toxins

MDPI

Article

Benchmarking Structural Models of α-Conotoxin–Nicotinic Acetylcholine Receptor Complexes: Apo-Conformer Selection, Interface Information, and Deep-Learning Co-Folding

SungYeop Jung [1] and Lee-Seung Kwon [2,*]

[1] School of Pharmacy and Medical Sciences, Griffith Health, Griffith University, Gold Coast, QLD 4222, Australia; sungyeop23@naver.com
[2] Department of Health Care Management, Eulji University, Daejeon 34824, Republic of Korea
* Correspondence: 20251477@eulji.ac.kr

Abstract

Conopeptides are a proven drug source, but translation stalls at one step: predicting how a peptide engages a human receptor before a clinical program commits. We benchmarked modelling of α-conotoxin–nicotinic acetylcholine receptor complexes at two deposited structures, kept separate because their data-leakage status differs: a human α7–ImI cryo-EM complex, rated MEDIUM leakage risk, and a non-human *Aplysia* acetylcholine-binding protein–RgIA crystal that lies inside the Boltz-2 training window and is therefore a positive control, not a held-out test. Apo-conformer choice fixes attainable accuracy before docking begins: the rigid-body floor spans 1.02–4.57 Å across sixty deposited ImI models, and no unbound conformer of either toxin can occupy its own bound site without atomic overlap. In the rigid-body, no-refinement LightDock configuration tested here, no pose among 3144 scored swarms reached 3 Å, and full disclosure of the interface gave a worse best pose than none; that ordering inverts when the receptor is truncated to its extracellular domain, so the sign of the effect belongs to the receptor construct, not the ligand class. Within that configuration, no scoring function ranked a near-native pose first, although five of nine ranked the deposited bound conformer first or within 1 Å when it was inserted. Co-folding (Boltz-2) was CAPRI High at both anchors, neither a fully clean test, and recovered the loop B arginine anchor that no docked pose reached; a temporal control, Boltz-1, run at A2 only, still exceeded every docked result. Confidence did not support selection in the comparisons tested: in neither Boltz-2 nor Chai-1 did it separate the true binder from a shuffled decoy, track the measured species potency gap, or distinguish which of two opposite subunit interfaces the toxin occupied ($p = 0.96$). With five diffusion samples per condition, these comparisons have limited power: they show that confidence failed where it was tested, not how often it fails.

Keywords: α-conotoxin; nicotinic acetylcholine receptor; ion channel; neuropharmacology; peptide drug discovery; molecular docking; co-folding; data leakage

Key Contribution: Deep-learning co-folding reaches CAPRI High accuracy for alpha-conotoxin-nicotinic acetylcholine receptor complexes where the rigid-body, no-refinement docking configuration tested here fails, yet in the comparisons tested its confidence does not separate a true binder from a shuffled decoy, does not track a measured species potency gap, and does not identify which of two opposite subunit interfaces the toxin occupies; the two anchors also differ in leakage status, so their accuracy is not equivalent evidence of generalisation. Accuracy is therefore attainable while selection is not, and the apo-conformer

Check for updates

Received: 19 October 2026
Revised: 9 November 2026
Accepted: 13 November 2026
Published: 16 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

chosen before docking, not the scoring function, sets the floor on what any pose can reach.

1. Introduction

Cone snails have already produced a marketed ion-channel drug. Ziconotide, the synthetic equivalent of ω-conotoxin MVIIA from *Conus magus*, is a 25-residue $Ca_V2.2$ blocker approved for intrathecal use in severe chronic pain, on the strength of a randomised controlled trial in patients with cancer- or AIDS-related pain refractory to other treatment [1]. It remains the only approved conopeptide, and the case that did not reach the clinic is instructive. α-Conotoxin Vc1.1 (ACV1), which targets the α9α10 nicotinic acetylcholine receptor (nAChR) for neuropathic pain, was well tolerated in a Phase I trial and advanced to Phase IIa before development was halted in 2005: subsequent work showed the peptide's IC_{50} at the human α9α10 receptor (320–1000 nM) to be substantially higher than at the rat receptor (19–109 nM), the species in which its analgesic activity had originally been characterised [2]. The chemistry was not the obstacle; predicting how the same peptide would behave at the human receptor subtype was.

ACV1 is not an isolated case of chemical abundance outrunning structural knowledge. The genus *Conus* comprised 761 valid Recent species as of the last comprehensive molecular phylogeny [3], each producing its own largely distinct venom, and that chemical diversity is reflected in the scale of what has been catalogued but not structurally resolved. ConoServer, the standard repository for conopeptide sequences and structures, currently lists 8529 protein entries and only 246 deposited three-dimensional structures—a structural coverage of about 2.9%, essentially unchanged since a 2024 review reported the same 246 structures across 48 species [4–6] (ConoServer accessed 6 August 2026). For the overwhelming majority of conopeptides, a computational model is the only structure that will ever exist, and any attempt to anticipate a peptide's behaviour at a human receptor subtype—the exact question that stopped ACV1—has to run through one.

Modelling conotoxin–nAChR complexes has an established lineage. ToxDock combined Rosetta FastRelax with FlexPepDock to dock α-conotoxin GID onto a homology model of the human α4β2 nAChR extracellular domain, correctly ranking the activity of four designed analogues, and validated the underlying protocol by redocking three experimentally determined complexes—including α-conotoxin PnIA[A10L,D14K] bound to an acetylcholine-binding protein (AChBP), a soluble homologue of the receptor's extracellular domain used as a surrogate—to Cα RMSDs of 0.6–1.3 Å [7]. Its authors were explicit about why a surrogate was necessary: "no crystal structure of a conotoxin and nAChR has been reported to date" [7]. Gulsevin and Meiler later redocked eleven AChBP and α7–AChBP chimera complexes with Rosetta, again modelling the receptor's extracellular domain by homology, and reported top-5 pose accuracies of 1.6 ± 0.2 Å [8]. Both studies report bound-to-bound redocking accuracy in absolute terms, without varying the interface information given to the search or normalising to the conformational change the search was, or was not, required to make.

That last constraint has been addressed on other targets. Introducing binding-site information in graded steps—from no restraints to the complete true interface—moved LightDock's top-10 success rate on 55 unbound protein–protein cases from 14.5% to 92.7% [9], using the same restraint-driven docking engine used here. To our knowledge, the same graded-disclosure design has not been applied to a disulfide-constrained peptide docked against a pentameric membrane receptor.

The missing experimental complex has since appeared. Cryo-EM structures of the human α7 nAChR bound to α-conotoxin ImI, and of the muscle-type receptor bound to

α-conotoxin ImII, were deposited in 2026 [10], alongside an earlier (2022) crystal structure of α-conotoxin RgIA bound to *Aplysia californica* AChBP [11]. For the first time, conotoxin–nAChR docking can be benchmarked against a genuine human receptor complex rather than a homology model or a soluble surrogate, and redocking accuracy can be reported against a rigid-body floor set by the toxin's own apo conformer rather than as an absolute number.

Deep-learning co-folding methods add a further option beyond template-based docking, but they raise a different question for any target deposited recently enough to matter: whether the model could have seen the answer during training. AlphaFold3 and related co-folders are routinely evaluated on structures released after their training cutoff for exactly this reason [12], but no equivalent evaluation exists for conotoxin–nAChR complexes; any co-folding result on this class needs its training cutoff and the target's release date reported side by side, not assumed clean.

Here we benchmark structural modelling of α-conotoxin–nAChR complexes against two deposited structures: the 2026 human α7 nAChR–ImI cryo-EM complex, and the 2022 *Aplysia californica* AChBP–RgIA crystal structure, retained as a non-human comparison point throughout rather than pooled with the human anchor. We ask three questions in sequence. First, how much of the achievable accuracy is fixed before docking begins by the choice of the apo conformer of the free toxin? Second, does supplying the search with interface information—none, a restraint set cross-donated from a different structure, or the target's own native contacts—improve the best pose obtained, and do standard scoring functions reliably identify that pose once it is found? Third, does a deep-learning co-folding model close the gap that docking and rescoring leave open, and does that result hold up against a check on its own training cutoff? Because a computational model is the only structure most conopeptides will ever have (as shown above), the answer to that third question bears directly on whether a marine venom peptide's behaviour at a human receptor subtype can be anticipated computationally before a discovery program commits to it—the step at which ACV1's clinical program was lost. The answers constrain what this benchmark can claim—two anchors and one docking engine are not a general verdict on the method class—but they are, to our knowledge, the first measured directly against an experimentally determined complex of a conotoxin and a human nicotinic receptor.

2. Results

2.1. Two Experimentally Determined Complexes: A Human α7–ImI Anchor and an AChBP–RgIA Surrogate

The benchmark uses two deposited complexes, evaluated separately throughout because their data-leakage status differs (Section 2.4): anchor A1 is the human α7 nicotinic acetylcholine receptor (nAChR) bound to α-conotoxin ImI (PDB 9NX0, cryo-EM, 3.06 Å, released 4 February 2026) [10]; anchor A2 is *Aplysia californica* acetylcholine-binding protein (AChBP), a soluble homologue of the receptor's extracellular domain, bound to α-conotoxin RgIA (PDB 7EGR, X-ray, 2.50 Å, released 6 April 2022) [11] (Figure 1A). At the shared orthosteric (+)/(−) subunit interface, 28 receptor residues at A1 and 33 at A2 lie within 5 Å of the toxin (Figure 1B).

One difference between the two anchors is a property of the deposited coordinates rather than of the biology, and it belongs with every A2 result reported below. Loop C of this receptor family carries a vicinal Cys–Cys disulfide at the tip of the α-conotoxin site. At A1, that bond is present and geometrically normal in all five receptor chains of 9NX0 (SG–SG 2.034–2.035 Å, and all five declared in the depositor's SSBOND records). At A2 it is present in only five of 7EGR's ten receptor chains: the depositor declares it for chains A, D, F, G and J (2.038–2.058 Å) and for no other chain, and the remaining five carry SG–SG separations of 3.630, 4.350, 4.412 and 4.733 Å—too far apart to be bonded—together with one chain

modelled at 1.600 Å, too close for a disulfide. The subunit that supplies the principal (+) face of the A2 interface used throughout this study, chain H, is one of the five without the bond (SG–SG 4.412 Å). A2's docking and co-folding results are therefore measured against a binding site whose loop C tip is, in the deposited model, in a different covalent state from the same loop at A1. Nothing in this study turns on that bond directly, and we make no claim about which state is correct in solution; it is reported because loop C geometry enters the conformational and contact analyses below (Sections 2.5, 5.9 and 5.11), and a reader comparing the two anchors should know the two are not matched in this respect.

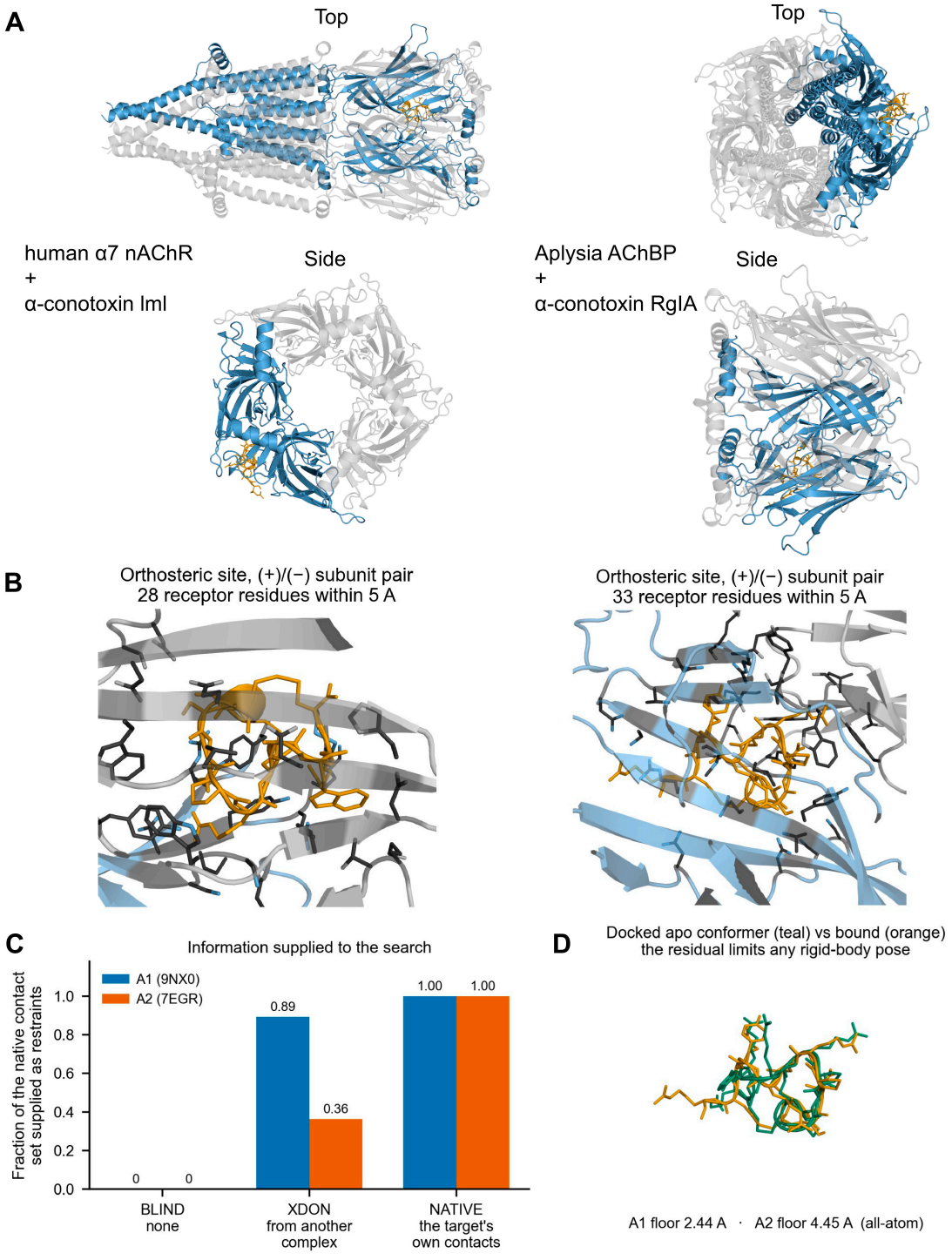


Figure 1. The two benchmark anchor complexes. (**A**) Top and side views of A1 (PDB 9NX0, human α7 nAChR + α-conotoxin ImI, cryo-EM, 3.06 Å) and A2 (PDB 7EGR, *Aplysia* AChBP + α-conotoxin RgIA,

X-ray, 2.50 Å) as cartoons: the docked (+)/(−) subunit pair blue, the remaining subunits grey, the toxin orange. (**B**) The orthosteric site of that subunit pair, with every receptor residue within 5 Å of the toxin drawn as dark sticks; the panel title gives the residue count. (**C**) Fraction of the anchor's native contact set supplied to the search as restraints at each rung of the disclosure ladder—BLIND (none), XDON (contacts donated from a different complex), NATIVE (the target's own contacts)—with A1 in blue and A2 in vermillion. (**D**) The docked apo NMR conformer (teal) superposed on the bound conformer of the same toxin (orange); the all-atom RMSD remaining after this optimal superposition is the rigid-body conformational floor, printed below for both anchors.

Docking was performed with LightDock [13], a restraint-driven, GSO-based multi-scale docking engine, using three levels of interface information in turn: BLIND (no restraints), XDON (a restraint set cross-donated from a different deposited structure—2BYP for A1, 6HY7 for A2; Table S4), and NATIVE (the target's own native contacts, giving 100% recall of the target's footprint by construction). By design, the cross-donated set does not recover the target's own footprint equally well at both anchors: it recovers 89.3% of A1's native contacts but only 36.4% of A2's (Table S4), a restraint-quality asymmetry between anchors that recurs in the docking results below (Figure 1C).

Neither the toxin's docked conformer nor the search itself can outrun a rigid-body limit set before docking begins. The apo (unbound) conformer of each toxin actually used in these runs—1IMI model 01 for ImI, 2JUT model 01 for RgIA—sits 2.439 Å (A1) and 4.453 Å (A2) from its own bound conformer by Kabsch superposition, an intrinsic all-atom conformational floor that no rigid-body search can close (Figure 1D). Section 2.2 examines how much this floor depends on which deposited apo entry is chosen, and how docking is performed against it.

2.2. The Toxin's Free-Solution Conformer Sets the Accuracy Limit; Describing the Binding Site Helps or Hurts Depending on the Receptor Construct

The 2.439 Å and 4.453 Å floors used above are not fixed properties of the toxins; they depend on which deposited apo NMR model is chosen to represent the free peptide. Across sixty deposited NMR models of ImI in three PDB entries, the rigid-body floor against 9NX0's bound conformer ranges from 1.022 Å (entry 1G2G, model 24) to 4.574 Å (entry 1CNL, model 06)—a spread wider than the 3 Å accuracy threshold used throughout this study (Figure 2D; Table S6; Figure S4A). The entry actually used for docking, 1IMI, spans a narrower 1.760–2.735 Å (median 2.19 Å) across its own twenty models. A bound-conformer control—the second, independently modelled copy of ImI already present within 9NX0 itself—differs from its counterpart by only 0.001 Å, confirming that the floor reflects a genuine property of the free-toxin conformer rather than an artifact of atom correspondence. The equivalent census for RgIA's 2JUT ensemble against 7EGR is given in Figure S4B. Because the floor moves by more than the 3 Å threshold itself depending on which entry is used, the apo conformer for a rigid-body benchmark must be fixed by a rule stated in advance of docking, not selected afterward by proximity to the answer.

Against these floors, docking accuracy at neither anchor came close to the 3 Å threshold on either basis, across any of the 3144 scored docking swarms (2 anchors × 3 rungs × 2 replicates; Figure 2A,B). On the all-atom basis, the best pose found across all swarms was 7.647 Å at A1 and 5.990 Å at A2—at A2 the all-atom threshold was unreachable by construction, since even the toxin's own apo conformer sits 4.453 Å from its bound pose (Section 2.1), but at A1 the floor (2.439 Å) left the threshold structurally attainable, and no swarm reached it regardless. On the backbone basis, where both anchors' floors (1.056 Å at A1, 2.397 Å at A2) left the 3 Å threshold reachable in principle, the closest any of the 3144 swarms came was 6.323 Å at A1 and 4.585 Å at A2 (per-run distributions in Figure S3; equal-swarm-count bootstrap correction for A1's larger per-run swarm count, 390–392

versus A2's 132–134, in Table S2; rank-sum tests in Table S3). LightDock's normal-mode option (ten receptor and ten ligand modes, default range) did not change this: the best A1 blind pose was 8.43 Å from the deposited ligand, against 8.74 Å for the rigid run, with no pose within 7 Å. The deformation range allowed is small, and the RMSD was computed without superposing the deformed receptor.

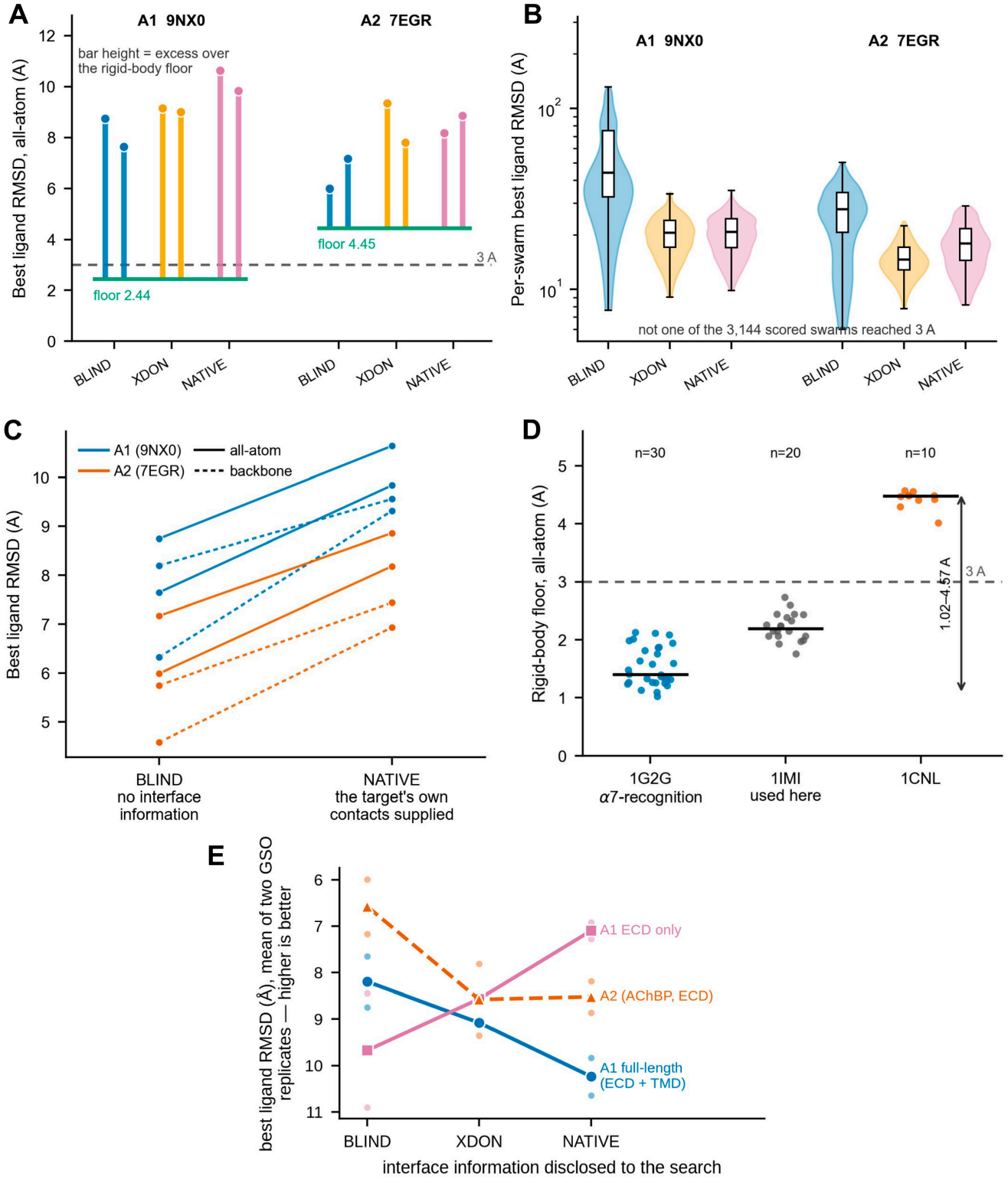


Figure 2. Best ligand RMSD across the three-rung disclosure ladder. Rung colours are consistent throughout: BLIND blue, XDON yellow, NATIVE pink. (**A**) Best all-atom ligand RMSD of each of

the 12 runs (2 anchors × 3 rungs × 2 GSO replicates), drawn as lollipops whose stems start at that anchor's rigid-body floor (green line), so stem length is the excess over the floor; the dashed line is the 3 Å criterion. (**B**) Distribution of the per-swarm-best RMSD within each rung, both replicates pooled, as violins with overlaid box plots on a log scale. (**C**) Paired BLIND-versus-NATIVE comparison of the best pose, A1 blue and A2 vermillion, solid lines all-atom and dashed lines backbone-only. (**D**) Rigid-body floor of every deposited apo NMR model of ImI, grouped by PDB entry with *n* above each group; horizontal bars are per-entry medians, the dashed line the 3 Å criterion, and the arrow the full range. (**E**) Ladder means for three receptor constructs—A1 full-length (blue circles), A1 extracellular domain only (pink squares) and A2 (orange triangles); large markers are the mean of two GSO replicates and the faint small points are the replicates themselves. The y-axis is inverted here, so higher is better.

Supplying the search with more interface information did not help at either anchor as constructed. Comparing the BLIND and NATIVE rungs directly, at both anchors and both search replicates, the fully informed NATIVE run produced a worse best pose than the completely uninformed BLIND run in every one of the 4 independent anchor × replicate comparisons, by a margin of 1.36–2.99 Å on the all-atom basis; the paired backbone comparison agrees in direction but is a second measurement of the same poses, not a fifth through eighth independent comparison (Figure 2C). Because each reported best pose is a minimum over an unequal number of swarms per anchor (Section 5.6), the more defensible evidence for this reversal is the equal-*N* bootstrap-of-minimum comparison in Table S2, which resamples both anchors to a common swarm count and reaches the same conclusion; the Mann–Whitney rank-sum test on the raw per-swarm-best distribution (Table S3) is reported only as a secondary check, since it treats swarms—a tiling of one receptor surface—as independent samples, which they are not.

The direction of that effect, however, is not a property of the ligand class. It reverses when the receptor is truncated. Repeating the full ladder on anchor A1 with the transmembrane domain removed—the same toxin, the same apo conformer, the same donor restraint set, the same search parameters, and a receptor of 3145 heavy atoms against A2's 3276 (Section 5.10)—inverts the ordering, and does so monotonically. Mean best all-atom RMSD across the two replicates runs 9.67 → 8.57 → 7.10 Å from BLIND through XDON to NATIVE on the extracellular-domain construct, against 8.20 → 9.08 → 10.24 Å on the same anchor at full length and 6.58 → 8.58 → 8.52 Å at A2. Each ladder is monotone within itself and consistent between replicates—the ECD XDON runs agree to 0.12 Å (8.624 and 8.507 Å) and the ECD NATIVE runs to 0.36 Å (7.278 and 6.922 Å)—so the inversion is not replicate scatter, and it runs against the order statistic rather than with it, since the winning ECD NATIVE runs carry fewer swarms (336) than the ECD BLIND runs they beat (390). Nor is it simply that truncated receptors behave one way and full-length receptors the other: A2 is itself an extracellular-domain-only construct and follows the full-length A1 direction, not the truncated A1 one.

What the ladder measures is therefore construct-conditional. On two of the three receptor constructs tested, disclosing the interface degraded the best pose; on the third, it improved it, monotonically and reproducibly. The practical implication for benchmark design is stronger than a flat negative would have been: a graded-disclosure ladder run on a single receptor construct can return either sign, so the sign obtained is not by itself evidence about the ligand class, and a study using this design should report the construct alongside the result. The mechanism behind the sign change is not established here; Section 2.6 shows that A1's shortfall at full length is search-and-scoring limited rather than conformer limited, and Section 5.10 records that the extracellular-domain construct brings the nearest swarm centre from 7.45 Å to 5.34 Å of a native ligand atom without bringing the native

centroid inside any swarm sphere, which A2 alone achieves at 0.69 Å—but connecting those observations to the sign of the disclosure effect would require more anchors than two.

2.3. The Poses Ranked First Reproduce None of the Native Interactions, and the Ranking Is Not Robust to a Non-Physical Term

LightDock's own scoring functions do no better at picking out these poses than the docking search does at finding them. Nine of LightDock 0.9.4's ten protein–protein scoring functions were applied to a fixed subset of poses drawn identically from each of the twelve runs (2742–2961 poses per run, 108 run × function combinations in total; Tables S7 and S8); the tenth, SIPPER, could not be evaluated because lightdock/scoring/sipper/driver.py references DEFAULT_CONTACT_RESTRAINTS_CUTOFF without importing it, and failed on all twelve runs with the same traceback. The scored subset was deliberately constructed to include each run's own best-RMSD pose—confirmed to equal the docking search's own best pose exactly, in all twelve runs—making the ranking task strictly easier than the original search; none of the nine functions nevertheless ranked a near-native pose first. The best top-1-by-score ligand RMSD achieved anywhere in the panel was 9.894 Å (DFIRE and fastDFIRE, on run A2_7EGR_BLIND_rep1), even though the same scored set contained poses of 5.990 Å (A2) and 7.647 Å (A1) (Figure 3B; the full 9 × 12 grid of rank correlations is given in Figure 3A and, on the full re-scored subset rather than a random 2000-pose sample, in Figure S1; funnels for the seven functions not shown in Figure 3 are given in Figure S2).

Because LightDock's score is maximised, a function that discriminates near-native poses should show a negative Spearman correlation between score and RMSD. On a seeded random subsample of 2000 poses per run, the median correlation was negative for five of the nine functions (DFIRE −0.202, fastDFIRE −0.203, vdW −0.077, PISA −0.074, MJ3h −0.070) and positive for four (cpydock +0.329, DFIRE2 +0.176, TOBI +0.096, SD +0.043), with no consistent relationship to top-1 accuracy: vdW—the function LightDock's own restrained search actually uses internally—had one of the more favourable median correlations (−0.077, third-strongest of the five negative functions) alongside one of the worst median top-1 RMSDs in the panel (25.06 Å; Figure 3C). DFIRE and fastDFIRE, two implementations of the same potential, agreed to three decimal places on both statistics, serving as an internal consistency check on the panel.

The ranking that LightDock does report is also not robust to a specific non-physical term in its own restrained-run objective. Re-scoring all twelve runs with that term forced off changed the top-ranked pose in every one of the eight runs that had carried restraints; the restraint-free winner's rank under the originally published, restraint-aware score was 10, 11, 11, 33, 68, 152, 156 and 287—never rank 1 (Figure 3D). This instability was not predicted by how much of the target's native footprint the restraint set recovered: among the four runs whose restraints recovered 100% of that footprint, removing the restraint term shifted top-1 RMSD by +0.219 Å and −0.092 Å in two runs and by +16.307 Å and +16.338 Å in the other two. Both large shifts occurred at anchor A2's NATIVE rung (+16.307 Å in replicate 2, +16.338 Å in replicate 1), and their near-identical magnitude across two independent GSO search seeds makes the effect a reproducible property of that rung rather than a single-seed accident; the next-largest shifts in the panel fall outside this 100%-recall subset, at A1's XDON rung (+10.397 Å and +9.175 Å, footprint recall 0.893), and likewise reproduced across both seeds (Table S9). In the four unrestrained (BLIND) runs, where this term should have had no effect at all, the reported winner still moved off rank 1 in three of the four (ranks 2, 13 and 18)—consistent with a separate, independently confirmed off-by-one defect in how LightDock 0.9.4 writes its per-step output (Table S11)—though with a negligible RMSD consequence in all four runs (−0.083 to +0.294 Å).

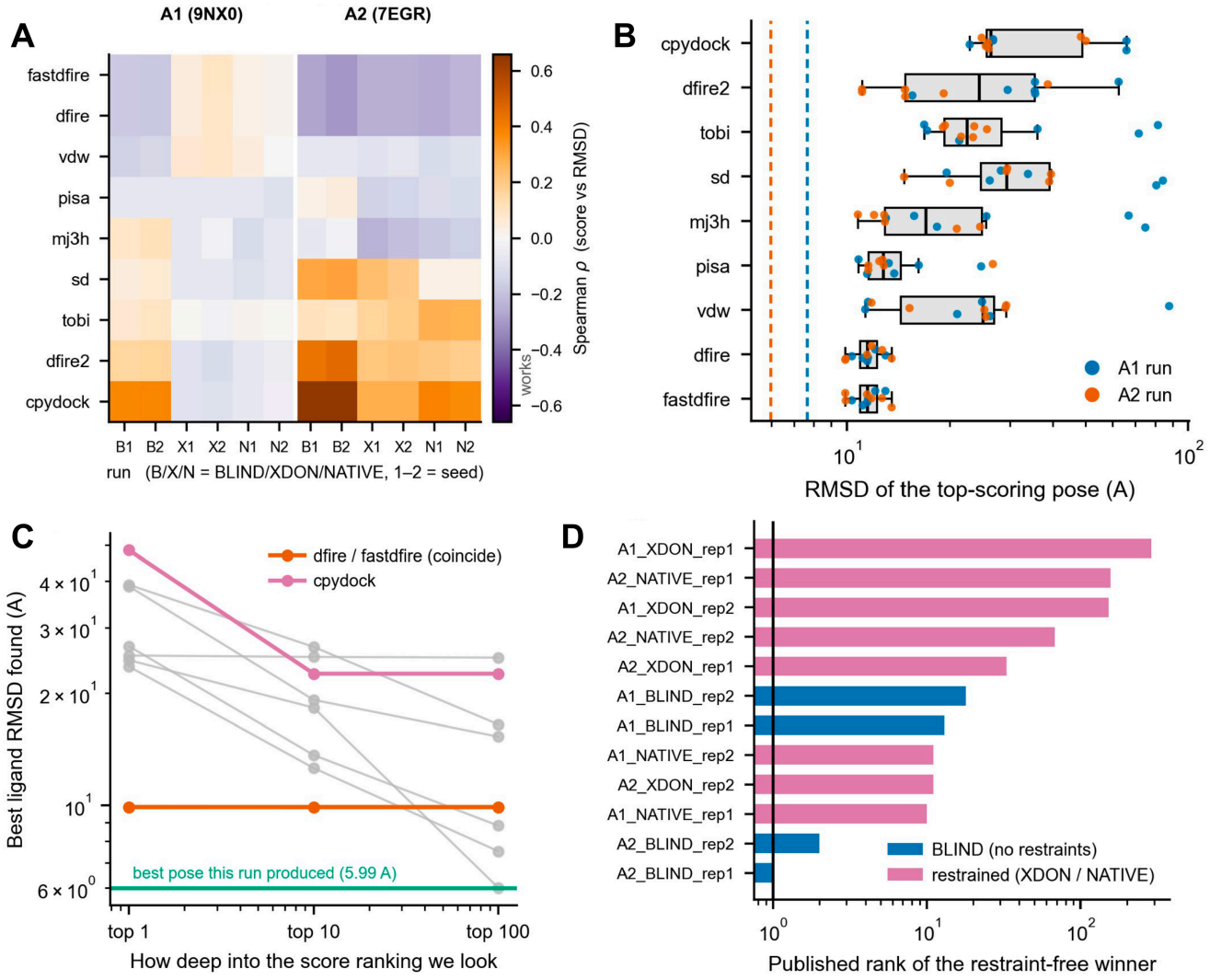


Figure 3. Score–RMSD rank correlation and rank-1 pose selection across nine LightDock scoring functions. (**A**) Spearman's ρ between score and ligand RMSD for 9 functions × 12 runs, on an unbiased random subsample of poses per run. Columns are runs, labelled B/X/N for BLIND/XDON/NATIVE and 1–2 for the GSO seed, grouped by anchor. LightDock maximises its score, so a function that works gives negative ρ; the diverging colour bar runs purple for negative through white to orange for positive and is annotated accordingly. (**B**) Ligand RMSD of the pose each function ranked first, all 12 runs as points over a box plot on a log axis, coloured by anchor (A1 blue, A2 vermillion); the two dashed vertical lines mark the best pose present in each anchor's scored set. (**C**) Best ligand RMSD recoverable as a function of how deep into the score ranking one looks (top 1, top 10, top 100) for a single run: DFIRE and fastDFIRE, which coincide, in orange; CPyDock in pink; the other seven functions in grey; the green line marks the best pose that run produced. (**D**) Rank given to the restraint-free winning pose by the originally published score, one bar per run on a log axis, coloured by whether that run was restraint-free (BLIND, blue) or restrained (XDON or NATIVE, pink).

A positive control places a known-good pose among the rescored poses. The deposited bound conformer of each toxin was inserted as pose 0, together with 300 rigid perturbations of it in six tiers of 50, and all were rescored with the same nine functions (Table S24A). At A1, vdw and SD ranked the deposited pose first and, at A2, SD did; for five of the nine functions (vdw, DFIRE, fastDFIRE, PISA and SD), the highest-scoring pose lay within 1 Å of the deposited pose at both anchors. The other four (DFIRE2, cpydock, MJ3h and TOBI) ranked a pose 4.09–26.68 Å from the deposited pose first. The rescoring pipeline therefore recognises a correct pose, and the failure of the nine functions on the poses the

search produced is not an artefact of pose reconstruction. The decoys in this control are perturbations of the deposited pose and are easier than the poses a search generates. The same control with the rigid fit of the apo conformer in place of the deposited conformer ranked that pose 22nd to 277th of 301 at A1 and 55th to 254th at A2 (Table S24B). The apo conformer overlaps the receptor (10 intermolecular heavy-atom pairs below 2.0 Å at A1, against none for the deposited ligand; Section 5.9), which is consistent with the interpretation in Section 3 that the functions rank degrees of overlap.

2.4. Co-Folding Places the Toxin in the Orthosteric Site at Both Receptors, with a Leakage Caveat at One

A deep-learning co-folding model, Boltz-2 [14]—given only the receptor and toxin sequences, no structural templates, and no restraints—placed the toxin substantially closer to its native pose than either the docking search or any of the nine rescoring functions did at either anchor (Figure 4A; per-sample results in Table S13). Across five diffusion samples at A1 (9NX0), the interface-superposed ligand all-atom RMSD ranged from 1.109 to 1.433 Å (median 1.290 Å, fnat 0.792–0.887); across five samples at A2 (7EGR), 1.57–3.04 Å (median 1.92 Å, fnat 0.84–0.89) (Figure 4B; Table S13). The rigid-body conformational floor used throughout Section 2.2 does not constrain these results, because a co-folding model predicts the ligand's conformation along with its pose rather than docking a fixed rigid conformer; the two kinds of methods are reported side by side with this difference stated explicitly, not on a common scale. Ranking the five samples within each block by Boltz's own reported confidence score did not reliably track ligand RMSD: the confidence–RMSD Spearman correlation was $\rho = 0.70$ at A1 · Boltz-2 (higher confidence paired with worse, not better, RMSD), $\rho = -0.30$ at A2 · Boltz-2, and $\rho = 0.60$ at A2 · Boltz-1 ($n = 5$ per block; none reach significance at this sample size, $p > 0.18$ throughout; Table S13). The sign is in the useful direction in only one of the three blocks, which is why every sample is reported individually here rather than only the single highest-confidence model. The result does not depend on the sampling settings tested, with two qualifications (Table S27). At A1, all 25 samples drawn from five seeds lay within 2.33 Å of the deposited ligand, but raising the recycling steps from three to six moved the median ligand RMSD from 1.37 to 1.99 Å and the median fnat from 0.85 to 0.77, so the reported accuracy depends on that setting by about 0.6 Å. At A2, the five samples reported here (1.57–3.04 Å) were reproduced exactly by a rerun, but 3 of the 35 distinct A2 samples drawn in the sensitivity analysis placed the toxin at a different site (ligand RMSD 36.0–36.6 Å, fnat 0); each was the lowest-ranked sample of its run by Boltz's confidence score. Statements in this paper that co-folding was accurate in every sample refer to the five samples reported.

The two anchors are not equally trustworthy as a test of Boltz-2's ability to generalise and are reported separately for that reason. A2 (7EGR) was released 6 April 2022, before Boltz-2's stated training cutoff of 1 June 2023, placing it inside the training window; its result is reported here as a positive control rather than a held-out test (Table S14). A1 (9NX0) was released 4 February 2026, about 32 months after the cutoff, but a pre-cutoff structure, 6HY7 (22 May 2019), exceeds 40% sequence identity to 9NX0 on both the receptor and toxin chains simultaneously—the same-complex homology criterion AlphaFold3 uses to exclude a test interface from evaluation [12]—so A1's leakage risk is rated MEDIUM rather than LOW (Table S15).

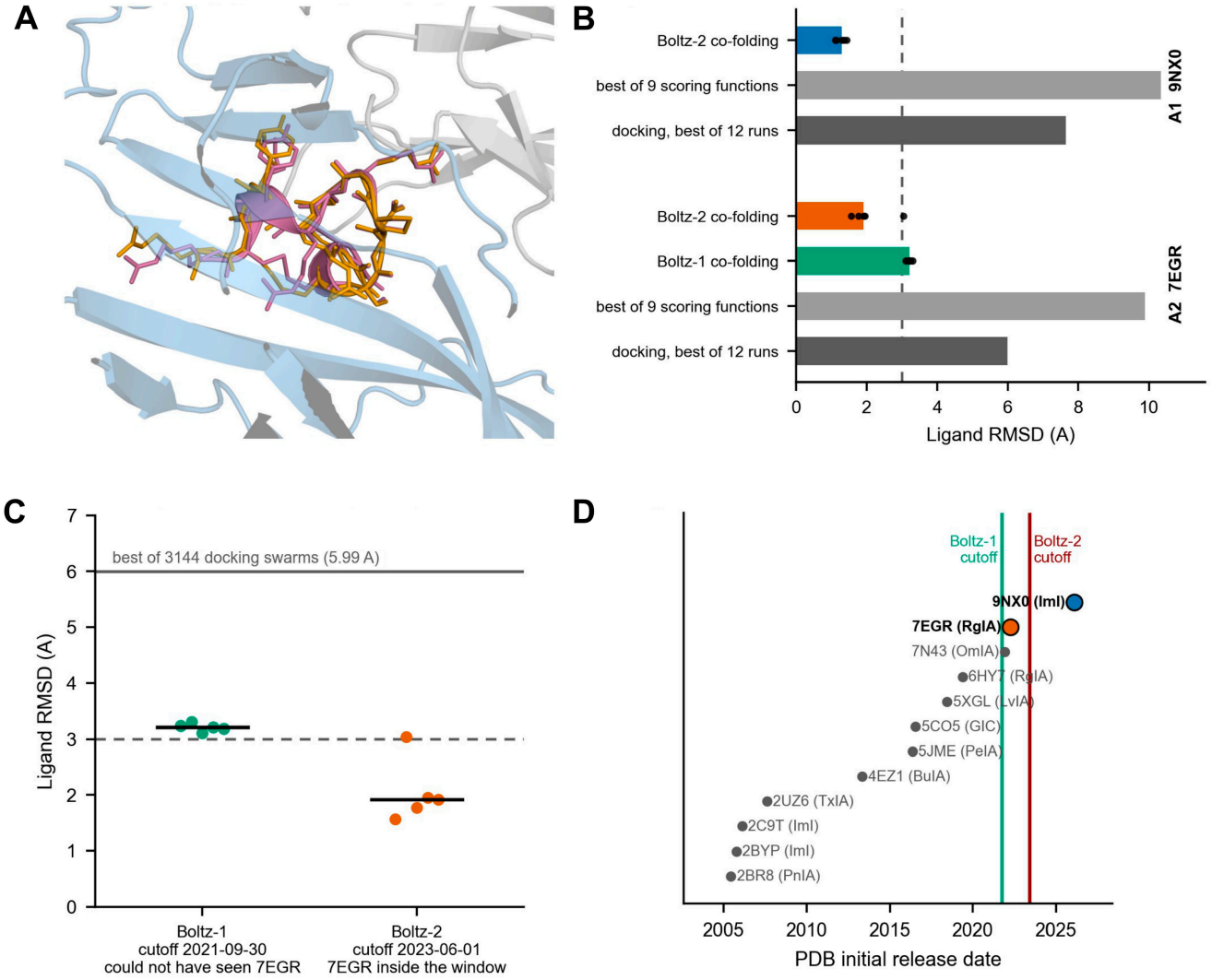


Figure 4. Co-folding compared with docking and rescoring, with a temporal control. (**A**) The Boltz-2 predicted toxin pose (pink) and the deposited toxin (orange) at A2 as sticks, after superposition on the receptor; the (+) and (−) subunits are light blue and grey cartoons. (**B**) Best ligand RMSD by method, grouped by anchor: Boltz-2 (A1 blue, A2 vermillion; $n = 5$ samples, shown individually as black points), Boltz-1 at A2 (green, $n = 5$), the best of nine rescoring functions (light grey) and the best of 12 docking runs (dark grey); the dashed vertical line is the 3 Å criterion. (**C**) The two co-folding models at A2 as individual samples with a bar at each mean, the axis labels giving each model's training cutoff and whether 7EGR fell inside it; the dashed line is 3 Å and the solid grey line the best pose found across all 3144 docking swarms. (**D**) PDB initial release date of every conotoxin–receptor complex used or considered, with the two anchors highlighted in colour and the rest in grey; the vertical lines are the training cutoffs, Boltz-1 in green and Boltz-2 in dark red.

Separately from this leakage-risk assessment, neither input file supplied a structural template, and the paired multiple-sequence alignment returned for either toxin contained no aligned homologues (a depth of three sequences, all copies of the query itself), against receptor MSA depths of 14,106 (9NX0) and 6193 (7EGR) sequences (Table S12). This rules out one specific mechanism—an inter-chain co-evolutionary signal already present in the input alignment—as the source of the predicted interface. It does not by itself reduce the leakage or memorisation concern assessed above: if the predicted interface is not read from a co-evolutionary signal in the alignment, the more likely alternative is that it is read from structural priors acquired during training, which is a route to leakage this MSA-depth observation does not close. If anything, the absence of a co-evolutionary alternative

explanation makes the training-window and homology audit above more load-bearing, not less. To ask where the predicted interface comes from, the A1 prediction was repeated with the receptor alignment, the toxin alignment, both, or neither removed (same seed and settings, five samples each; Table S26). Removing the toxin alignment, which contains only the query sequence, left the prediction essentially unchanged: ligand RMSD 1.104–1.885 Å against 1.133–1.874 Å with it, and ipTM 0.698–0.738 against 0.702–0.727. Removing the receptor alignment destroyed the prediction, with or without the toxin alignment: ligand RMSD 12.6–37.5 Å and fnat 0–0.245, and the two receptor-alignment-free conditions gave byte-identical structures. The ipTM of one of those wrong poses (0.826) was higher than that of any sample with the alignment (maximum 0.727). The toxin alignment therefore carries no usable information, and the interface prediction depends on the receptor alignment. These runs do not separate a contribution of the receptor alignment's co-evolutionary content from a contribution of the receptor structure it helps the model build (whole-receptor Cα RMSD 24.0–26.5 Å without it, 7.2–7.5 Å with it), and they do not test the peptide's own sequence and disulfide framework in isolation.

A temporal control separates two possible readings of the A2 result—generalisation from homologous training data, versus memorisation of 7EGR itself. Boltz-1 [15], whose training cutoff (30 September 2021) predates 7EGR's release (6 April 2022) and which therefore could not have seen this structure during training, still placed the toxin in the correct pocket: across five samples with otherwise identical inference settings to Boltz-2 (five diffusion samples, 200 sampling steps, three recycling steps, no templates), Boltz-1 achieved 3.107–3.301 Å interface-superposed ligand RMSD (fnat 0.679–0.768)—worse than Boltz-2, as expected of an earlier model version, but better in all five samples than the best pose found among the 3144 scored docking swarms (5.990 Å) reported in Section 2.2 (Figure 4C). The gap between Boltz-1 and Boltz-2 reflects both models' different training data and different architecture and scale together, and is not by itself a quantitative measure of how much 7EGR's presence in Boltz-2's training window contributed to its result; what the comparison does establish is that a model provably unable to have seen 7EGR already reaches an accuracy neither the docking search nor rescoring approached at either anchor. The full release-date timeline of every conotoxin–nAChR-family complex used or considered in this study, relative to both training cutoffs, is given in Figure 4D and Table S14; residue-level detail of the predicted interfaces at both anchors is given in Figure S6, and the swarm-level docking geometry underlying Section 2.2's docking result is given in Figure S5.

A Boltz-1 control at A1 could not be run on the full-length construct: without the memory-efficient kernel package, which cannot be installed here without replacing the PyTorch (version 2.4.1+cu124.) build that produced the reported results, the 796-residue input exhausted GPU memory and returned no prediction. We therefore compared Boltz-1 and Boltz-2 on an extracellular-domain construct of A1 (206 residues per chain; same sequence inputs, alignments, seed and sampling settings, five samples each; Table S25). Boltz-2 placed the toxin at 1.20–1.94 Å ligand RMSD in all five samples, and Boltz-1 at 3.05–3.59 Å in four and at a different site in the fifth (34.3 Å). This construct differs from the full-length construct of the main analysis and from the truncated receptor of Section 5.10, so these values are compared only with each other. The control does not remove the MEDIUM rating of A1: the 2019 structure 6HY7 that exceeds 40% identity to 9NX0 predates the Boltz-1 training cutoff (30 September 2021) as well as that of Boltz-2.

2.5. What Each Method Reproduces at the Interface: The Loop B Arginine Anchor

Ligand RMSD and fnat summarise how close a model is, not what it gets right. Because the α-conotoxin site is a defined set of loops, the comparison can be made contact by contact (Figure 5; Section 5.7). Residue numbering below is that of the deposited coordinates, which

for 9NX0 runs one lower than the mature-protein numbering commonly used in the α7 literature (deposited Tyr92, Trp148, Cys189–Cys190, Glu192 and Tyr194 correspond to Tyr93, Trp149, Cys190–Cys191, Glu193 and Tyr195 in that convention); deposited numbering is used throughout for internal consistency with the coordinate files.

At A1 (9NX0, human α7–ImI), the deposited complex makes 53 receptor–ligand residue–residue contacts at a 5.0 Å heavy-atom cutoff. Nine of these were reproduced by at least four of the five Boltz-2 samples and by none of the eighteen docked poses examined (the best-RMSD pose, the top-1-by-score pose, and the best-fnat pose of each of the six A1 runs): seven are loop B contacts in which Ser147, Trp148, Ser149, Tyr150 and Gly151 engage ImI Arg7 and Pro6, and two are loop E contacts of Leu118 with ImI Pro6 and Cys3. Restricting the comparison to the headline BLIND replicate widens the set of contacts made by co-folding and missed by docking to 32 of 53, adding loop A Tyr92–Arg7, the loop C contact Tyr187–Asp5 that is the closest approach in the interface, the single receptor–ligand ion pair Glu192–Arg11, and the complementary-face pocket that receives ImI Trp10 (Trp54, Lys75, Thr76, Leu108, Asn110). Every one of these under-reproduced contacts involves a residue independently identified as functionally essential by mutagenesis: Asp5, Pro6 and Arg7 form the loop 1 triad in which conservative substitution of any one residue lowers affinity at an α7/5-HT3 chimeric receptor by two to three orders of magnitude, and Trp10 is the loop 2 residue for which an aromatic side chain is specifically required [16]—the contact set co-folding recovers and docking misses at every disclosure level is accordingly the set mutagenesis had already marked as the one that matters, not an arbitrary subset of the interface.

Aggregated over the full scored ensembles, no docked pose approached the co-folding result at either anchor. Across all 468,800 scored poses of the six A1 runs the highest fnat reached by any pose was 0.245 (13 of 53 contacts), and across all 160,000 poses at A2 it was 0.554 (31 of 56), against 0.792–0.887 (A1) and 0.839–0.893 (A2) for the five Boltz-2 samples and 0.679–0.768 for the five Boltz-1 temporal-control samples at A2. Those docking figures are oracle maxima over an entire ensemble rather than what a user obtains: the pose each run ranked first reached fnat 0.000–0.208 at A1 and 0.000–0.125 at A2, 88% (A1) and 70% (A2) of all scored poses reproduce no native contact whatsoever, and the median pose reproduces none at either anchor. The A1 maximum of 0.245 was attained by exactly one pose out of 468,800.

The failure is therefore not a shifted interface but, for the pose a user would actually receive, a different site altogether. The top-ranked pose of A1 BLIND replicate 2 sits 87.99 Å from the native ligand, on the transmembrane domain, sharing no contact residue with the native interface. Two qualifications belong with this comparison. First, the direction is not uniform: six A1 contacts were missed by all five Boltz-2 samples, and at A2 the Thr53–Arg409 hydrogen bond—the third-closest approach in that interface at 2.58 Å—was missed by every Boltz-2 sample while every Boltz-1 sample and the best-RMSD poses of two docking rungs reproduced it, so at the leakage-exposed anchor the temporal control recovers a contact the later model does not; these are pooled observations across poses, however, and no single docked pose exceeded fnat 0.245 (A1) or 0.554 (A2). Second, interaction-type counts are reported interaction by interaction in Tables S18 and S19 with their geometric criteria stated, and hydrogen-bond counts are given both with and without those whose acceptor is a disulfide-bonded cysteine sulfur—four of the thirteen at A1 and three of the nineteen at A2—since these are geometrically identified but chemically questionable.

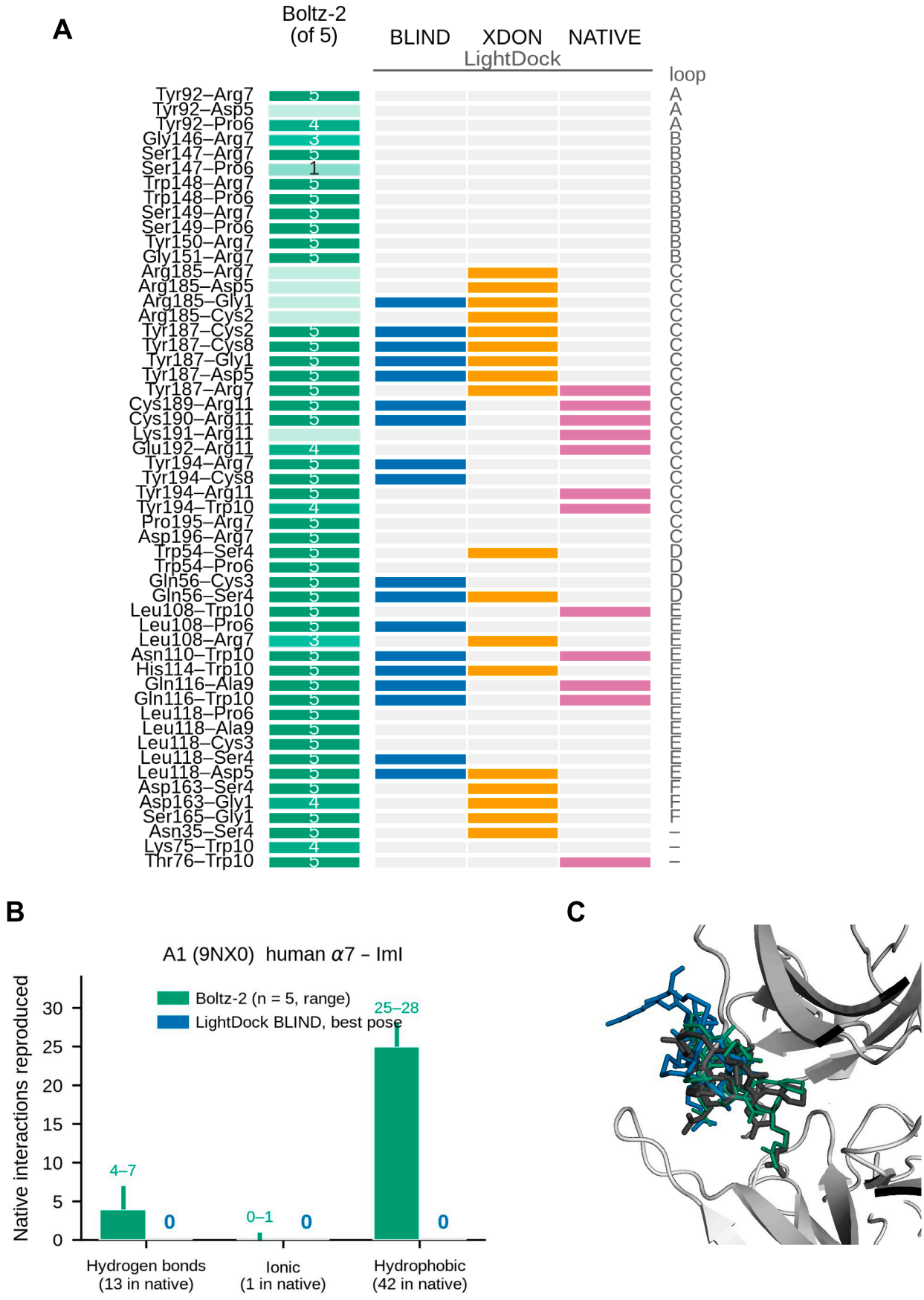


Figure 5. Interface reproduction at anchor A1 (PDB 9NX0, human α7 nAChR + ImI). (**A**) Every native receptor–ligand residue contact (5.0 Å heavy-atom cutoff), one per row, ordered by binding-site loop,

which is named in the right-hand column. The green column counts how many of the five Boltz-2 samples reproduced each contact and is shaded by that count; the three columns to its right record whether any reported pose of the BLIND, XDON, or NATIVE docking rung did, coloured by rung where yes and left pale grey where no. Residue numbering follows the deposited coordinates. (**B**) The same comparison by interaction type: Boltz-2 in green with the range over its five samples as an error bar, and the best-RMSD blind docking pose in blue; native totals are given in the axis labels. (**C**) The orthosteric site with the deposited ImI (black), the best Boltz-2 sample (green) and the best docked pose (blue) as sticks on the grey receptor cartoon.

2.6. Model Quality on a Common Scale, and How Much of the Native Interface a Rigid Toxin Conformer Can Reach

Reporting docking and co-folding on the CAPRI scale places both against the wider literature (Figure 6; Section 5.8). At A1, no rigid-body run produced a Medium-or-better model: the best pose of each of the six runs reached DockQ 0.249–0.387, CAPRI Acceptable throughout, and the pose ranked first by the scoring function reached only DockQ 0.004–0.237. At A2, the two blind runs did reach CAPRI Medium (DockQ 0.604 and 0.536; L-RMSD 4.585 and 5.749 Å; I-RMSD 1.550 and 1.893 Å), whereas the restraint-guided rungs did not (DockQ 0.300–0.395). Every top-ranked-by-score pose at both anchors was CAPRI Incorrect or low Acceptable, which locates the failure in scoring rather than in sampling. Co-folding was CAPRI High at both anchors and in every one of the five samples reported (Boltz-2 DockQ 0.826–0.872 at A1 and 0.909–0.939 at A2; Boltz-1 0.808–0.821 at A2). Under the stricter CAPRI peptide-target bands, appropriate to a twelve-to-thirteen-residue ligand, every docked pose at both anchors is Incorrect, and Boltz-2 falls from High to Medium at A1; both classifications are reported because the choice of band changes the reading.

How much of that shortfall was reachable at all is a separate question, and it has a different answer at the two anchors. Maximising fnat and DockQ directly over the six rigid-body degrees of freedom of the docked apo conformer—rather than evaluating it at its RMSD-optimal superposition, which is not a maximum because neither metric is monotone in RMSD—gives a rigid-placement ceiling of 47 of 53 contacts (fnat 0.887) at A1, where one placement maximises both metrics and reaches DockQ 0.921, and of 43 of 56 contacts (fnat 0.768) at A2. At A2 the two metrics are maximised by different placements: the fnat-optimal one reaches DockQ 0.761, while the DockQ ceiling of 0.781 is reached at 40 of 56 contacts. Each metric's ceiling is therefore quoted from its own maximisation, and no single placement attains both. At A1, that ceiling lies far above what the search achieved (13 of 53, DockQ 0.387): 34 native contacts were reachable by a rigid placement of the very conformer being docked and were not found, so the A1 shortfall is attributable to search and scoring rather than to the conformer. At A2, the ceiling instead lies below the weakest Boltz-2 sample (43 versus 47 of 56). The 16-contact gap between the best docked pose and the weakest co-folded one there therefore divides into 12 contacts (75%) that a rigid search could have found and did not, and 4 contacts (25%) that no rigid placement of that apo conformer can reach; in DockQ, 0.177 of the 0.305 gap is search-limited, and 0.128 is unreachable. Because these maxima come from a stochastic search, they are lower bounds on the true ceiling, and the only rigorous upper bound available here is near-trivial (Section 5.8), so the A2 ordering rests on the convergence controls reported there rather than on proof.

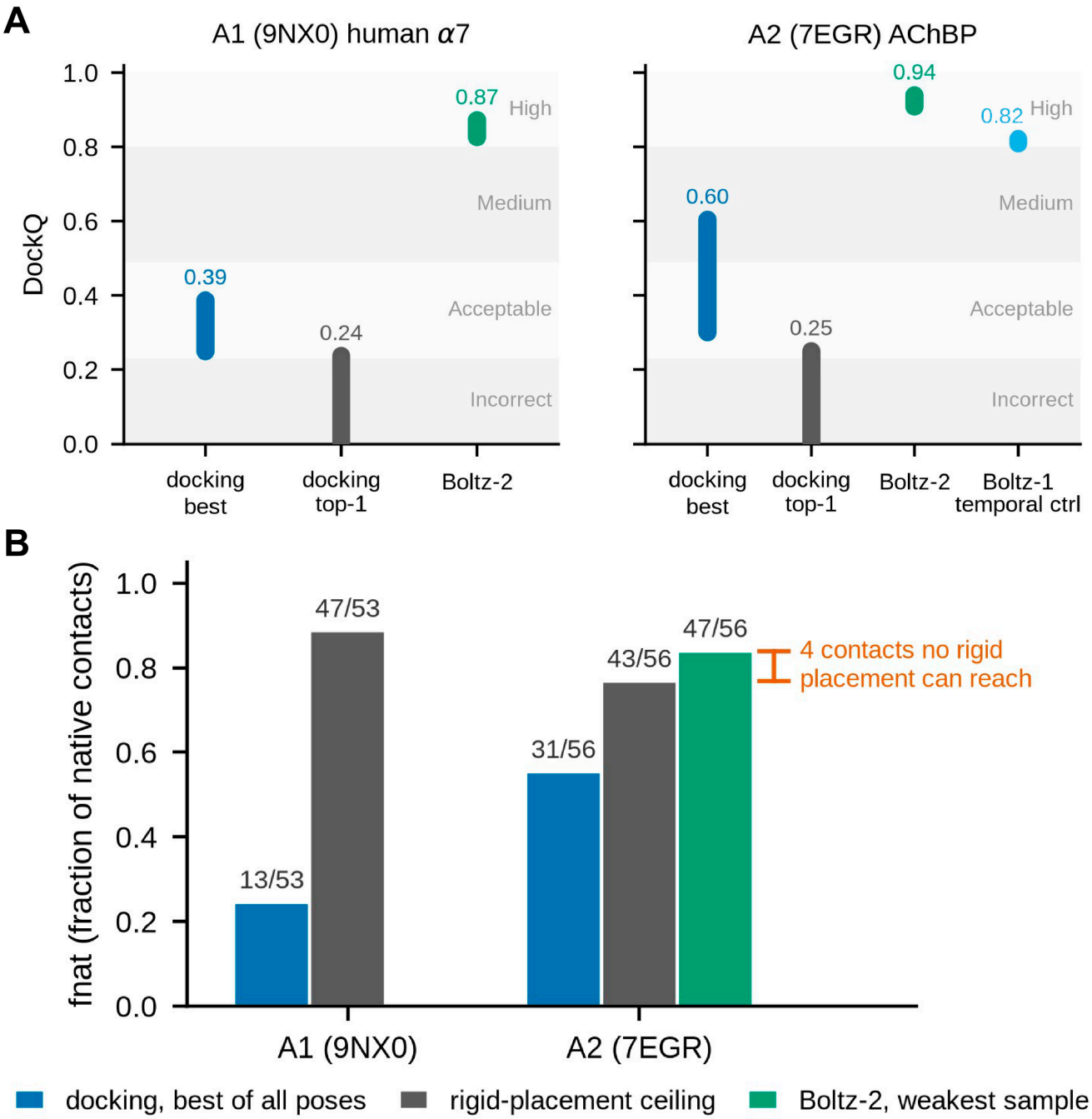


Figure 6. Model quality on the CAPRI scale, and the rigid-placement ceiling. (**A**) DockQ for each method at both anchors, drawn as a vertical bar spanning the range over runs (docking) or samples (co-folding) with the top of that range printed: docking best pose blue, docking top-1-by-score grey, Boltz-2 green, and at A2 only Boltz-1 light blue. The background bands mark the CAPRI High, Medium, Acceptable and Incorrect intervals of DockQ. (**B**) fnat at both anchors: achieved by docking over all poses (blue), the maximum reachable by any rigid placement of the docked apo conformer (grey), and the weakest of the five Boltz-2 samples (green). Reproduced and native contact counts are printed above each bar, and the bracket at A2 spans the last two bars.

2.7. Whether Confidence Separates Orthologous Receptors Depends on the Model; Separating Binders from Non-Binders Did Not Succeed in Either Model in the Comparisons Tested

Every comparison in this section rests on five diffusion samples per condition, so its statistical power is limited, and each result is a statement about the comparisons made here and not about a rate of failure. The preceding sections evaluate models against a known

answer. The use that motivates this work is different: ranking a peptide's engagement of a human receptor against its non-human counterpart, where no experimental structure exists for either and the model's own confidence output is the only available decision variable. We tested that directly on the case that stopped ACV1, and in two independent co-folding models (Figure 7).

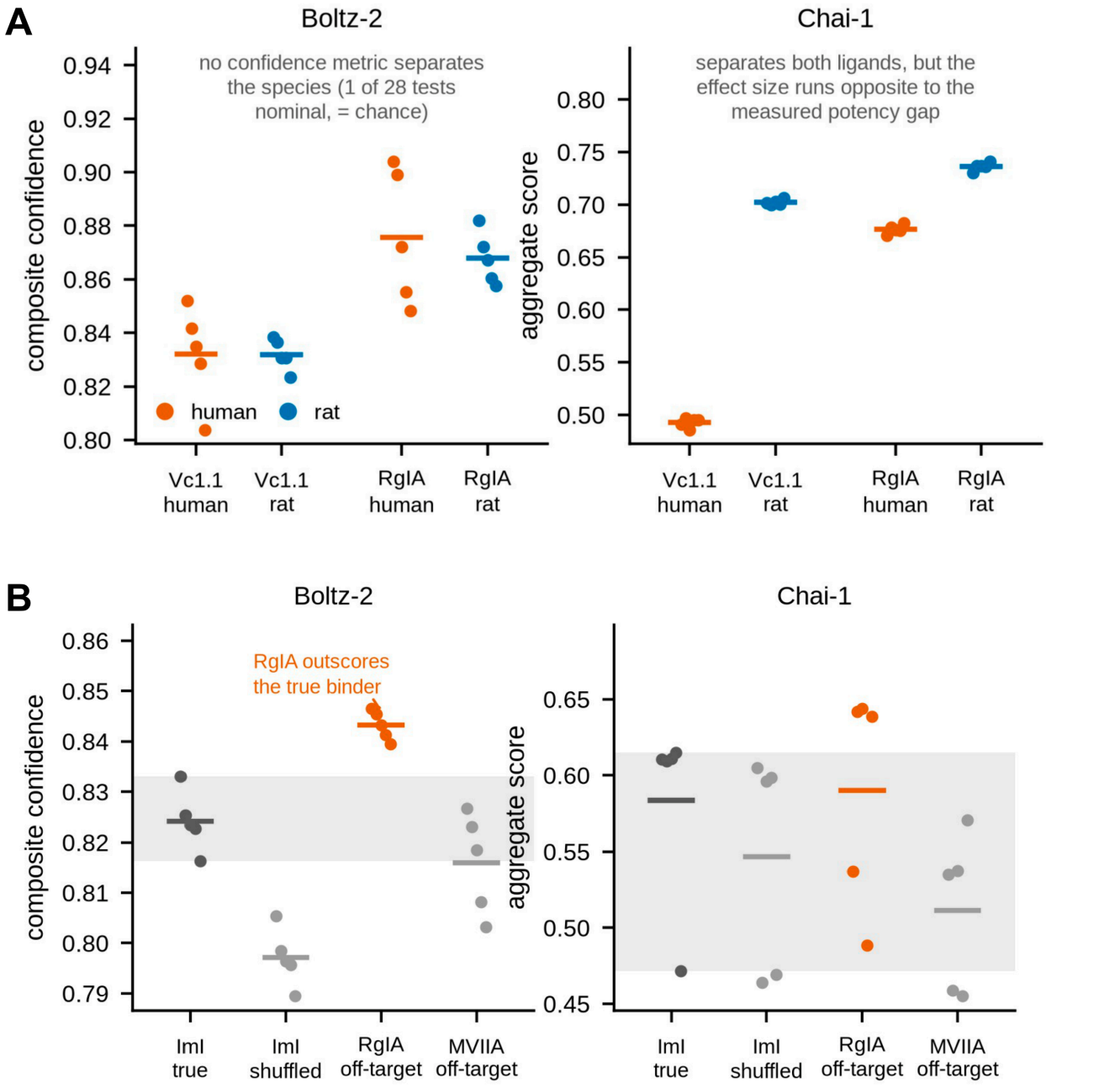


Figure 7. Confidence-based discrimination in two co-folding models. Each experiment is shown separately for Boltz-2 (composite confidence) and Chai-1 (aggregate score); the two quantities are on different scales and the panels are not to be read across models. (**A**) Per-sample confidence for Vc1.1 and RgIA co-folded with the human (orange) and the rat (blue) α9α10 receptor, five diffusion samples per condition, with horizontal bars at the condition means. (**B**) The same two metrics at the A1 α7 site for four ligands—the true binder ImI (black), a shuffled decoy (grey), RgIA (orange, an α9α10-selective conotoxin reported to also block α7 at reduced potency [17]), and MVIIA (grey, a $Ca_V2.2$ ligand unrelated to any nicotinic receptor)—again five samples each with bars at the means. The shaded band spans the range of the true binder.

Vc1.1 was co-folded with rat and with human α9α10 under three matched conditions (full-length extracellular plus transmembrane domains, extracellular domain only, and extracellular domain with explicit disulfide restraints), five diffusion samples each, all settings and the seed held identical between species. As a second and independent test with a different ligand, RgIA—reported to be 300-fold more potent at rat than at human α9α10, a difference attributed largely to a single α9 Thr56→Ile substitution [18]—was run under the same extracellular-domain protocol. Because no experimental α9α10 structure exists, absolute accuracy is not evaluable at these targets, and no RMSD to a native pose is reported; only the relative ordering of the two species is.

None of the model's confidence outputs separated the species beyond diffusion-sampling noise. Across 28 species comparisons (four conditions × seven metrics: ipTM, pTM, composite confidence, mean interface PAE, ligand pLDDT, interface contact count, and complex interface pLDDT; exact permutation tests, $n = 5$ per species), exactly one cleared $p < 0.05$ uncorrected (ligand pLDDT, Vc1.1 extracellular-domain condition, $p = 0.040$)—the number expected by chance at that threshold across 28 tests. A pose-reproducibility diagnostic was more suggestive for Vc1.1, replicate rat poses clustering tightly while human poses did not (exact permutation $p = 0.016$ full-length, 0.040 extracellular-domain, 0.056 with disulfide restraints), but with five samples per species the exact test's attainable minimum is $p = 0.0079$ over 126 labellings, so the strongest value obtained is twice the floor and an order of magnitude above a Bonferroni threshold for the 32 tests performed. Two further checks argue against reading even that as a species effect. The result is not robust to the superposition used to define it: computing the same test after superposing on the α9 subunit alone, rather than on both receptor chains, raises it to $p = 0.079$ (full-length) and $p = 0.159$ (extracellular-domain), so which frame the poses are compared in decides whether the asymmetry is nominally significant. And the asymmetry co-varies with which of the two dimer interfaces each sample occupied: in both Vc1.1 conditions all five rat samples sit on the α9(+)/α10(−) interface while the human samples split two to three across the two interfaces, so the tighter rat clustering is in part a statement about interface consistency rather than about the orthologue, and it is largely the same observation as the interface-assignment result reported below. Restricting the comparison to samples sharing one interface leaves two human samples and no usable power. We therefore report the diagnostic and do not interpret it as evidence of species discrimination. Decisively, the asymmetry did not reproduce for RgIA ($p = 0.476$; within-human, within-rat and between-species median pairwise ligand RMSD 31.2, 30.6 and 30.5 Å respectively), even though RgIA carries the larger and better-characterised species difference of the two ligands. For scale, the same model's within-run pose reproducibility at the two anchors where an experimental structure exists is 0.949 Å (A1) and 1.726 Å (A2) median; the 30 Å spread seen at α9α10 indicates broad uncertainty at this target rather than a confident species-specific prediction.

Because a single model cannot distinguish a property of the method from a property of that model, both experiments were repeated with an architecturally independent co-folder, Chai-1 [19], on the identical input files and with matched inference settings (Section 5.4). Its accuracy at the two anchors where a deposited structure exists is comparable to Boltz-2's (ligand RMSD to the deposited pose 1.16–1.49 Å at A1 and 1.32–1.74 Å at A2 across five samples), so the species comparison below is not a report on a malfunctioning model.

The two models agree on the pose-reproducibility asymmetry for Vc1.1 and disagree on everything else. Chai-1 reproduces the Vc1.1 pattern and more strongly (median within-species pairwise ligand RMSD 1.97 Å for rat against 31.63 Å for human, exact permutation $p = 0.008$, at the test's attainable floor), but for RgIA it lands in a different regime from Boltz-2 altogether: both species' replicate poses cluster tightly (3.22 Å rat, 4.34 Å human) and the species test is marginal ($p = 0.056$), where Boltz-2 had scattered both species and returned

a complete null (30.6 and 31.2 Å, $p = 0.476$). Unlike Boltz-2, Chai-1's scalar confidence outputs do separate the species, cleanly and in the direction the experimental potencies would predict—rat above human for both ligands, on composite confidence, pTM and ipTM, every test at the permutation floor, with a between-species difference 15–30 times the within-species sampling spread. Two observations stop that from being read as species discrimination working. First, the effect size runs opposite to the experiment across the two ligands: RgIA carries the larger measured species gap (300-fold, against roughly ten- to thirty-fold for Vc1.1) but the smaller predicted one (0.060 in composite confidence, against 0.209 for Vc1.1). Second, one metric reverses: ligand pTM is higher for human than rat for Vc1.1. A receptor-side artefact is not the explanation—the paired MSA depths differ by only 3.6% between species (5967 human, 6183 rat), against a 42% confidence difference, and the Vc1.1 and RgIA runs of a given species use the identical receptor alignments yet give species gaps differing 3.5-fold, so a ligand-dependent component is real. What the two models together establish is therefore narrower than either alone would suggest: whether a co-folding model's confidence separates orthologous receptors is a property of the model, not of the method, and where it does separate them, the magnitude does not track the measured potency difference.

A negative control at A1 gives the corresponding result for binder discrimination. Against the deposited α7 receptor, co-folded ImI reproduced its native pose closely and consistently (ligand RMSD 1.134–1.882 Å, native-site residue recovery 0.893–0.964, interface Cα RMSD 0.672–0.897 Å across five samples). Three controls were run identically: ImI with its non-cysteine positions shuffled, and its cysteine framework held fixed; MVIIA, a conotoxin selective for the voltage-gated calcium channel $Ca_V2.2$ and unrelated to any nicotinic receptor; and RgIA, an α-conotoxin selective for α9α10 nAChR that is also reported to block α7, far less potently than α9α10 [17]. No single confidence metric separated all three controls from the true binder; each was separated by a different metric, and in one comparison a control outscored the true binder—RgIA was placed within 1.5–2.1 Å of the native ImI centroid with site recovery, ipTM and interface PAE statistically indistinguishable from the true binder, and a composite confidence score higher than ImI's (0.840–0.847 versus 0.816–0.833). We do not report that ordering as inverted, because the ordering it would have to invert is not established. The comparison the primary source makes is of one peptide across two receptors—RgIA is a much more potent inhibitor of its own primary target, α9α10, than of α7, its α7 values being given in that paper's supplementary material [17]—whereas the confidence scores above compare two peptides at one receptor, and no head-to-head measurement of ImI and RgIA at α7 on a common scale is reported in that source's main text. The two comparisons run on different axes, and the first does not fix a direction for the second. The claim supported here is the narrower one: the rank order of these two peptides at α7 is not established in the literature, so a confidence score that ranks them cannot be scored correct or incorrect against it, and confidence alone cannot settle the ranking either way. What the score does track is pose plausibility—RgIA was placed accurately in the site and scored accordingly—which is a different quantity from the binding strength a ranking would need. RgIA and ImI share an identical eight-residue N-terminal segment, which may bear on this, though that interpretation is not tested here. One control the panel does not contain is an α-conotoxin verified to be inactive at α7, which is the hardest case—a close structural relative that genuinely does not bind. The limitation is simply that no condition in the present panel meets that requirement, and it is not a property of this target that would make such a control uninformative: α7 is homopentameric, its five subunit interfaces are equivalent, and the two-interface ambiguity described below for the α9α10 heterodimer does not arise here. These three controls therefore establish that confidence failed in the comparisons actually performed, and do not establish a rate at which it fails. The practical

consequence is scoped to the same range: within the comparisons made here, the model's confidence output cannot by itself be used to decide whether a conopeptide binds at this site, still less which species it prefers.

A third, independent line of evidence points the same way. Across all eight run conditions of the Vc1.1 and RgIA species experiments above (five diffusion samples each, 40 samples total), each α9α10 dimer admits two structurally distinct docking modes for an α-conotoxin: the toxin can occupy the α9(+)/α10(−) interface, with α9 supplying the principal (+)-face loop B/C contacts and α10 the complementary (−)-face, or the α10(+)/α9(−) interface, with the two subunits' roles exchanged. Classifying every sample by which mode it occupies (Methods §5.4), 24 of 40 land on the α9(+)/α10(−) interface and 12 on α10(+)/α9(−); the remaining 4 do not resolve clearly to either. ipTM does not separate the two modes—α9(+)/α10(−) 0.612–0.874 (mean 0.722) against α10(+)/α9(−) 0.633–0.806 (mean 0.723), a Monte Carlo permutation test on the mean difference (N = 200,000 shuffles) returning $p = 0.96$. Which assignment is correct is not resolved by these models, and it is not settled in the literature either. α9α10 is reported to present three potential binding sites—α9(+)/α9(−), α9(+)/α10(−) and α10(+)/α9(−)—and molecular-dynamics simulations of human α9α10 extracellular-domain pentamers, built on a crystal structure of the monomeric α9 extracellular domain in complex with RgIA, support a favourable binding of RgIA at α9(+)/α9(−) or α10(+)/α9(−) rather than at the α9(+)/α10(−) interface; the same study records that earlier modelling attempts had reached opposing conclusions, one favouring α9(+)/α10(−) and another α10(+)/α9(−) [20]. That evidence is simulation-based and is offered by its authors as support rather than as a settled assignment, so it does not adjudicate between the two modes classified here, and this study is not the place to settle the question either. The third of those interfaces, α9(+)/α9(−), requires two α9 subunits and could not arise in the models classified here at all: each of the 40 is an α9α10 heterodimer carrying a single copy of each subunit (Section 5.4), a construction that admits only the two modes above, and no sample was assigned to a third. The ambiguity that does arise is not cosmetic. The α9 residue to which the measured rat–human RgIA difference is principally attributed—the Thr/Ile substitution of Azam and McIntosh [18], at position 64 of the mature subunit and 61 of the construct used here (Table S23)—lies on α9's complementary face, 26.7 Å from that same subunit's loop C. It is therefore part of the modelled interface only in the α10(+)/α9(−) mode: across all 40 samples it makes no contact with the toxin in any α9(+)/α10(−)-mode model. Whichever mode is right, a confidence score that cannot separate the two cannot indicate whether the residue carrying the species difference is in the interface at all, which is a sharper limitation than the potency- and species-discrimination failures reported above and is independent of them.

Two assumptions underlie the two predicted interfaces. First, each model contains exactly one α9 and one α10 chain, so only the α9(+)/α10(−) and α10(+)/α9(−) interfaces can form (Section 5.4); the α9(+)/α9(−) interface is excluded by construction, and the subunit arrangement of the native pentamer, which determines which interfaces exist, is not tested. Second, the input specifies sequence only, so no receptor conformational state is imposed, and both interfaces are treated as accessible. The models therefore say nothing about which interface is occupied in a native receptor. The residue carrying the rat–human RgIA difference (position 64 of the mature α9 subunit, 61 in the construct; Table S23) contacts the toxin only in the α10(+)/α9(−) mode, which makes it the most direct test. If that mode is functional, substituting this residue between the rat and human identities should change the potency of RgIA and Vc1.1 in two-electrode voltage-clamp recordings of Xenopus oocytes expressing α9α10 (IC50 shifts, with matched controls for expression); if the α9(+)/α10(−) mode is functional, it should not. Substitutions at residues that contact the toxin in one mode only, taken from the 40 classified models, would extend

the test to the interface as a whole, and radioligand or fluorescence competition binding on the same mutants would separate binding from gating. Receptors with a fixed subunit arrangement, such as concatemeric constructs, would test each interface separately. None of these experiments was performed; they are proposed designs.

3. Discussion

Taken together, the four results above split cleanly along one line. In the LightDock configuration tested here (rigid receptor, a single fixed apo conformer, no post-search refinement), rigid-body docking and its rescoring functions failed to reach the study's own accuracy threshold at either anchor, on either RMSD basis, however much interface information the search was given (Sections 2.2 and 2.3). A co-folding model reached that threshold, or came close to it, at both anchors, and did so even in a run of the same model whose training data provably predate the deposition of the more leakage-sensitive anchor (Section 2.4). The interesting question this benchmark raises is not simply "which method is more accurate"—with a training-data question hanging over one anchor and no such question over the docking result, that comparison is already asymmetric—but why more information helped the docking search not at all, and whether that failure mode is likely to be specific to this ligand class, this docking engine, or both.

One candidate explanation connects the two docking-side findings directly. Section 2.3 shows that a single non-physical term in LightDock's restrained-run scoring function determines which pose that function ranks first, changing the winner in every restrained run and never leaving the true top-1 pose in place; Section 2.2 shows that the restraints built from that same term did not improve, and in every paired comparison worsened, the best pose the search actually found. Read together, these are consistent with restraints acting on this ligand class as a term the scoring function can satisfy independently of pose accuracy, rather than as information that narrows the search toward the native interface. This is the opposite of what graded interface disclosure did for LightDock on globular protein–protein complexes, where moving from no restraints to the complete true interface raised top-10 success from 14.5% to 92.7% [9]. A disulfide-locked twelve-to-nineteen-residue peptide docked against a pentameric membrane-receptor extracellular domain differs from that benchmark's unbound protein pairs in scale, in backbone flexibility, and in how much of the receptor surface a cross-donated restraint set can plausibly cover; the present study cannot separate which of these drives the reversal, only that on this specific target class the same restraint-driven mechanism that helps on globular complexes did not.

A second explanation operates before the search begins. It bounds what any rigid search could have achieved at A2, but it does not by itself account for the A1 result, and the two claims should be kept at different scopes. The α-conotoxin orthosteric site is formed between two subunits and is rearranged when the peptide binds, so a rigid docking protocol asks a free-solution conformer to enter a pocket shaped around a different conformer. Measured against deposited unbound structures of the same two proteins, that rearrangement is an opening of loop C rather than a closure around the peptide: the loop C tip Cα lies 10.32 Å from the toxin centroid in the toxin-bound α7 subunit against 8.40–9.49 Å in the fifteen unbound α7 subunits of PDB 7EKI, 9J37 and 8V89, and 9.94 Å in the toxin-bound AChBP subunit against 5.36–7.79 Å in the five unbound subunits of PDB 2BYN; all twenty unbound subunits move in the same direction, by 0.83–1.92 Å at A1 and 2.15–4.57 Å at A2 (Section 5.11). Either way, the pocket a rigid apo conformer is asked to occupy is not the pocket that conformer was solved in, and here that is not a matter of degree: superposing each unbound NMR conformer of either toxin onto its deposited bound counterpart at that conformer's own optimal rigid fit, and then measuring against the deposited receptor, not one of the 60 deposited ImI models and not one of the 20 RgIA models is free of atomic

overlap (Section 5.9). The docked conformers carry 10 (A1) and 45 (A2) intermolecular heavy-atom pairs below 2.0 Å, with closest approaches of 0.91 and 0.56 Å, against a deposited bound ligand that has none below 2.5 Å at either anchor and a closest approach of 2.52 Å. A search that cannot place its ligand without overlap will not reach the native pose no matter how the site is described to it, which is consistent with what the restraint ladder found (Section 2.2), and a scoring function evaluating such poses is ranking degrees of overlap rather than degrees of nativeness, which is consistent with the rescoring result (Section 2.3). How much of each anchor's shortfall this accounts for, however, differs between the two anchors, and the difference is arithmetic rather than interpretive. At A2 the 3 Å all-atom criterion is closed to every deposited conformer, not only to the one docked: the best of the twenty 2JUT models leaves a 3.248 Å floor after optimal superposition and the model actually docked leaves 4.453 Å (Table S6), so 0 of 20 available conformers make the criterion reachable and the A2 outcome is bounded before any search runs—the conformer-limited fraction of the A2 accuracy gap quantified in Section 2.6 is that same effect measured from the other side. At A1 the criterion was open—the docked conformer's floor is 2.439 Å and the best deposited ImI model's is 1.022 Å—and the best of the 3144 scored swarms still returned 7.647 Å, 5.208 Å above its own floor. The conformer limit therefore explains why A2 could not pass the criterion; it does not explain why A1 did not, and that residual is the search and scoring failure quantified in Section 2.6 rather than a second consequence of the same cause. The co-folded ligand conformers, by contrast, lie 0.97–1.21 Å (A1) and 1.50–2.63 Å (A2) from the bound conformer, because a co-folding model predicts the ligand's conformation together with its pose and is not required to reuse a free-solution structure. What this does not establish is that adding flexibility would fix the docking result: it shows that a rigid unbound conformer cannot be placed without overlap, not that a flexible protocol would find the native pose. Testing that requires the refinement stage discussed below.

The scoring-function result is reported for nine classical potentials inside one docking package, and should be read at that scope. One instrumentation caveat belongs with it. Of the nine functions evaluated, eight negate their computed energy before returning it, so that LightDock's maximising convention selects the more favourable pose; DFIRE2 alone does not (scoring/dfire2/c/cdfire2.c returns the energy scaled but unnegated, whereas for example scoring/fastdfire/c/cdfire.c returns (energy × 0.0157 − 4.7) × −1 under a comment stating that it changes the sign), while the ranking routine sorts in decreasing score order in both cases. Since DFIRE-family energies are favourable when negative, the panel's DFIRE2 column therefore reports a quantity whose sign opposes the other eight, which is the most likely origin of its positive score–RMSD correlation. This is reported as a property of the installed build rather than of the potential, alongside the two other defects this study encountered in the same package (Tables S7 and S11), and the DFIRE2 row should be read with it in mind. LightDock 0.9.4 exposes ten protein–protein scoring functions of the ten possible, of which nine were evaluated (Section 2.3); this study did not run an equivalent panel through a second docking engine, so whether the same failure to rank near-native poses would recur under HADDOCK's or another engine's own scoring is a direct extension of this work rather than a claim already established by it. What can be said from the data in hand is that the failure was not confined to a weak subset of the panel—the function LightDock's own production search used internally, vdW, showed a favourable rank correlation and one of the worst top-1 selections at the same time (Section 2.3)—so a better score-RMSD correlation on this ligand class did not translate into a better selected pose even within the one engine tested.

The docking protocol tested here is also narrower than the phrase "template-based docking" might suggest, and the two studies already cited in the Introduction as this ligand class's docking precedent make the point directly. ToxDock combined Rosetta FastRelax

with FlexPepDock—flexible receptor side chains and full ligand backbone flexibility during a dedicated refinement stage—to redock α-conotoxin PnIA[A10L,D14K] onto an AChBP surrogate at 0.6–1.3 Å Cα RMSD, and Gulsevin and Meiler reported 1.6 ± 0.2 Å top-5 accuracy across eleven AChBP and α7–AChBP chimera complexes using the same family of Rosetta refinement tools [7,8]. This study's LightDock configuration held the receptor fully rigid (use_anm: false), docked a single fixed apo conformer, and applied no post-search local refinement—only a global search stage. The failure reported in Sections 2.2 and 2.3 is accordingly a statement about that specific configuration—a rigid-body, restraint-driven global search with no refinement step—and not about template-based docking as a method class; a version of this benchmark that added the kind of flexible refinement ToxDock and Gulsevin and Meiler used is a direct, comparatively inexpensive extension of this work, and every docking-side conclusion in this paper is scoped to the configuration actually tested, not beyond it.

The comparison between LightDock and co-folding differs in modelling assumptions as well as in performance, and the two cannot be fully separated here. Three results in this paper bound the contribution of rigidity. At A1, a rigid placement of the docked apo conformer can reproduce 47 of 53 native contacts, against 13 found by the search (Section 2.6), so rigidity of the ligand does not explain the A1 shortfall. At A2, the rigid-placement ceiling (43 of 56 contacts) lies below the weakest co-folded sample (47 of 56), and of the 16-contact gap between the best docked pose and that sample, 12 contacts (75%) were reachable by a rigid placement, and 4 (25%) were not. No unbound conformer of either toxin can be placed in the bound site without atomic overlap (Section 5.9). What was tested is flexibility inside the same search: with LightDock's normal-mode option (ten receptor and ten ligand modes at the default deformation range) the best ligand RMSD in the A1 blind search was 8.43 Å, against 8.74 Å for the rigid run, and no pose came within 7 Å (receptor and ligand deformation 0.49 and 0.54 Å RMS at the best pose; Section 2.2). That range of deformation does not reach the loop C opening of Section 5.11, and a comparison with flexible peptide-docking programs was not performed. A balanced assessment of the two approaches needs both, and the docking-side conclusions are accordingly limited to the configuration tested.

The co-folding result is more encouraging, and the temporal control is what makes it interpretable rather than merely suggestive. Because Boltz-1's training cutoff precedes 7EGR's release, its 3.107–3.301 Å result at that anchor cannot be entry memorisation; the most it can reflect is generalisation from homologous training data, and it still beat every one of the 3144 scored docking swarms (Section 2.4). That is consistent with a broader pattern reported for co-folding methods generally—accuracy declining with a target's dissimilarity to the training set, most directly quantified across protein–peptide interfaces by FoldBench [21]—without this study being able to quantify that dependence for the conotoxin–nAChR class specifically, since it tests only two anchors at two points on that dissimilarity axis. Boltz-2's own result at A2 remains a positive control rather than independent evidence of generalisation, precisely because 7EGR falls inside its training window (Section 2.4); and A1's MEDIUM leakage rating means even the more recent anchor is not a fully clean test, since a 2019 structure exceeds the same-complex homology threshold AlphaFold3 itself uses to exclude test interfaces, by a margin the underlying identity search puts just above that threshold rather than comfortably clear of it. Reporting that margin rather than rounding it to "leakage-free" is the more defensible choice available with two anchors. A benchmark of structurally dissimilar, experimentally validated complexes would provide stronger evidence of generalisation than either anchor, and this study does not provide one.

This benchmark is necessarily narrow, and the narrowness bounds what can be concluded from it. Two deposited complexes, one non-human, are not a representative sample of α-conotoxin–nAChR interfaces, and neither the apo-conformer floor result (Section 2.2) nor the restraint-information result should be extrapolated to conotoxins with different lengths, disulfide topologies, or receptor subtypes without independent testing. The apo-ensemble floors reported here (Section 2.2) also carry a standing methodological warning that applies to any future use of this design: because the floor is a function of which deposited NMR model is chosen, and because a model with a low floor is, by construction, one whose unbound conformer already resembles the bound one, selecting an ensemble member after seeing how close it comes to the native answer would constitute the same kind of leakage this study documents for co-folding training data, applied instead to conformer choice. A benchmark of this design should fix its apo-conformer selection rule before any docking is run, not after. The confidence results of Section 2.7 carry their own version of the same limit. They rest on three negative controls against one true binder at one site, and on two orthologue pairs for the species comparison, with five diffusion samples per condition; that is enough to show that confidence failed in the specific comparisons made, and not enough to put a rate on how often it fails. The binder-discrimination panel in particular contains no α-conotoxin verified to be inactive at α7; that control is absent from the panel rather than precluded by the target, since α7 is homopentameric and does not present the two-interface ambiguity that complicates α9α10 (Section 2.7), so every binder-discrimination statement here is bounded by the comparisons actually run. The species side is further bounded by the two ligands available with a measured human–rat potency gap, so the observation that predicted and measured gaps run in opposite directions rests on two points and would be overturned or confirmed by a third. Extending either panel—a set of α-conotoxins verified inactive at the test receptor, and further orthologue pairs with measured selectivity—is the cheapest way to convert these results from a demonstration that the failure occurs into an estimate of how common it is. The alignment ablation was run at A1 only, with one seed and five samples per condition.

The co-folding output is a structure with confidence scores, not a binding free energy, and several quantities that bear on affinity are absent from the modelling framework used here. No binding free energy is computed. Receptor conformational dynamics enter only as a single static structure per sample, although loop C differs between unbound and toxin-bound subunits by 0.83–1.92 Å at A1 and 2.15–4.57 Å at A2 (Section 5.11). Solvation and side-chain rearrangement are not modelled explicitly. Boltz-2 includes an affinity module, but in the installed version (2.2.1) it accepts only a small-molecule ligand as the binder, and the toxins are supplied as protein sequences, so it was not run. Physics-based rescoring of the predicted complexes, such as end-point free-energy methods with molecular dynamics, was not performed and is a task for later work.

For conopeptide drug discovery, the practical reading is closer to the docking result than the co-folding one. ACV1's clinical discontinuation followed from a potency gap between the rat receptor used in its discovery and the human receptor the drug needed to act on [2]. A decade later, dedicated docking protocols for this ligand class were still being built and validated without any experimentally determined conotoxin–nAChR complex to test them against, using acetylcholine-binding-protein surrogates instead because none had yet been solved [7]. That structural gap is now partly closed for two complexes, and on those two complexes, rigid docking with restraint information did not reproduce the native interface reliably enough to have flagged such a gap in advance, while a co-folding model, evaluated with its own training-data caveats stated rather than assumed away, came far closer. It would be tempting to read that as evidence that co-folding could have caught ACV1's species gap prospectively. It does not follow, and the distinction matters

for anyone planning to use these methods that way: reproducing a deposited interface accurately is a statement about pose geometry against a known answer, whereas flagging a species gap requires the method to rank two receptors it has never seen a structure for, using its own confidence output as the decision variable. Those are different tasks, and accuracy on the first does not license the second. What this benchmark establishes is the first: on the two complexes tested, co-folding reproduced the native interface where this rigid-body docking configuration did not. Whether either method can support the second—prospective species discrimination—is a separate question requiring receptors for which no experimental structure exists, and is addressed in Section 2.7 rather than assumed here.

4. Conclusions

On two deposited α-conotoxin–nicotinic receptor complexes, the rigid-body, no-refinement LightDock configuration tested here and its rescoring functions did not reach this study's accuracy threshold, and among the poses its search produced from the apo conformer, no scoring function ranked a near-native pose first, although five of the nine ranked the deposited bound conformer first or within 1 Å when it was inserted (Section 2.3); the pose each run ranked first reproduced none of the native interactions at the human anchor. A deep-learning co-folding model reached CAPRI High at both anchors, including at the anchor whose own deposition postdates its training data but which is rated MEDIUM leakage risk, and recovered the loop B arginine anchoring that no docked pose reached at any level of interface disclosure. Two limits on the docking side are now separable: the free-toxin conformer, which no deposited unbound model places in the bound site without atomic overlap, and the search and scoring, which at the human anchor left 34 of 53 reachable native contacts unfound. Apo-conformer choice must therefore be fixed by a rule stated before docking, not selected afterward by proximity to the answer.

Two results constrain how this benchmark's design should be reused. First, the direction of the interface-disclosure effect is a property of the receptor construct: on the same anchor, with the same toxin and protocol, disclosure degrades the best pose at full length and improves it monotonically when the receptor is truncated to its extracellular domain. A graded-disclosure ladder run on one construct can return either sign, so the construct must be reported with the result. Second, pose accuracy and confidence-based selection are different capabilities and were not acquired together: in two independent co-folding models, confidence failed to separate a true binder from a shuffled sequence or from a conotoxin not selective for this receptor, separated orthologous receptors in one model but not the other with an effect size that ran opposite to the measured potency difference across two ligands, and did not distinguish which of the two opposite subunit interfaces the toxin occupied in a well-powered sample of the same α9α10 models (α9(+)/α10(−) versus α10(+)/α9(−), 24 versus 12, $p = 0.96$)—a distinction that decides whether the residue carrying the measured species difference is in the interface at all. These confidence findings are bounded by the comparisons actually performed: the binder-discrimination panel contains no α-conotoxin verified to be inactive at the test receptor, so they establish that confidence failed where it was tested, not how often it fails.

These are the properties of one ligand class evaluated against two structures, one docking engine and two co-folding models; extending the comparison to further conotoxin–receptor complexes as they are deposited, and to a docking protocol that permits receptor and ligand flexibility, is the direct next step. For a peptide family where more than 97% of catalogued sequences have no deposited structure to model from, a co-folding model checked against its own training cutoff is, on the evidence gathered here, the more promising building block for structural work on this ligand class than the rigid-body docking configuration tested here. That endorsement covers reproducing an interface whose answer

is known. It does not extend to the task that ended one clinical program for this peptide family—ranking a peptide's engagement of a human receptor against its non-human counterpart, prospectively and without an experimental structure for either. On the evidence in Section 2.7, the confidence outputs of the models tested here do not support that task, and reporting pose accuracy as though they did would misdirect the readers most likely to act on it.

5. Materials and Methods

5.1. Structural Anchors and Apo Conformers

Two deposited complexes were used as docking targets. Anchor A1 is the human α7 nicotinic acetylcholine receptor (nAChR) bound to α-conotoxin ImI (PDB 9NX0; cryo-EM, 3.06 Å; released 4 February 2026) [10]; the docked receptor pair was chains D and E (residues 1–478, extracellular and transmembrane domains, 6326 heavy atoms). Anchor A2 is *Aplysia californica* acetylcholine-binding protein bound to α-conotoxin RgIA (PDB 7EGR; X-ray, 2.50 Å; released 6 April 2022) [11]; the docked receptor pair was chains H and G (residues 19–224, extracellular domain only, 3276 heavy atoms). Both pairs were cut from their deposited homopentamer and represent a complete orthosteric (+)/(−) subunit interface; A1's pair additionally carries the transmembrane domain, which is the source of the swarm-count asymmetry between anchors described below. Receptors were treated as rigid throughout (normal-mode flexibility disabled, use_anm: false).

Each toxin was docked as an apo (unbound) conformer taken from a deposited NMR ensemble: ImI from PDB 1IMI, model 01 (12 residues, GCCSDPRCAWRC, 90 heavy atoms after stripping the model's C-terminal amide cap nitrogen to match the 90-atom footprint resolvable in 9NX0); RgIA from PDB 2JUT, model 01 (13 residues, 105 heavy atoms). This model was fixed before any docking was performed and was not chosen by proximity to the native pose. Residue correspondence between the docked apo ligand and the native bound ligand was established by Needleman–Wunsch alignment on three-letter residue names, followed by atom-name matching within each aligned residue pair, giving 90/90 matched atoms at A1 (one native-only atom, the bound structure's amide cap nitrogen, is unmatched) and 105/105 at A2.

An intrinsic, docking-independent conformational floor was computed for every apo model as the Kabsch-superposition RMSD to the corresponding bound conformer, using the same matched-atom correspondence. For the two apo models actually docked, this floor is 2.439 Å all-atom/1.056 Å backbone at A1 and 4.453 Å all-atom/2.397 Å backbone at A2. Throughout, a run's excess RMSD is its best-RMSD pose's RMSD minus this floor (best—floor); this is a naming convention for RMSD above the conformational floor, not a strict decomposition, because the floor is computed after optimal (Kabsch) superposition while the docking RMSD it is subtracted from is not (Section 5.2), so the difference still carries whatever translational and rotational error the docking search left in, on top of the RMSD attributable to search and scoring. The same floor was computed for every other deposited NMR model of each toxin—all 30 models of PDB 1G2G, all 20 of 1IMI, and all 10 of 1CNL for ImI (60 models total, three entries); all 20 models of 2JUT for RgIA—to characterise how much this floor depends on apo-entry choice (Section 2.2, Figures 2D and S4, Table S6). A bound-conformer control was computed from the second, independently modelled copy of ImI already present in 9NX0 itself (chain F versus chain I).

5.2. Rigid-Body Docking

Docking used LightDock 0.9.4 (pip package; the package's own __version__ string, "3.0", is a branding artifact and does not indicate the algorithm version) [13], a glowworm swarm optimisation (GSO) algorithm. Each run placed 200 glowworms per swarm and

optimised for 100 GSO steps, scored during the search by LightDock's van der Waals potential (lightdock3.py setup.json 100 -c 20 -s vdw); swarm placement covered the receptor surface at a swarm radius of 10 Å, giving 390–392 swarms per run at A1 and 132–134 at A2, the difference reflecting A1's larger receptor surface (extracellular plus transmembrane domain versus extracellular domain only). Twelve runs were performed in total: two anchors × three information rungs × two independent GSO search replicates (distinct random seeds for swarm placement and glowworm initialization). These values were not chosen from a convergence test. To test sampling, the 20 swarms whose starting positions are nearest the deposited ligand were rerun at A1 with five times the glowworms (1000) and four times the steps (400), with the same swarm placement and seed; the best ligand RMSD in those swarms fell from 8.75 Å to 7.06 Å at 100 steps with the denser sampling and was not lower at later steps (8.09 Å at 400 steps), and no pose came within 7 Å (Table S28). The test covered 20 of the 390 swarms of one run, because a run of all swarms at those settings exhausted memory and would have taken about twenty times longer.

The three information rungs varied only the restraint set passed to LightDock's restraint-driven search; docking parameters were otherwise identical within an anchor. BLIND supplied no restraints. NATIVE supplied the anchor's own native contact footprint: receptor residues with any heavy atom within 5.0 Å of the deposited toxin (28 residues at A1, 33 at A2) were passed as active receptor restraints, giving 100% recall of the native footprint by construction. XDON supplied a restraint set built by transplanting a different structure's interface footprint onto the target—9NX0 received a restraint set built from PDB 2BYP, 7EGR from PDB 6HY7—recovering 89.3% of A1's own native footprint but only 36.4% of A2's (Table S4); the full cross-donation audit, including donor/target combinations not carried through to docking, is reported in Table S4. Swarm counts were matched across rungs within an anchor by setting swarms_per_restraint explicitly. The XDON rungs used LightDock's default of 20; at that default the NATIVE rungs would have produced 404 swarms at A1 and 475 at A2, against 390 and 134 for the corresponding BLIND runs—a 3.5-fold excess at A2—so swarms_per_restraint was set to 14 at A1 and 4 at A2, giving 392 and 132 swarms respectively. Rung comparisons within an anchor are therefore made at matched swarm counts by construction (390/390/392 at A1; 134/134/132 at A2); within an anchor, the two GSO search replicates of a given rung used matched swarm counts.

Accuracy was reported as ligand RMSD in the deposited (native) coordinate frame after algebraically reconstructing each pose from LightDock's own receptor-centred working frame (translation and quaternion rotation per glowworm, read from gso_100.out) plus the fixed translation LightDock applied when centring the receptor at setup; no additional superposition was performed, since superposing would hide a correctly scored but rigidly mistranslated pose. The pose-reconstruction and RMSD pipeline was independently cross-checked by regenerating full swarms as real PDB coordinates via LightDock's own lgd_generate_conformations.py and recomputing RMSD through a code path sharing no logic with the analytic one; the two agreed to within 8×10^{-5} Å, consistent with PDB fixed-point coordinate precision. Three RMSD bases are reported throughout: all-atom, backbone (N, Cα, C, O), and Cα-only.

5.3. Rescoring

Nine of LightDock 0.9.4's ten protein–protein scoring functions (DFIRE, DFIRE2, fastDFIRE, MJ3H, PISA, SD, TOBI, cpydock, vdW) were applied, independently of the search that produced them, to a fixed pose subset drawn identically from each of the twelve runs: each run's top 500 poses by its own published score, top 500 by RMSD to native, and 2000 additional poses drawn at random with a fixed seed (20260730), yielding 2742–2961 scored poses per run after de-duplication (108 run × function combinations in

total). The tenth function, SIPPER, was not evaluated: lightdock/scoring/sipper/driver.py in this package version references the name DEFAULT_CONTACT_RESTRAINTS_CUTOFF without importing it, and raised the same exception on all twelve runs. Rank correlation between score and RMSD (Spearman's ρ) was computed both on this full re-scored subset per run and, to obtain an unbiased estimate uncontaminated by the deliberately RMSD-selected half of the subset, on a further seeded random sub-sample of 2000 poses per run.

A restraint-free re-score was additionally computed for all twelve runs by forcing LightDock's restraint-satisfaction multiplier to zero regardless of how many restraints a run actually used, isolating the contribution of that term to each run's originally published ranking (Section 2.3). This uses the same fixed pose subset as the rescoring panel above.

5.4. Deep-Learning Co-Folding

Co-folding used Boltz-2 (v2.2.1) [14] and, as a temporal control at anchor A2 only, Boltz-1 [15], run with identical inference settings (three recycling steps, 200 sampling steps, five diffusion samples per target, structural kernels disabled, no MSA subsampling). Input files specified only the receptor and toxin amino-acid sequences, with no structural template and no user-supplied restraints; the use_templates flag recorded in each run's logged hyperparameters is a model-capability setting and does not indicate that a template was supplied. Multiple-sequence alignments were generated automatically by Boltz's own MSA search; the resulting paired alignment for either toxin contained no aligned homologues beyond the query itself (depth of three sequences), while the corresponding receptor chains returned deep alignments (14,106 sequences for 9NX0's receptor chain, 6193 for 7EGR's), of which the model reads at most its default limit of 8192 sequences, so that for 9NX0 the receptor alignment actually used held 8192 sequences; the three-sequence toxin alignment is read as a single sequence. Each condition was sampled five times (five diffusion samples), and every sample is reported individually (Table S13) rather than only the single highest-confidence or lowest-RMSD model, because the model-internal confidence ranking of the same predicted poses did not reliably track ligand RMSD (Section 2.4; Table S13). The sampling steps (200) and recycling steps (three) are the defaults of the software version used; five diffusion samples were drawn per target, where the default is one, so that every sample could be reported individually. A sensitivity analysis changed one setting at a time at both anchors, with the same seed: 400 sampling steps, six recycling steps, and 25 samples (Table S27). The 25 samples were drawn as five runs of five samples with five seeds (the reference seed and four others), not as one batch, because a batch of 25 was not attempted. Sampling steps and recycling steps were not raised beyond these values, and each of those two conditions was run with one seed.

Predicted poses were superposed onto their deposited receptor in two ways, both reported: a whole-receptor Cα superposition, and an interface-local superposition restricted to receptor Cα atoms within 10 Å of the native ligand—following exhaustive enumeration of the homomeric chain pairing between predicted and deposited receptor chains, since naive sequential chain assignment inflates the whole-receptor RMSD more than twenty-fold at both anchors by mis-pairing chemically identical subunits. Ligand RMSD was computed after each superposition on the same matched-atom correspondence used for the intrinsic-floor calculation (Section 5.1), and fnat (the fraction of the deposited complex's native receptor–ligand residue–residue contacts, a residue pair counting as in contact if any heavy-atom pair lies within 5 Å, reproduced by the predicted pose) was computed independently. The rigid-body conformational floor computed in Section 5.1 does not apply to co-folding output, because Boltz predicts the ligand's conformation together with its pose rather than docking a fixed rigid conformer.

The species and negative-control experiments of Section 2.7 were run in both Boltz-2 and, as an architecturally independent second co-folder, Chai-1 (v0.6.1) [19]. Chai-1 received the identical input files, converted only in container format, with no re-derivation of any sequence, and inference settings matched as closely as the two interfaces allow: five diffusion samples, 200 diffusion timesteps, three trunk recycles, the same MSA server and the same seed. Chai-1's ESM embedding stage has no Boltz counterpart and was left at its default; this is recorded rather than absorbed. Because no experimental structure of α9α10 exists, no RMSD to a native pose is reported for those targets, and the species comparison rests entirely on the relative ordering of the two orthologues within one model. Pose reproducibility was quantified identically for both models as the pairwise ligand all-heavy-atom RMSD among replicate samples after superposing each sample's receptor onto the first sample's receptor, tested by exact permutation over all 126 distinct labellings of the ten samples, so that the two models' p-values share the same attainable floor of $1/126 = 0.0079$. Scalar confidence metrics were compared by the same exact test. Because a pose-reproducibility statistic defined this way depends on the frame the poses are compared in, the same test was recomputed under three further superposition schemes—the α9 subunit alone, the α10 subunit alone, and both chains truncated to the extracellular domain—and the α9-only result is reported alongside the primary one in Section 2.7 wherever the two disagree on nominal significance.

Pose mode for each of these 40 samples (Section 2.7) was classified from receptor–ligand heavy-atom contacts (5.0 Å cutoff): a sample was scored α9(+)/α10(−) if chain A (α9) made at least three contacts within its own principal-face loop A/B/C span (construct positions 93–201) and chain B (α10) either made at least two contacts within its complementary-face loop D span (construct positions 55–61) or made no contact at all; α10(+)/α9(−) if the same two criteria held with the two chains' roles exchanged; and unresolved if neither pattern was met. This is a coarse structural classifier operating on already-generated poses, not a fitted pose-clustering or re-docking procedure, and is reported as such. Because every input specifies exactly one α9 chain and one α10 chain, the α9(+)/α9(−) interface also reported for α9α10 [20] cannot be formed in these models and was therefore not an available classification outcome; the classifier distinguishes the only two arrangements this construction can generate. The permutation test on ipTM comparing the two modes was a Monte Carlo test (200,000 random relabellings of the pooled two-mode set, mean-difference statistic, seed fixed for reproducibility), distinct from the exact enumeration used elsewhere in this section for the five-sample species comparisons, because the two mode groups (24 and 12 samples) are too large for exhaustive enumeration.

Receptor sequences for these species and negative-control experiments were the human and rat α9 and α10 subunits (UniProt Q9UGM1, P43144, Q9GZZ6 and Q9JLB5 respectively; precursor lengths 479, 479, 450 and 447 residues, accessions verified live against the UniProt REST API). Each subunit's construct begins at the same precursor position in both species—position 26 for α9, position 25 for α10—rather than at either species' own SignalP-predicted signal-peptide cleavage site, which UniProt places identically for α10 in both species (1–24) but differently for α9 (human 1–22, rat 1–25) despite the two α9 precursors being the same length; the extracellular-domain condition truncates both subunits at precursor position 237 in both species, and the full-length condition runs to each precursor's own C-terminus. Because construct numbering is accordingly a shared positional reference rather than either species' own mature-chain numbering, and differs by a further fixed offset from the "Thr56/Ile56" convention this study cites from Azam and McIntosh [18] for the same substitution, full accession, boundary and numbering-correspondence detail—including the offset itself—is given in Table S23 rather than restated here.

The three co-folding models serve different purposes. Boltz-2 is the primary model. Boltz-1 is used only as a temporal control: its stated training cutoff (30 September 2021) precedes the release of 7EGR (6 April 2022), so 7EGR cannot have been in its training data, whereas Boltz-2's cutoff (1 June 2023) falls after it. Chai-1 is used as an independently developed co-folder to test whether the confidence results of Section 2.7 belong to the method or to one model. The only difference in training data that this study uses is the pair of stated cutoffs; the architectural and training-procedure differences between the models are described in their original reports [14,15,19] and are not used in the interpretation. The design cannot separate an algorithmic improvement from a difference in training data. Boltz-1 and Boltz-2 differ in both, so a difference in their accuracy at A2 cannot be assigned to either; Boltz-1 supports one inference only: that its result at A2 is not memorisation of 7EGR. Separating the two factors would require models that differ in one of them, which this study does not have.

5.5. Data-Leakage Assessment

Because a co-folding result on a complex deposited close to the evaluated model's training period cannot be interpreted without knowing whether that period could have included the target, every co-folding result is reported together with the relevant training cutoff and target release date, determined as follows. Release dates were taken from the RCSB PDB entry metadata for 9NX0 and 7EGR. Boltz-2's training cutoff (1 June 2023) is taken verbatim from its preprint's stated data cutoff [14]; Boltz-1's (30 September 2021) from its own preprint [15]. A2 (7EGR, released 6 April 2022) falls before the Boltz-2 cutoff and is accordingly reported as a positive control rather than a held-out test. A1 (9NX0, released 4 February 2026) falls after the Boltz-2 cutoff, but same-complex homology to pre-cutoff structures was additionally checked, following the same-complex exclusion criterion used by AlphaFold3 [12]: every pre-cutoff PDB entry containing either ImI or a human α7-nAChR-family receptor chain was identified by an RCSB sequence-similarity search, and its sequence identity to each of 9NX0's two chains was recorded independently (receptor-chain identity from the RCSB search/v2 sequence-similarity service; toxin-chain identity by direct pairwise alignment, since ImI is too short for the service's own minimum query length). One pre-cutoff entry, 6HY7 (22 May 2019, a human α9-nAChR extracellular domain in complex with a different α-conotoxin, RgIA), exceeds 40% sequence identity to 9NX0 on both chains simultaneously, the threshold AlphaFold3 uses to exclude a training-set interface as same-complex with a test target; A1 is accordingly rated a MEDIUM leakage risk rather than LOW. The two anchors' leakage status is never pooled: results at A1 and A2 are reported and interpreted separately throughout (Section 2.4). The ratings follow one rule. An anchor released before the model's training cutoff is rated HIGH. An anchor released after the cutoff is rated LOW if no pre-cutoff entry exceeds 40% sequence identity to it on both the receptor and the toxin chain simultaneously, and MEDIUM if one does. A MEDIUM or HIGH rating means that the accuracy reported at that anchor cannot be read as evidence of generalisation to unseen interfaces.

5.6. Statistical Analysis

Because the reported best-RMSD for a run is a minimum over swarms—an order statistic sensitive to sample size—and the number of swarms per run differs between anchors (390–392 at A1, 132–134 at A2), the primary between-rung comparison used an equal-N bootstrap correction: each run's per-swarm-best array was resampled without replacement to the common minimum swarm count (N = 132) 20,000 times, with a fixed seed (20260730), and the minimum retaken within each resample (Table S2). A two-sided Mann–Whitney U test on the raw per-swarm-best RMSD distribution, computed independently

within each anchor (Table S3), is reported alongside this as a secondary check; it treats swarms as independent samples despite their being a tiling of one receptor surface, so it is not the primary basis for any between-rung claim in this study. Spearman's ρ was used throughout for score–RMSD rank correlation (Section 5.3). All statistical computation, pose algebra, and rescoring were implemented in Python 3; code and the complete set of run outputs underlying every reported number are available as described in the Data Availability Statement.

5.7. Interface Contact Analysis

Native receptor–ligand contacts were enumerated from the deposited coordinates at a 5.0 Å heavy-atom cutoff, a residue pair counting as in contact if any heavy-atom pair falls within that distance; this is the definition used for fnat throughout (Section 5.4), so docking and co-folding are compared on one contact definition. Contacts were assigned to the conventional loops of the nicotinic binding site from the residue numbering of each deposited structure. Interaction types were classified geometrically, without an angular term, since the toxin chains of both deposited entries carry no hydrogen atoms: a hydrogen bond as an N/O/S donor–acceptor pair within 3.5 Å, an ionic interaction as a charged Arg/Lys/His nitrogen within 4.0 Å of an Asp/Glu carboxylate oxygen, and a hydrophobic contact as an apolar carbon pair within 4.5 Å. These are inventories of geometrically compatible pairs rather than validated interactions, and are reported as such. Hydrogen-bond counts are given both including and excluding pairs whose acceptor is a disulfide-bonded cysteine sulfur, and ionic-interaction counts at both 4.0 and 4.5 Å, because in this interface a single 0.34 Å change in the cutoff alters the count. One pair may satisfy two definitions and is then counted under both, so the type counts are not additive. Docked poses were reconstructed from the recorded per-glowworm transforms and compared on the same atom correspondence used elsewhere (Section 5.1). For each run, three poses are reported—best by RMSD, first by score, and best by fnat—because these are generally different poses and only the second is what a user receives.

5.8. CAPRI Classification, DockQ, and the Rigid-Placement Ceiling

Models were scored on the CAPRI scale using fnat, ligand RMSD after receptor superposition (L-RMSD, backbone atoms) and interface backbone RMSD (I-RMSD, interface defined in the native structure at a 10 Å heavy-atom cutoff), combined into DockQ by its published definition [22]. The implementation used here reproduces the maintained reference implementation (DockQ v2.1.3) exactly across all evaluated model × mode comparisons, to 0.000 in DockQ, fnat and I-RMSD and 0.001 Å in L-RMSD. The receptor subunit pair was treated as one rigid body; for co-folded models, whose two sequence-identical receptor chains may map onto the deposited chains in either order, the assignment maximising DockQ was used. Because the α-conotoxins are twelve-to-thirteen-residue peptides, every model was additionally classified under the CAPRI peptide-target variant, which alters both the interface definition and the thresholds, and both classifications are reported.

Two relationships between these metrics and the ligand RMSD used elsewhere in this work are worth stating. First, the docking runs hold the receptor at its deposited coordinates and the receptor file used for docking is byte-identical to the one used to build the native complex, so the CAPRI receptor superposition is the identity (Kabsch fit RMSD 0.0000 Å for every evaluated pose) and CAPRI L-RMSD is numerically the interface-frame ligand backbone RMSD reported in Section 2.2; the all-atom values quoted elsewhere differ from L-RMSD by atom basis alone. No such identity holds for co-folded models, which are returned in an arbitrary global frame and are therefore compared only after superposition. Second, neither fnat nor DockQ is monotone in ligand RMSD, so evaluating a conformer at its RMSD-

optimal superposition yields a lower bound rather than a maximum. The rigid-placement ceilings in Section 2.6 were therefore obtained by maximising fnat and DockQ directly over the six rigid-body degrees of freedom, by multi-start stochastic search (3–5 × 10^8 screened placements, 194–297 independent restarts, a no-improvement certificate over 6.76 × 10^6 local placements, and deterministic one-dimensional sweeps of all six degrees of freedom). An independent replicate with a disjoint random stream reproduced all four maxima exactly, and a targeted search seeded on the co-folded poses did not exceed them. Every maximum occurs at a ligand RMSD above the RMSD minimum, which is a direct demonstration of the non-monotonicity. These values remain lower bounds on the true ceiling: the only rigorous upper bound available, from the distance geometry of pairwise-compatible contacts, is near-trivial (1.000 at A1, 0.982 at A2) and does not constrain them, so the A2 ordering in Section 2.6 rests on these convergence controls rather than on proof.

5.9. Close-Contact Census of Unbound Conformers

To test whether a rigid unbound conformer can occupy the bound site at all, each deposited NMR model of each toxin was superposed onto its deposited bound counterpart by Kabsch fit on the matched-atom correspondence of Section 5.1, and intermolecular heavy-atom distances to the deposited receptor were then counted. All measurements were made in the deposited frame; the receptor file is byte-identical to the deposited coordinates, so no frame conversion was applied to either body. The deposited bound ligand itself was put through the identical procedure as a control, and must and does come out with no pair below 2.5 Å at either anchor. Counts are reported at 2.0, 2.5, 3.0 and 4.0 Å together with the closest approach. This census establishes only that a rigid unbound conformer cannot be placed without overlap; it does not establish that a flexible protocol would recover the native pose.

5.10. Extracellular-Domain Control at Anchor A1

A1 presents both the extracellular and transmembrane domains (6326 receptor heavy atoms), while A2 presents an extracellular domain only (3276), so the two anchors' docking results are not obtained on comparable surfaces. To separate receptor extent from anchor identity, the full three-rung ladder was repeated at A1 on a truncated receptor. The domain boundary was not asserted from sequence annotation: the extracellular domain was defined as those A1 residues that structurally align to the AChBP chains of A2, AChBP being an extracellular-domain protein by construction, so that the truncated A1 corresponds to what A2 already is. The alignment used the same routine that produced the cross-donated restraint sets (190 and 189 residues aligned on the two chains, refined RMSD 1.429 and 1.459 Å, reproducing the values recorded for those chains in the cross-donation audit). The resulting receptor carries 3145 heavy atoms, within 4% of A2's. All 28 native-contact restraint residues survive the truncation, so the NATIVE rung is restrained by the same footprint as at full length. Ligand, apo conformer, glowworm count, surface density, swarm radius, starting-point seed and scoring function were copied from the corresponding full-length runs, and the ligand file is byte-identical to the published A1 runs'; replicates differ only in the GSO search seed, as at full length. Swarm counts were matched as closely as the restraint counts permit (390 BLIND, 377 XDON, 336 NATIVE), and where they differ, the difference disadvantages the rung that wins.

5.11. Loop C Displacement Between Unbound and Toxin-Bound Receptors

To establish the direction and the magnitude of the receptor rearrangement that a rigid docking protocol holds fixed, loop C of each anchor's principal (+)-face subunit was compared with the same loop in deposited unbound structures of the same protein;

no homologue was substituted for either receptor. The unbound set was PDB 7EKI, 9J37 and 8V89 for human α7 (five subunits each, fifteen in total) and PDB 2BYN for *Aplysia* AChBP (five subunits), compared against chain D of 9NX0 and chain H of 7EGR, the toxin-bound principal-face chains of the two anchors. Loop C was defined structurally rather than from sequence annotation, as the sequence-adjacent Cys–Cys pair plus eight residues on either side (eighteen residues: 181–198 around Cys189–Cys190 at A1, 199–216 around Cys207–Cys208 at A2), and the loop tip was taken as the residue of maximal Cα displacement. This definition uses the Cys pair only as a sequence landmark and does not require the two residues to be covalently bonded, which matters at A2: in 7EGR's chain H the two sulfurs are 4.412 Å apart, and the depositor declares no disulfide there (Section 2.1). Each unbound subunit was superposed onto the bound subunit by Cα fit on the subunit core with loop C excluded from the fit (371–372 fitted atoms and 0.90–1.26 Å fit RMSD at A1; 184–186 atoms and 0.38–0.57 Å at A2), and the distance from the loop C tip Cα to the centroid of the deposited bound toxin was then measured for the bound and the superposed unbound subunit in the same frame. Two further superposition schemes were computed as controls—a 30 Å local core (47–50 fitted atoms) and the two flanking β-strands alone (18–20 atoms)—and returned loop C tip displacements of the same magnitude as the subunit-core fit (1.80–3.17 Å at A1, 2.99–5.93 Å at A2, against 1.85–3.49 Å and 3.25–5.89 Å respectively). This measurement compares deposited states; it does not simulate the binding transition and carries no claim about the order in which conformational change and binding occur.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/toxins18110498/s1, Figure S1: Score–RMSD rank correlation on the full re-scored subset; Figure S2: Score–RMSD funnels for the seven scoring functions not shown in Figure 3; Figures S3–S6; Tables S1–S28. The machine-readable source files underlying Tables S1–S28 (docking, rescoring and co-folding outputs) are in the code repository given in the Data Availability Statement.

Author Contributions: Conceptualization, S.J. and L.-S.K.; methodology, S.J.; software, S.J.; validation, S.J. and L.-S.K.; formal analysis, S.J.; investigation, S.J.; data curation, S.J.; writing—original draft preparation, S.J.; writing—review and editing, S.J. and L.-S.K.; visualisation, S.J.; supervision, L.-S.K.; project administration, L.-S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available at https://github.com/leeseungkwon804/conotoxin-nachr-benchmark (accessed on 12 November 2026).

Acknowledgments: During the preparation of this manuscript, the authors used Claude (Anthropic) (Claude Code version 2.1.322.) for analysis scripting, figure generation, literature searching and drafting of the text. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

nAChR	Nicotinic acetylcholine receptor
AChBP	Acetylcholine-binding protein
ECD	Extracellular domain

TMD	Transmembrane domain
RMSD	Root-mean-square deviation
PDB	Protein Data Bank
GSO	Glowworm swarm optimisation
MSA	Multiple sequence alignment
fnat	Fraction of native contacts recovered
DockQ	Continuous docking quality score
CAPRI	Critical Assessment of PRediction of Interactions

References

1. Staats, P.S.; Yearwood, T.; Charapata, S.G.; Presley, R.W.; Wallace, M.S.; Byas-Smith, M.; Fisher, R.; Bryce, D.A.; Mangieri, E.A.; Luther, R.R.; et al. Intrathecal ziconotide in the treatment of refractory pain in patients with cancer or AIDS: A randomized controlled trial. *JAMA* **2004**, *291*, 63–70. [CrossRef] [PubMed]
2. Tae, H.-S.; Hung, A.; Clark, R.J.; Adams, D.J. Molecular determinants of the selectivity and potency of α-conotoxin Vc1.1 for human nicotinic acetylcholine receptors. *J. Biol. Chem.* **2025**, *301*, 108017. [CrossRef] [PubMed]
3. Puillandre, N.; Bouchet, P.; Duda, T.F., Jr.; Kauferstein, S.; Kohn, A.J.; Olivera, B.M.; Watkins, M.; Meyer, C. Molecular phylogeny and evolution of the cone snails (Gastropoda, Conoidea). *Mol. Phylogenet. Evol.* **2014**, *78*, 290–303. [CrossRef] [PubMed]
4. Kaas, Q.; Westermann, J.-C.; Halai, R.; Wang, C.K.L.; Craik, D.J. ConoServer, a database for conopeptide sequences and structures. *Bioinformatics* **2008**, *24*, 445–446. [CrossRef] [PubMed]
5. Kaas, Q.; Yu, R.; Jin, A.-H.; Dutertre, S.; Craik, D.J. ConoServer: Updated content, knowledge, and discovery tools in the conopeptide database. *Nucleic Acids Res.* **2012**, *40*, D325–D330. [CrossRef] [PubMed]
6. Li, R.; Hasan, M.M.; Wang, D. In Silico Conotoxin Studies: Progress and Prospects. *Molecules* **2024**, *29*, 6061. [CrossRef] [PubMed]
7. Leffler, A.E.; Kuryatov, A.; Zebroski, H.A.; Powell, S.R.; Filipenko, P.; Hussein, A.K.; Gorson, J.; Heizmann, A.; Lyskov, S.; Tsien, R.W.; et al. Discovery of peptide ligands through docking and virtual screening at nicotinic acetylcholine receptor homology models. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E8100–E8109. [CrossRef] [PubMed]
8. Gulsevin, A.; Meiler, J. An Investigation of Three-Finger Toxin–nAChR Interactions through Rosetta Protein Docking. *Toxins* **2020**, *12*, 598. [CrossRef] [PubMed]
9. Roel-Touris, J.; Bonvin, A.M.J.J.; Jiménez-García, B. LightDock goes information-driven. *Bioinformatics* **2019**, *36*, 950–952. [CrossRef] [PubMed]
10. Bhattacharjee, B.; Noviello, C.M.; Rahman, M.M.; Mayer, J.P.; Gajewiak, J.; McIntosh, J.M.; Hibbs, R.E.; Stowell, M.H.B. Shape-shifting conotoxins reveal divergent pore-targeting mechanisms in nicotinic receptors. *Structure* **2026**, *34*, 463–470.e4. [CrossRef] [PubMed]
11. Wang, X.Q.; Pan, S.; Fan, Y.X.; Xue, Y.; Zhu, X.P.; Luo, S.L. Co-Crystal Structure of Ac-AChBPP in Complex with RgIA [sic; Official Deposited Title]. PDB Entry 7EGR. 2022. Available online: https://www.rcsb.org/structure/7EGR (accessed on 12 November 2026).
12. Abramson, J.; Adler, J.; Dunger, J.; Evans, R.; Green, T.; Pritzel, A.; Ronneberger, O.; Willmore, L.; Ballard, A.J.; Bambrick, J.; et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493–500. [CrossRef] [PubMed]
13. Jiménez-García, B.; Roel-Touris, J.; Romero-Durana, M.; Vidal, M.; Jiménez-González, D.; Fernández-Recio, J. LightDock: A new multi-scale approach to protein–protein docking. *Bioinformatics* **2018**, *34*, 49–55. [CrossRef] [PubMed]
14. Passaro, S.; Corso, G.; Wohlwend, J.; Reveiz, M.; Thaler, S.; Somnath, V.R.; Getz, N.; Portnoi, T.; Roy, J.; Stark, H.; et al. Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. *bioRxiv* **2025**, 2025.06.14.659707. [CrossRef] [PubMed]
15. Wohlwend, J.; Corso, G.; Passaro, S.; Getz, N.; Reveiz, M.; Leidal, K.; Swiderski, W.; Atkinson, L.; Portnoi, T.; Chinn, I.; et al. Boltz-1: Democratizing Biomolecular Interaction Modeling. *bioRxiv* **2024**, 2024.11.19.624167. [CrossRef] [PubMed]
16. Quiram, P.A.; Sine, S.M. Structural elements in α-conotoxin ImI essential for binding to neuronal α7 receptors. *J. Biol. Chem.* **1998**, *273*, 11007–11011. [CrossRef] [PubMed]
17. Ellison, M.; Feng, Z.-P.; Park, A.J.; Zhang, X.; Olivera, B.M.; McIntosh, J.M.; Norton, R.S. α-RgIA, a novel conotoxin that blocks the α9α10 nAChR: Structure and identification of key receptor-binding residues. *J. Mol. Biol.* **2008**, *377*, 1216–1227. [CrossRef] [PubMed]
18. Azam, L.; McIntosh, J.M. Molecular basis for the differential sensitivity of rat and human α9α10 nAChRs to α-conotoxin RgIA. *J. Neurochem.* **2012**, *122*, 1137–1144. [CrossRef] [PubMed]
19. Chai Discovery; Boitreaud, J.; Dent, J.; McPartlon, M.; Meier, J.; Reis, V.; Rogozhnikov, A.; Wu, K. Chai-1: Decoding the molecular interactions of life. *bioRxiv* **2024**, 2024.10.10.615955. [CrossRef]

20. Zouridakis, M.; Papakyriakou, A.; Ivanov, I.A.; Kasheverov, I.E.; Tsetlin, V.; Tzartos, S.; Giastas, P. Crystal structure of the monomeric extracellular domain of α9 nicotinic receptor subunit in complex with α-conotoxin RgIA: Molecular dynamics insights into RgIA binding to α9α10 nicotinic receptors. *Front. Pharmacol.* **2019**, *10*, 474. Corrigendum in *Front. Pharmacol.* **2022**, *13*, 980202. https://doi.org/10.3389/fphar.2022.980202. [CrossRef] [PubMed]
21. Xu, S.; Feng, Q.; Qiao, L.; Wu, H.; Shen, T.; Cheng, Y.; Zheng, S.; Sun, S. Benchmarking all-atom biomolecular structure prediction with FoldBench. *Nat. Commun.* **2025**, *17*, 442. [CrossRef] [PubMed]
22. Basu, S.; Wallner, B. DockQ: A Quality Measure for Protein–Protein Docking Models. *PLoS ONE* **2016**, *11*, e0161879. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.