graded as Poor and Medium, and no result against Ts1, where all 10 designs tested stayed unresolved.

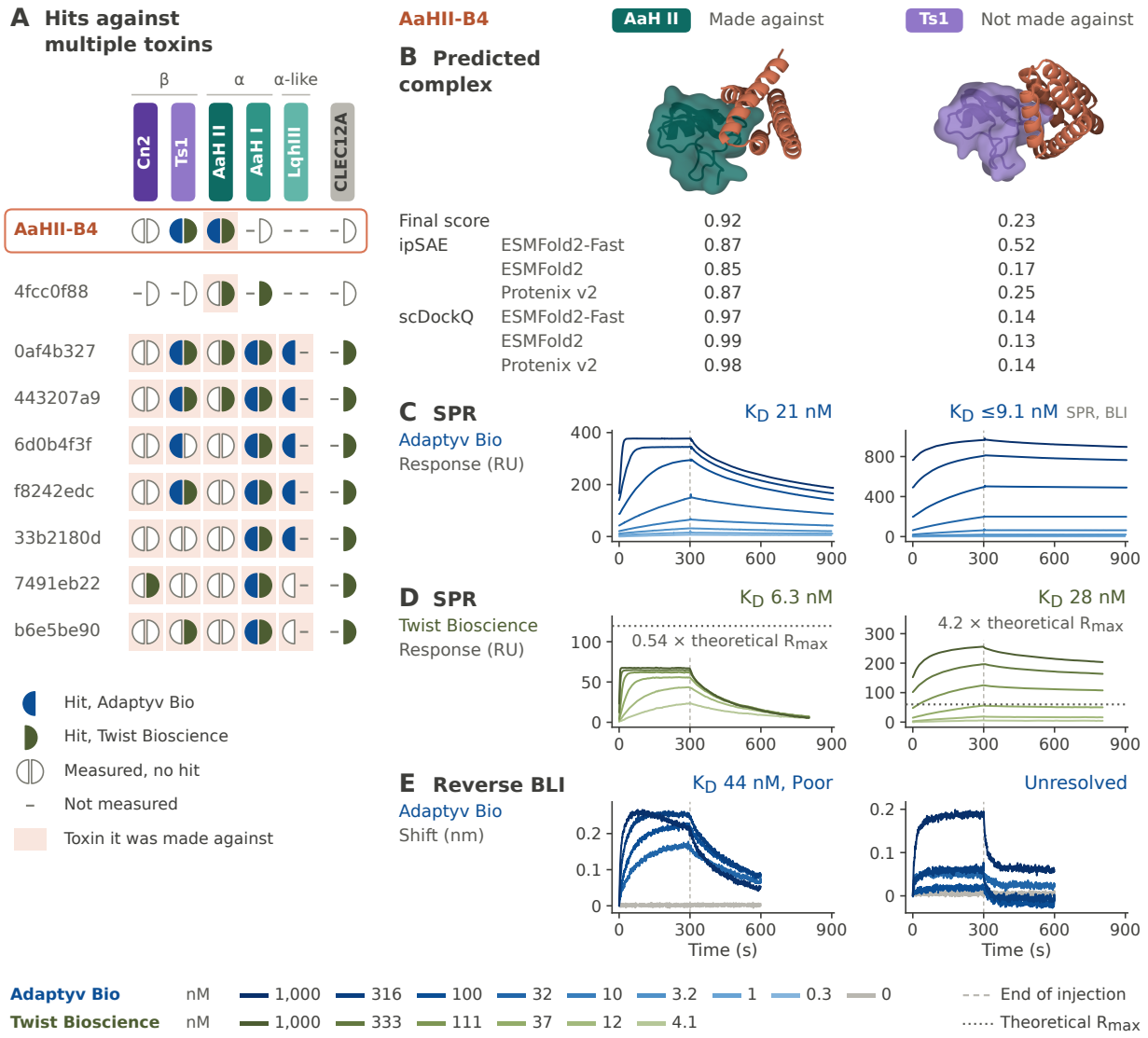


Figure S24. AaHII-B4, made against AaH II, was also a hit against Ts1 at both labs; we label it a binder of AaH II only, because at Twist Bioscience its fitted R_{max} against Ts1 was 4.2 times the theoretical R_{max}. A) The nine designs that were hits against multiple toxins; CLEC12A, off-target binding. **B)** Predicted complexes (ESMFold2), final score, ipSAE, and scDockQ. **C and D)** Main-screen SPR at Adaptyv Bio and Twist Bioscience; in C, the gray tag marks a mean over SPR and BLI replicates; in D, the dotted line marks the theoretical R_{max}, and the number gives the fitted R_{max} as a multiple of it. **E)** Reverse BLI, replicate 1 of two: against AaH II, under its own K_D and fit grade (the two average 60 nM); against Ts1, both were unresolved. ≤, upper limit.

S4.4 SPR against ovalbumin

The control. The Adaptyv Bio team tested 52 designs in SPR against ovalbumin, an egg-white glycoprotein of 45 kDa [91] unrelated to the toxins. They were 48 of the 49 designs that were hits against any toxin at either lab (all but 925ee271), all 17 binders among them, and four others. As in the main screen (Section S1.2), the design was captured, here after purification, and ovalbumin was in solution, at 10, 31.6, 100, 316.2, and 1,000 nM. The concentrations rose without regeneration, so each injection started from the level left by the one before and the response at 1 µM is cumulative. Each design had two replicates over two runs, with two blank spots per run, and the designs' median capture was about 700 RU in both runs, within the range of the main screen's runs. Adaptyv Bio reported hit or no hit for each design and fitted no model. Their team advises against reporting a K_D or binding strength from this control, because it gives hit or no hit only. The responses, and the shares of the theoretical R_{max} derived from them, are our own readings of their traces. The runs included no positive control for ovalbumin, and the other 159 designs were not tested, so the rate of hits among designs that bind no toxin is unknown. We read all capture levels from the loading traces and left out the 3 of the 104 series whose spots held no design, so AaHI-B3, 6a9bcd5b, and 569c6be9 have one replicate each (Table S24; Figure S26). One more loading trace could not be read (4eae42fc, replicate 1), so that series is counted in RU only.

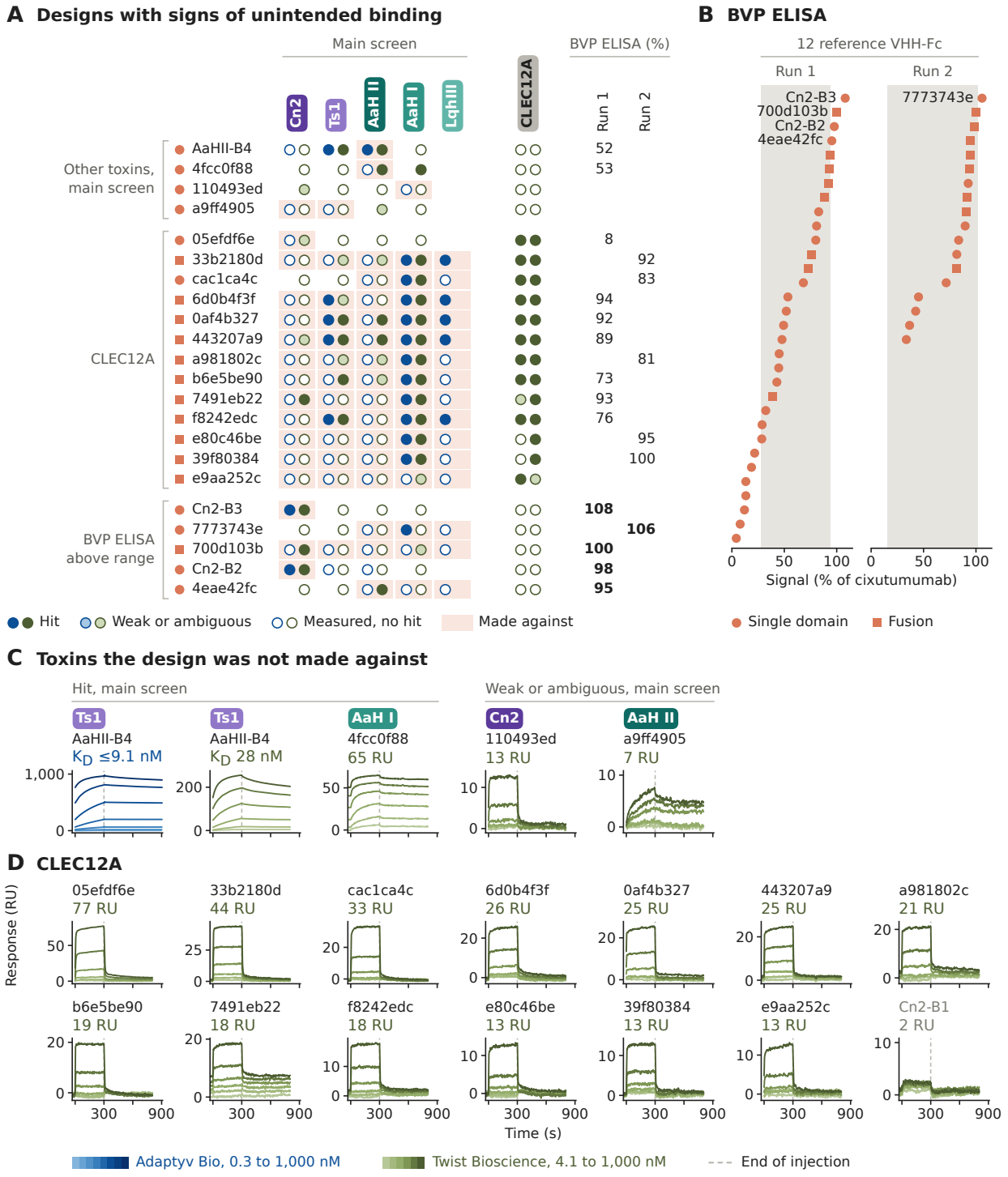


Figure S25. Signs of unintended binding were rare among single domains (10 of 179) and more common among fusions (12 of 32). A) The 22 designs with at least one sign (a hit or a weak or ambiguous signal against another toxin; a hit against CLEC12A; or a BVP ELISA signal above the reference range of its run, in bold), grouped by the sign (brackets; other toxins, toxins the design was not made against); blue, Adaptyv Bio; green, Twist Bioscience (CLEC12A: two runs). Fusions were made against all five toxins, so among these signs only CLEC12A and the BVP ELISA can show unintended binding by them. Eleven additional designs gave only a weak, ambiguous, or near-background signal against CLEC12A (not drawn). **B)** BVP ELISA, 50 designs in two runs, mean of two plates; gray bands, range of Twist Bioscience's 12 reference VHH-Fc on the day of each run (28 to 94% and 16 to 102%); named, the designs above it. **C)** Sensorgrams of the pairs in A's first group. **D)** CLEC12A, the run with the larger response; gray, Cn2-B1, not a hit. RU, response at 1,000 nM.

Hits. Eleven of the 52 designs were hits. Two were binders, Cn2-B1 and AaHI-B4, and nine came from the 35 other designs. At the end of the 1 µM injection, Cn2-B1 gave 246 and 254 RU (one replicate per run), and AaHI-B4 gave 181 and 210 RU (both replicates in one run). The other 15 binders gave 0 to 53 RU (median 19 RU), the nine other hits gave 78 to 734 RU, and the blank spots stayed at or under 5 RU. Replicates differed 1.4-fold at the median, and the response did not follow the capture level (Spearman's ρ 0.07 over 100 spots). The two LqhIII binders were not hits (12 to 22 RU); they were measured only at Adaptyv Bio and had CLEC12A at Twist Bioscience as their only other off-target control.

Size and shape. Ovalbumin is about four to five times heavier than the two toxin constructs compared here (Cn2, 9.6 kDa; AaH I, 11.2 kDa; Table S11), so an equal response in RU means far fewer molecules of ovalbumin. We therefore compared responses as shares of the theoretical R_{max} for one molecule per captured design. At 1 µM ovalbumin, the share was 13% for Cn2-B1, 11% for AaHI-B4, 0.4 to 2.9% for the other 15 binders, and 7 to 41% for the nine other hits. Cn2-B1 reached the same share against Cn2 at 0.45 and 0.59 nM in its two SPR replicates, about 2,000 times less toxin than ovalbumin. AaHI-B4 reached the same share against AaH I at 52 nM in its SPR replicate, about 20 times less. Among the other hits, 9a8cdc7a reached its share (7%) against Cn2 only at 64, 290, and 400 nM in its three replicates, about 3 to 16 times less. In all 11 hits the response rose with every injection, showed no plateau, and kept 88 to 97% of its level through the 10 min wash. The data therefore give no K_D. This shape is more typical of nonspecific adsorption than of binding at a defined site, but these data cannot tell the two apart.

What the response tracked. The response did not track the other specificity assays. Four of the 11 hits had a signal against CLEC12A at Twist Bioscience, and 12 of the 16 tested designs with such a signal were not hits against ovalbumin. With the BVP ELISA, ρ was −0.05 (50 designs). We compared the mean response at 1 µM with 25 measures of sequence, predicted structure, production, and the other assays (Figure S27), and three passed a Bonferroni limit of 0.002. They were the share of lysine and arginine (ρ 0.50), the number of lysines and arginines (0.43), and exposed lysines and arginines per surface area (0.42). The calculated isoelectric point fell just short (0.42), and measures of hydrophobicity ran the other way (−0.27 to −0.31). Ovalbumin is acidic (isoelectric point 4.5) [71], as are 16 of the 17 binders (4.0 to 4.8, calculated without tags from the EMBOSS pK_a values [92]). The exception is Cn2-B1 (7.8), which is near neutral: with the pK_a set of Table S17, its net charge at pH 7.4 is −0.6. The response also tracked the domain. Design 9a8cdc7a (21% lysine and arginine) and all three fusions carrying it were hits, while none of the four fusions carrying 47d04ba8 (15%) was. Charge does not explain every hit, though. AaHI-B4 (4.2) and the strongest hit, 173c84c2 (4.4), are acidic. Two relatives of AaHI-B4, 91% and 77% identical to it and with the same number of lysines and arginines, were not hits. The correlation also rested on lysine (0.40) rather than arginine (−0.06). The analysis was exploratory, and the designs are not independent, since six of the 11 hits share one backbone family. With one value per family (32 families), ρ for the share of lysine and arginine was 0.41 and no longer passed the limit. A repeat at higher salt would test the part played by charge. Ovalbumin is not among the four criteria, and hits against it are flagged. Requiring no hit against ovalbumin as well would leave 15 of the 17 binders (13 of the 15 pairs measured at both labs in Table S16, which does not count the two LqhIII pairs).

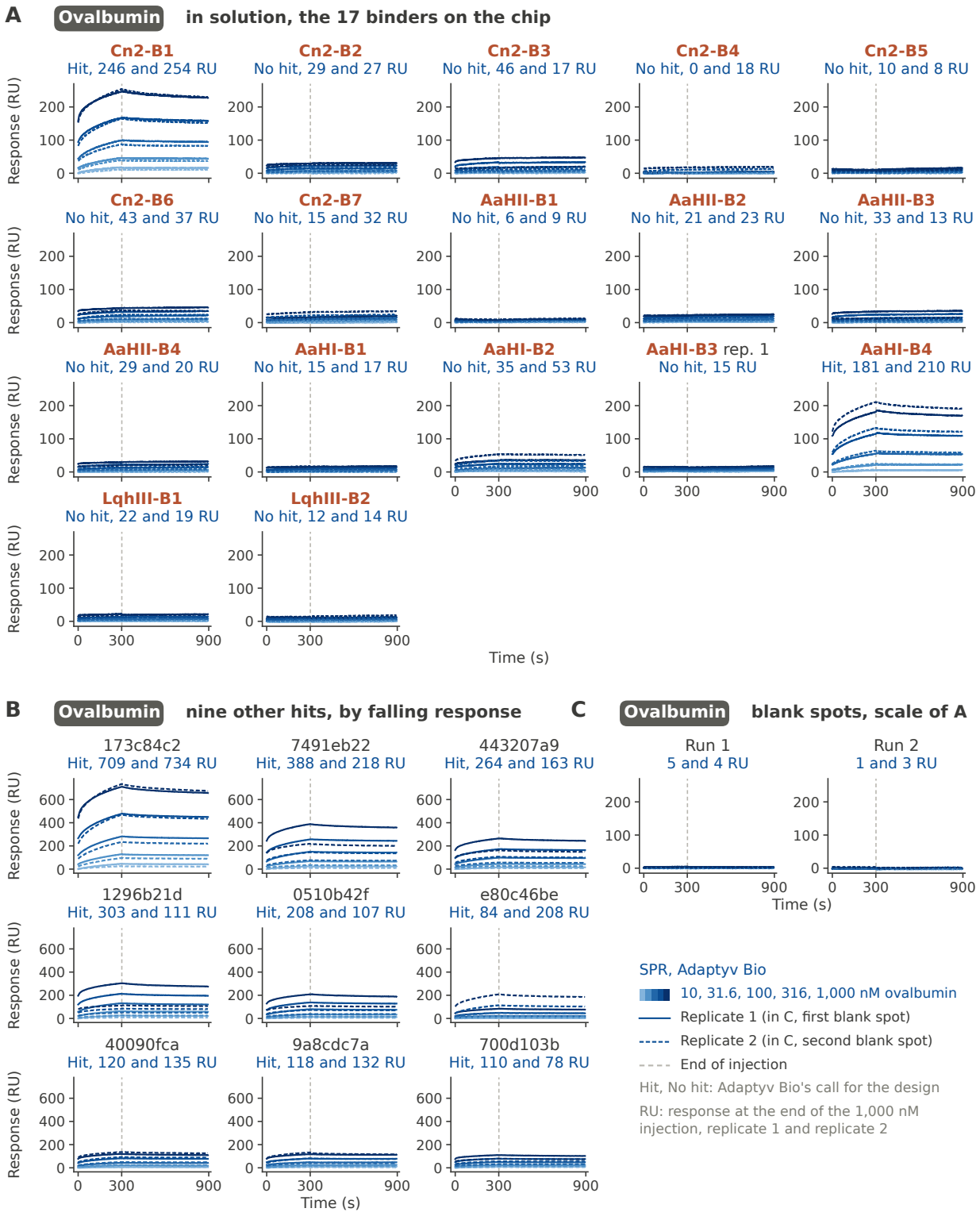


Figure S26. Against ovalbumin, Cn2-B1 and AaHI-B4 gave clear responses, and the other 15 binders stayed near the baseline. SPR at Adaptyv Bio with the design captured and ovalbumin in solution at 10 to 1,000 nM, in rising order without regeneration, so each injection starts where the last one ended; both replicates in each plot. **A)** The 17 binders, on one scale. **B)** The nine other designs that were hits, by falling response, on one wider scale. **C)** The blank spots of the two runs, on panel A's scale. Headers give Adaptyv Bio's result and the response (RU) at the end of the 1 µM injection in each valid replicate. AaHI-B3 has one valid replicate. Values in Table S24.

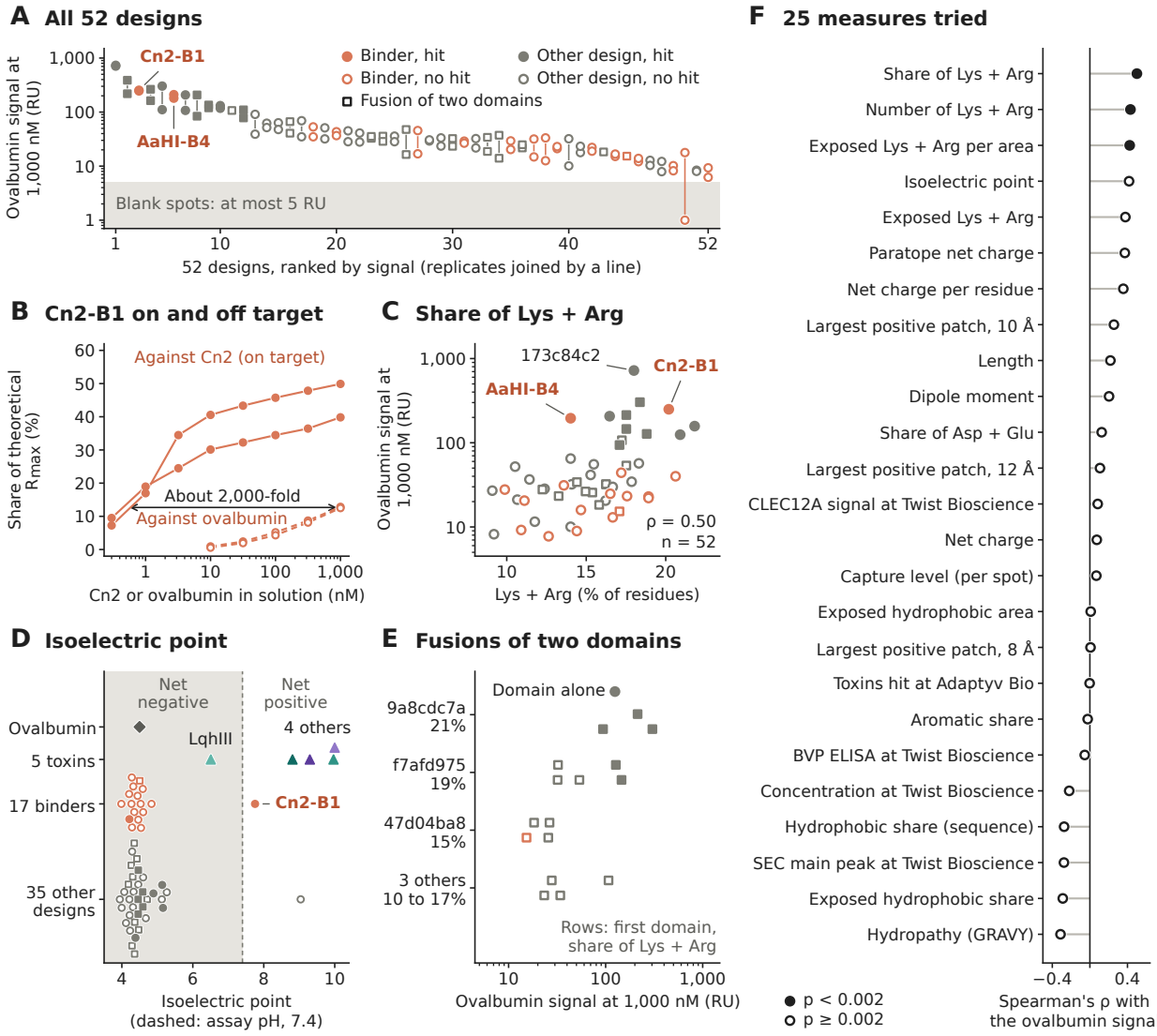


Figure S27. The response against ovalbumin rose with the designs' share of positively charged residues, which fits Cn2-B1 but not AaHI-B4. **A)** Response at 1 µM ovalbumin for the 52 designs, ranked, with replicates joined by a line; gray band, the range of the blank spots (at most 5 RU); the three series whose spots held no design are left out, and one value under 1 RU is drawn at 1. **B)** Cn2-B1's response as a share of the theoretical R_{max} against Cn2 (solid lines, two main-screen replicates) and against ovalbumin (dashed lines, two replicates); the arrow joins equal shares. **C)** Mean response against the share of lysine and arginine in the sequence (Spearman's ρ 0.50, $n = 52$). **D)** Isoelectric points: ovalbumin, as reported [71]; the toxins, calculated from the mature sequences with the cysteines left out; the designs, calculated without tags (EMBOSS pK_a values; with the set of Table S17, Cn2-B1 carries −0.6 at pH 7.4, so it is near neutral); dashed line, pH 7.4, the pH of Adaptyv Bio's running buffer in the main screen. **E)** The 16 fusions tested, grouped by their first domain, with that domain's share of lysine and arginine; the circle is the single domain 9a8cdc7a itself. **F)** Spearman's ρ between the mean response and each of 25 measures; filled, $p < 0.002$ (0.05 divided by 25). Marks in A, C, D, and E: orange, binders; gray, other designs; filled, hits against ovalbumin; squares, fusions. Structure-based measures come from ESMFold2 models of the designs alone, and the paratope from the complex with an intended toxin.

Table S24. Two of the 17 binders and nine of the 35 other designs tested were hits against ovalbumin. SPR at Adaptyv Bio, design captured and ovalbumin in solution. Rows list the 17 binders, then the nine other designs that were hits, by falling response. Design, name for the 17 binders, UUID for the others. Format, single domain or fusion of two domains. Ovalbumin, Adaptyv Bio's result. Response (RU), at the end of the 1 µM injection, cumulative over the five injections, one value per valid replicate. Share (%), mean response as a share of the theoretical R_{max} for one ovalbumin molecule of 45 kDa per captured design, with the capture level read from the loading trace. CLEC12A, result at Twist Bioscience over both runs (hit, signal for a weak or ambiguous signal, or no hit). pI, isoelectric point calculated from the design's sequence without tags. Lys + Arg, % of residues. The 26 other designs tested were not hits (8 to 107 RU; 0.5 to 9.6%).

Design	Format	Ovalbumin	Response (RU)	Share (%)	CLEC12A	pI	Lys + Arg (%)
The 17 binders							
Cn2-B1	single domain	hit	246, 254	12.7	no hit	7.8	20
Cn2-B2	single domain	no hit	29, 27	1.6	no hit	4.0	10
Cn2-B3	single domain	no hit	46, 17	1.3	no hit	4.5	14
Cn2-B4	single domain	no hit	0, 18	0.6	no hit	4.3	14
Cn2-B5	single domain	no hit	10, 8	0.4	no hit	4.2	11
Cn2-B6	single domain	no hit	43, 37	2.2	no hit	4.6	21
Cn2-B7	single domain	no hit	15, 32	1.0	no hit	4.4	18
AaHII-B1	single domain	no hit	6, 9	0.5	no hit	4.3	13
AaHII-B2	single domain	no hit	21, 23	1.0	no hit	4.5	19
AaHII-B3	single domain	no hit	33, 13	1.3	no hit	4.5	19
AaHII-B4	single domain	no hit	29, 20	1.7	no hit	4.3	17
AaHI-B1	single domain	no hit	15, 17	1.0	no hit	4.8	15
AaHI-B2	single domain	no hit	35, 53	2.9	no hit	4.3	17
AaHI-B3	fusion	no hit	15	1.9	no hit	4.5	17
AaHI-B4	single domain	hit	181, 210	11.0	no hit	4.2	14
LqhIII-B1	single domain	no hit	22, 19	1.2	no hit	4.6	11
LqhIII-B2	single domain	no hit	12, 14	0.5	no hit	4.3	17
Other designs that were hits							
173c84c2	single domain	hit	709, 734	41.2	no hit	4.4	18
7491eb22	fusion	hit	388, 218	27.6	hit	4.6	18
443207a9	fusion	hit	264, 163	22.0	hit	4.5	18
1296b21d	single domain	hit	303, 111	10.7	no hit	4.9	16
0510b42f	single domain	hit	208, 107	8.8	no hit	5.2	22
e80c46be	fusion	hit	84, 208	21.0	hit	4.5	18
40090fca	fusion	hit	120, 135	15.0	signal	4.6	19
9a8cdc7a	single domain	hit	118, 132	6.8	no hit	5.1	21
700d103b	fusion	hit	110, 78	13.3	no hit	4.5	17

S5 *In silico* scores and measured binding

We tested whether the campaign's scores could have predicted the binders. The final score tracked how success rates differed between design sets, but within a set it did not pick out the binders (Table S25). Scores and binding fell from the single-toxin to the multi-toxin sets (Figures 4 and 7), and Set 8, the lowest-scoring set against Ts1, AaH I, and LqhIII, had no hits. Within a set and toxin, neither the final score nor the campaign's rank separated binders from the rest, with one exception: the campaign's four top-ranked designs of Set 1 (Cn2) are all binders. ipSAE from ESMFold2, the term shown in Table S25, came out highest of the six terms of the final score: within a set and toxin, the ipSAE terms reached an area under the receiver operating characteristic curve (AUROC) of 0.71, 0.68, and 0.51 (ESMFold2, ESMFold2-Fast, and Protenix v2) and the scDockQ terms 0.56, 0.60, and 0.50. The mean of the three ipSAE terms, which involves no choice among terms, reached 0.70 (0.53 to 0.83). In a paired bootstrap over designs, ipSAE from ESMFold2 exceeded the final score by 0.11 (95% interval −0.07 to 0.32; $p = 0.26$), so 15 binders do not show that any term outperforms the final score. Average precision (AP) led to the same conclusion (Table S25). Within a set and toxin, and in the single-toxin sets, every AP interval includes the chance level. All 60 single-toxin designs scored 0.88 to 0.95 against their toxins, including the 12 made against Ts1, which yielded no binder, and the score did not separate the 11 binders among the 60 from the other 49. The analysis covers the 510 pairs of the 180 campaign designs with their intended toxins for which at least one team returned a usable result. Two of the 17 binders are against LqhIII, which only the Adaptyv Bio team measured, so the first two rows of Table S25 do not count them as binders. Counting them, or leaving out all pairs against LqhIII, changed the AUROC values there by up to 0.02 and the AP values by up to 0.04. The 31 designs added later are left out, as the scores of 28 of them come from a re-run after the campaign.

S5.1 Designs predicted as monomers

A design that folds alone into its designed shape presents a pre-formed binding surface, whereas one that reaches that shape only with the toxin would pay for it in affinity. We predicted every design as a monomer, without toxin, template, or sequence alignment,

with the three predictors and five seeds each. The self-consistency RMSD (scRMSD) is the Cα RMSD to the design's chain in the design model after superposition (median over seeds; for fusions, the larger of the two domains' values), and pLDDT is the mean over seeds. Of the 180 campaign designs, 123 had an scRMSD of 2.0 Å or less with ESMFold2-Fast, 123 with ESMFold2, and 118 with Protenix v2 (Figure S28B). Of the 17 binders, 16, 17, and 14 met this threshold (medians 0.55, 0.63, and 0.70 Å), and all 17 reached a monomer pLDDT of 72 or more (Figure S28A). With ESMFold2, 107 of the 120 designs in Sets 1 to 7 are within 2.0 Å, but only 1 of 30 in Set 8 (median 3.3 Å) and 15 of 30 in Set 9, the campaign's two sets made against all five toxins. Of the 30 Set 8 designs, 28 are variants of one binding mode on one backbone (Section S2.1), so Set 8 is close to a single case. It keeps its fold (median TM-score 0.86, 1.1 Å over the helices alone), so the differences lie in loops and chain ends, and it yielded no binder. Some designs, among them binders, have kinked helices. A kink that the predictor formed only in the complex, to fit the toxin, would leave the free design without the designed surface. Helix kink angles nearly always match between the monomer and the complex from the same predictor (Figure S28C), so the kinks are part of the predicted fold. The campaign selected designs whose predicted complexes matched their design models (scDockQ), and a design model can itself be a complex predicted for an earlier sequence by ESMFold2 or Protenix v2, two of the three monomer predictors (Section S1.1), so part of this agreement is expected.

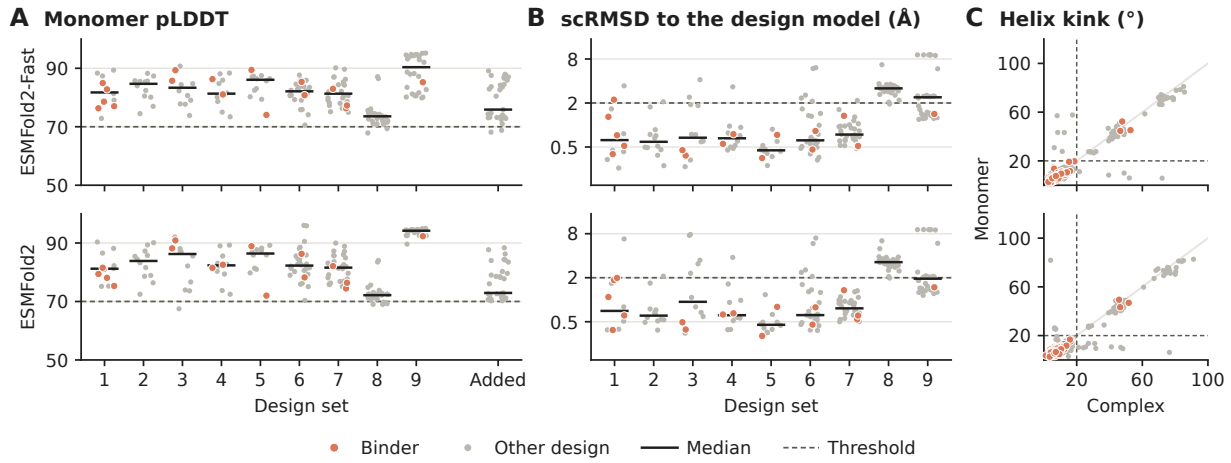


Figure S28. Most designs are predicted to fold on their own into the designed shape; most exceptions are in the campaign's two sets made against all five toxins. Monomers are designs predicted without toxin, template, or sequence alignment, five seeds each, by ESMFold2-Fast (top) and ESMFold2 (bottom). Numbers, the sets of Table S6; Added, Set 10. **A)** Monomer pLDDT, mean of the five seeds; dashed line, the campaign's cutoff of 70. **B)** scRMSD, the Cα RMSD between the monomer and the design's chain in the design model, median of the five seeds (log scale; 180 campaign designs; for fusions, the larger of the two domains' values); dashed line, 2.0 Å. **C)** Kink angles of helices of 12 or more residues in the monomer (seed of highest pLDDT) and in the same predictor's complex; dashed lines, 20°; gray line, equal angles. Orange, the 17 binders; bars, medians.

Table S25. Within a set and toxin, the final score did not separate the binders from the rest. Area under the ROC curve, AUROC (95% interval), of each score for binding, on the measured pairs of the 180 campaign designs with their intended toxins; an interval that includes 0.5 means no separation, and none is corrected for multiple comparisons. Each comparison has a second line with the average precision, AP (95% interval), the mean over the binders of the share of binders among the pairs that score at least as high as that binder. Within a set and toxin, that share is taken among the pairs of the binder's own set and toxin. The value in parentheses after AP is the chance level, the AP expected for a random order, and an AP interval that includes it means no separation. Ties lower the chance level of the rank to 0.035 when pooled and to 0.21 in the single-toxin sets. Binders, the pairs that met the criteria at both labs out of all measured pairs; in the last row, the binders out of all designs. Rank within set, the campaign's recorded rank, as (rank − 1)/(n − 1); in the single-toxin sets, 31 of the 60 designs hold the rank of a design they replaced. Blank, not reported.

Comparison	Binders	Measure	Final score	ipSAE, ESMFold2	Rank within set
Pooled over sets	15/510	AUROC	0.83 (0.72 to 0.91)	0.89 (0.82 to 0.94)	0.58 (0.42 to 0.74)
		AP (0.04)	0.12 (0.06 to 0.23)	0.22 (0.10 to 0.47)	0.04 (0.02 to 0.11)
Within set and toxin	15/510	AUROC	0.60 (0.38 to 0.76)	0.71 (0.55 to 0.84)	0.57 (0.39 to 0.76)
		AP (0.32)	0.32 (0.22 to 0.60)	0.43 (0.31 to 0.71)	0.43 (0.30 to 0.71)
Single-toxin sets, binders versus the other 49	11/60	AUROC	0.49 (0.32 to 0.66)		0.63 (0.43 to 0.82)
		AP (0.23)	0.19 (0.10 to 0.37)		0.26 (0.13 to 0.50)

Table S26. Antibodies with a solved complex scored far above the decoys against Cn2 and narrowly above them against AaH II. ipSAE, best of five seeds, by predictor; the antibody values are those drawn in Figure 4. In parentheses, the PDB entry of the complex. Highest decoy, the highest-scoring of the 28 to 31 decoys and negative controls for that toxin and predictor. Threshold, the ipSAE threshold (fixed in advance) that a design had to reach (Figure S2); with Protenix v2, the highest decoy lies above it for all five toxins.

Toxin and complex	ESMFold2-Fast	ESMFold2	Protenix v2
Cn2			
9004G (2YC1)	0.85	0.82	0.91
LR (4V1D)	0.88	0.83	0.91
RU1 (4V1D)	0.81	0.82	0.84
Voltage-sensor domain II (25IH)	0.00	0.00	0.09
Highest decoy	0.01	0.01	0.70
Threshold	0.15	0.15	0.54
Ts1			
Serrumab	0.34	0.00	0.31
Highest decoy	0.32	0.12	0.31
Threshold	0.34	0.15	0.25
AaH II			
4C1 (4AEI)	0.67	0.49	0.79
Voltage-sensor domain IV (6NT4)	0.00	0.00	0.67
Highest decoy	0.66	0.15	0.76
Threshold	0.67	0.17	0.71
AaH I			
9C2	0.01	0.01	0.89
Highest decoy	0.11	0.02	0.87
Threshold	0.15	0.15	0.72
LqhIII			
Serrumab	0.02	0.00	0.24
Voltage-sensor domain IV (7K18)	0.00	0.00	0.55
Highest decoy	0.02	0.02	0.75
Threshold	0.15	0.15	0.54

S5.2 Co-folded positive controls

The three predictors co-folded each antibody control with its toxin as three chains (the variable domains of the heavy and light chains, VH and VL, and the toxin), to test whether the scores against each toxin separate a known complex from decoys (Table S26). Against Cn2, the scores separate known complexes from decoys: 9004G, LR, and RU1 scored above every decoy with each predictor. Against AaH II, 4C1 scored only narrowly above the highest decoy with two of the three. In all 12 cases (four complexes, three predictors), the best seed matched the crystal pose (DockQ 0.82 to 0.95). These four complexes come from three PDB entries that may have been in the predictors' training data.

Ts1, AaH I, and LqhIII lack such a test, because no complex of serrumab or 9C2 has been solved. Both antibodies scored at most 0.34, close to or below the highest decoy, except 9C2 with Protenix v2, which scored just above it against AaH I (Table S26). In these predictions both partners folded and were in contact. The toxins' own channel domains scored below the highest decoy with each of the three predictors (Table S26).

S6 Published antibodies and nanobodies against the five toxins

Published antibody fragments and nanobodies against the same toxins remain the standard that designed binders have to beat (Table S27). They bind Cn2 and AaH II with sub-nanomolar reported K_D values, and several protected mice in challenge studies with lethal doses of toxin or venom. Against these two toxins, our tightest binders by K_D at Twist Bioscience were Cn2-B1, with K_D upper limits of 3.1 nM at Adaptyv Bio and 1.7 nM at Twist Bioscience, and AaHII-B1, with 78 nM at Adaptyv Bio and 1.7 nM at Twist Bioscience (Table 1). Against AaH I, AaHI-B1 gave 26 nM at Twist Bioscience and about 6.5 nM at Adaptyv Bio, the mean of one SPR and one BLI replicate (9.7 and 3.2 nM). Our designs have not been tested in functional assays or mice, so the comparison covers affinity only.

Table S27. Selected published binders against the five toxins came from immunized animals or antibody libraries, and several protected mice in challenge studies. Two of our controls, 9004G and 4C1, are listed in Table S12. Values are as reported by the cited groups (blank, value not found); methods, routes, and doses differ between studies. AahI′ is a natural variant of AaH I (Val to Ile at position 36 of UniProt P01479). LD_{50}, median lethal dose; scFv, single-chain variable fragment.

Toxin	Binder and origin	K_D (nM)	Neutralization	Reference
Cn2	LR, a human scFv derived from 9004G, which was matured by directed evolution from library clone 3F	0.11	Rescued mice from 3 LD_{50} of *C. noxius* or *C. suffusus* venom (90 to 100% survival)	[42, 43]
Cn2	RU1, a human scFv matured from library clone C1		Neutralized 2 LD_{50} of Cn2 and 3 LD_{50} of *C. noxius* venom	[18]
Cn2	10FG2, a human scFv matured further from RU1; binds toxins of several *Centruroides* species	0.70	Neutralized 4 LD_{50} of *C. noxius* venom premixed; rescued mice from 3 LD_{50}	[21]
Ts1	15e, a human scFv from a phage library		Partly protected mice from 1 LD_{50} of Ts1	[93]
Ts1, LqhIII	Serrumab, a human scFv selected on *T. serrulatus* venom		By two-electrode voltage clamp on Na_V channels, equimolar serrumab reduced the effect of Ts1 by 43%, abolished the small effect of LqhIII on $Na_V1.2$, and halved its larger effect on an insect channel	[40]
AaH II	NbAahII10, a nanobody from a dromedary immunized with AaH II	0.49	Neutralized 7 LD_{50} of AaH II	[19]
AaH I (AahI′)	NbAahI′F12, a nanobody from immunized dromedaries		Protected mice from 100 LD_{50} of AahI′, intracerebroventricular	[20]
AaH I	9C2, an scFv from a murine monoclonal antibody	nanomolar	Protected mice from AaH I, intracerebroventricular	[41]
AaH I, AaH II	NbF12-10, NbAahI′F12 joined to NbAahII10		Neutralized 5 LD_{50} of venom, subcutaneous; 85 µg intravenous rescued all mice after a lethal dose	[20]

S7 Limitations of this study and what would resolve them

Table S28 lists the main limitations of this study and the experiment that would resolve each.

Table S28. Limitations and remedies. Ranked by how strongly each affects our conclusions.

Limitation	What would resolve it
No design was tested for neutralization or for any effect on toxin activity (Section S4)	Electrophysiology, such as patch clamp, on sodium channels with active toxin
No design was tested *in vivo*, so protection from envenomation, serum half-life, and immunogenicity were not measured	Protection of mice from toxin and from whole venom, with pharmacokinetics and immunogenicity
All binding was measured against recombinant toxins with no documented activity, tagged wherever the construct is documented, and never against toxin purified from venom (Section S3.3)	Binding against toxin purified from venom; for each recombinant preparation, activity on sodium channels and the intact mass, to confirm the construct and its four disulfide bonds
Positive controls worked at both labs for only Cn2 and AaH I, and none worked at either lab for Ts1, AaH II, or LqhIII (Table S11; Section S3.4)	For each toxin, a control antibody that binds it in each assay
Each K_D rests on one concentration series at Twist Bioscience and on two to four fitted replicates at Adaptyv Bio, all run without regeneration; the replicates differ 3.0- to 13-fold for the AaH I binders, and competition was measured one to three times per design and antibody (Sections S3.2, S3.6, and S4.1)	Independent replicates from separate protein preparations, in an assay optimized for clean sensorgrams, with flow rate and surface density varied
The four AaH I binders meet the theoretical R_{max} criterion at the toxin mass we derived from Twist Bioscience's values, but not at the lower mass of the mature toxin (Table S10)	The binding stoichiometry of 9C2 and of each of the four binders with AaH I, such as by isothermal titration calorimetry, on a toxin lot of confirmed mass
Competition could be assessed only against Cn2 and, with one antibody of unknown epitope, for three AaH I binders; Adaptyv Bio's competition assay detects only tight binders, and their reverse BLI included no control antibody and left Ts1 unresolved, so the binding sites of the binders against AaH II, AaH I, and LqhIII remain unknown (Sections S4.1 and S3.10)	Competition with two antibodies of known, separate epitopes that bind each toxin in the assay, and reverse BLI with a working control against all five toxins
LR also competed with seven designs that gave weak responses against Cn2 but did not meet the criteria for binding, so some may bind Cn2 very weakly, and calls near the edge of our definition could go either way (Figure S20)	Replicate concentration series of these seven designs against Cn2, extended to higher concentrations, with and without LR
With the toxin on the sensor, all 17 binders were hits at Adaptyv Bio but only 12 of the 15 tested were hits at Twist Bioscience, and, Ts1 aside, 28 design-toxin pairs that we do not label binders were also hits (Section S3.10)	Both orientations for the three discordant binders and the 28 pairs, on shared protein preparations, with monovalent designs in solution
Off-target binding was tested in SPR only against two unrelated proteins, CLEC12A at Twist Bioscience and ovalbumin at Adaptyv Bio (two binders, Cn2-B1 and AaHI-B4, were hits against ovalbumin), and polyreactivity was tested only in one BVP ELISA run per design, which gave high signals (Sections S4.2, S4.3, and S4.4)	A panel of unrelated proteins and related toxins, and repeated BVP ELISA runs at several concentrations
Production cost, thermal stability, and the benefit of selection over unselected single-toxin designs were not measured	Yield and cost of production at scale, the melting temperature of each binder, and the hit rate of unselected single-toxin designs tested alongside the selected ones
We ran one campaign without a baseline, so Claude's contribution cannot be separated from that of its tools, nor its time and cost compared with a conventional campaign	The same prompt given to experts, or a fixed pipeline, with the same tools and resources