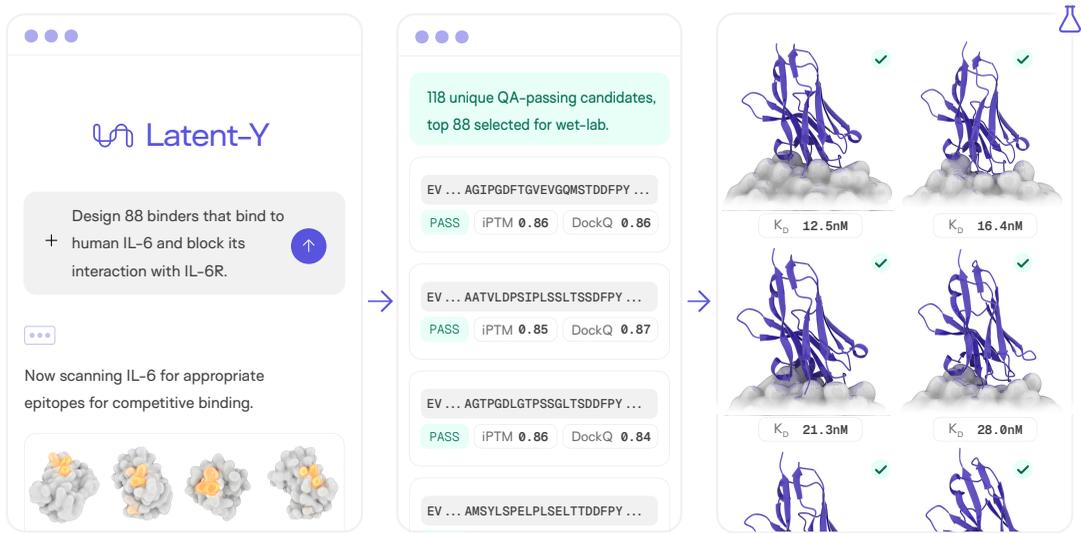


# Latent-Y: A Lab-Validated Autonomous Agent for De Novo Drug Design

Latent Labs Team

London, UK & San Francisco, USA

23 March 2026



Starting from a text prompt Latent-Y autonomously designs lab-ready antibodies and peptides.

## Abstract

Drug discovery relies on iterative expert workflows that are slow to parallelize and difficult to scale. Here we introduce Latent-Y, an AI agent that autonomously executes complete antibody design campaigns from text prompts, covering literature review, target analysis, epitope identification, candidate design, computational validation, and selection of lab-ready sequences. Latent-Y is integrated into the Latent Labs Platform, where it operates in the same environment as drug-discovery experts with access to bioinformatics tools, biological databases, and scientific literature. The agent can run fully autonomously end-to-end, or collaboratively, where researchers review progress, provide feedback, and direct subsequent steps. Candidate antibodies are generated using Latent-X2, our frontier generative model for drug-like antibody design. We demonstrate the agent's capability across three distinct campaign types: epitope discovery guided by therapeutic specifications, cross-species binder design, and autonomous design from a scientific publication targeting human transferrin receptor for blood-brain barrier crossing. Across nine targets, Latent-Y produced lab-confirmed nanobody binders against six, achieving a 67% target-level success rate with binding affinities reaching the single-digit nanomolar range, without human filtering or intervention. In user studies, experts working with Latent-Y completed design campaigns 56 times faster than independent expert time estimates, compressing weeks of work into hours. Because Latent-X2 is a general-purpose atomic-level model for biologics design, the same agent architecture naturally extends to macrocyclic peptide and mini-binder design campaigns, broadening autonomous discovery across therapeutic modalities. Latent-Y is available to selected partners at [platform.latentlabs.com](https://platform.latentlabs.com).

## 1. Introduction

Frontier AI models such as Latent-X2 [1] and other molecular design models [2, 3] have recently demonstrated the feasibility of zero-shot biologics design, directly producing antibody and peptide candidates that have drug-like properties and advancing prior work in protein structure prediction [4, 5] and generative protein design [6–14].

In a world in which we can now design lab-ready molecules on the computer, the rate-limiting step for early drug discovery is no longer finding a candidate molecule for a drug program. The rate-limiting step instead becomes the bandwidth of drug-discovery organizations and access to the PhD-level domain expertise required to do so at scale.

To address this bottleneck, we developed Latent-Y, an agentic system for *de novo* drug design, capable of designing antibodies, peptides, and mini-binders from text descriptions of drug design goals. Latent-Y directly addresses the exploration and scale bottleneck by delivering autonomous drug design research that can be executed in parallel. We demonstrate, for the first time, that an autonomous agent can design lab-confirmed *de novo* antibody binders purely from text descriptions of goals and requirements, without requiring manual intervention in the design process. Latent-Y-designed nanobodies display strong binding affinities reaching single-digit nanomolar ranges, with highly competitive target-level success rates, hitting 6 of 9 targets attempted in the lab.

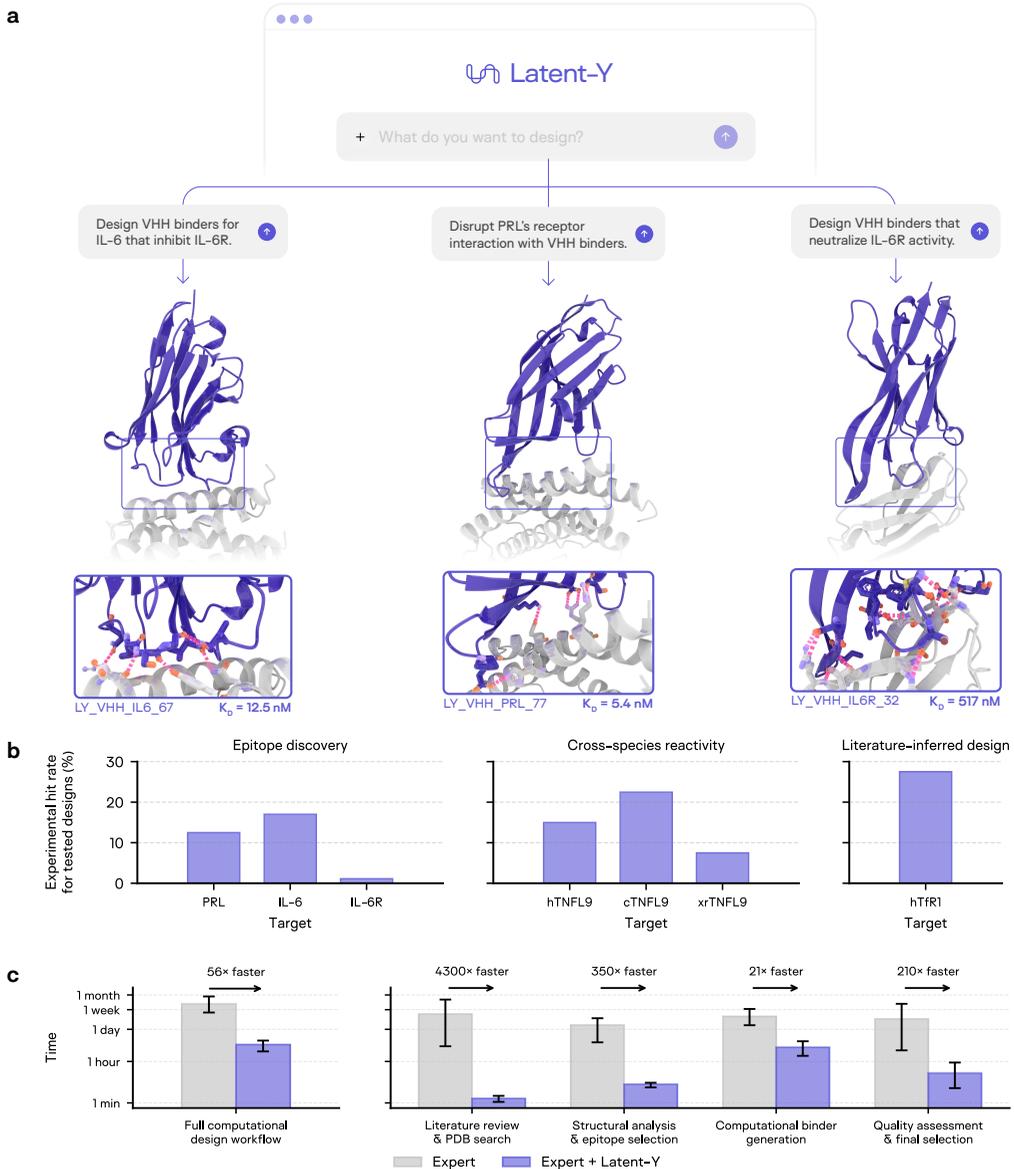
Where Latent-X2 reasons at the atomic level to design precise molecular interactions, Latent-Y operates in the same environment as human experts, applying expert-level reasoning to navigate from research objective to lab-ready candidates. It queries molecular and literature databases, orchestrates Latent-X2 to generate and score candidates, and employs bioinformatics tools to analyse the sequence and structure of targets, epitopes, and designed molecules. It intelligently explores a large combinatorial design space. It tests different hypotheses, analyses results, and integrates insights across parallel runs, plans appropriate next steps, and works towards user-provided goals, such as achieving an intended functional mechanism or producing the requested number of lab-ready designs.

In this work, Latent-Y is run both fully autonomously and collaboratively. The majority of campaigns were executed end to end without human intervention, while the cross-species campaign demonstrates tight human-agent collaboration, with the researcher steering strategy and the agent adapting in real time. Both modes are supported across the full spectrum of expert involvement. We show that Latent-Y significantly accelerates the work of protein design experts, reducing the time required for expert research and computational workflows from weeks to hours. This acceleration compounds when running Latent-Y instances at scale and in parallel, for example to find development candidates across multiple pre-clinical drug programs simultaneously.

### Our main contributions are:

1. The first autonomous agent for *de novo* biologics design, delivering lab-ready sequences from text input.
2. Lab-validated *de novo* nanobody design from text, with a 67% target-level success rate across 9 targets.
3. Latent-Y-designed *de novo* nanobodies with single-digit nanomolar binding affinities, confirmed in the lab.
4. A 56-fold acceleration of expert-led design campaigns with Latent-Y versus without, measured against independent expert time estimates.
5. Lab-validated cross-species binder design via Latent-Y-generated custom generative code.

**Availability.** Latent-Y is integrated and available on the Latent Labs Platform at [platform.latentlabs.com](https://platform.latentlabs.com).



**Fig. 1 | Latent-Y autonomously designs nanomolar-affinity antibodies from text prompts, accelerating expert workflows by over 50-fold. (a)** Three independent VHH design campaigns, each initiated from a single natural-language prompt targeting IL-6, PRL, and IL-6R respectively. Designed structures of the top binder–target complex are shown alongside binding interface detail, with insets highlighting designed non-covalent interactions at the binding interface (pink dashed lines). All designs were experimentally characterized by surface plasmon resonance (SPR). **(b)** Per-target experimental hit rates across all successful targets. cTNFL9, cynomolgus TNFL9; hTNFL9, human TNFL9; xrTNFL9, cross-reactive TNFL9. **(c)** Estimated time required to complete a full computational protein design campaign, comparing expert alone versus expert assisted by Latent-Y. Expert baseline times were obtained by polling independent protein designers across academia and industry ( $n = 10$ ); Latent-Y-assisted times reflect the average across agent-assisted runs ( $n = 5$ ). Error bars indicate minimum and maximum values. **(d)** Time breakdown by major design stage, including literature review and Protein Data Bank (PDB) search, structural analysis and epitope selection, computational binder generation, and quality assurance and final selection. Error bars indicate minimum and maximum values.

## 2. Results

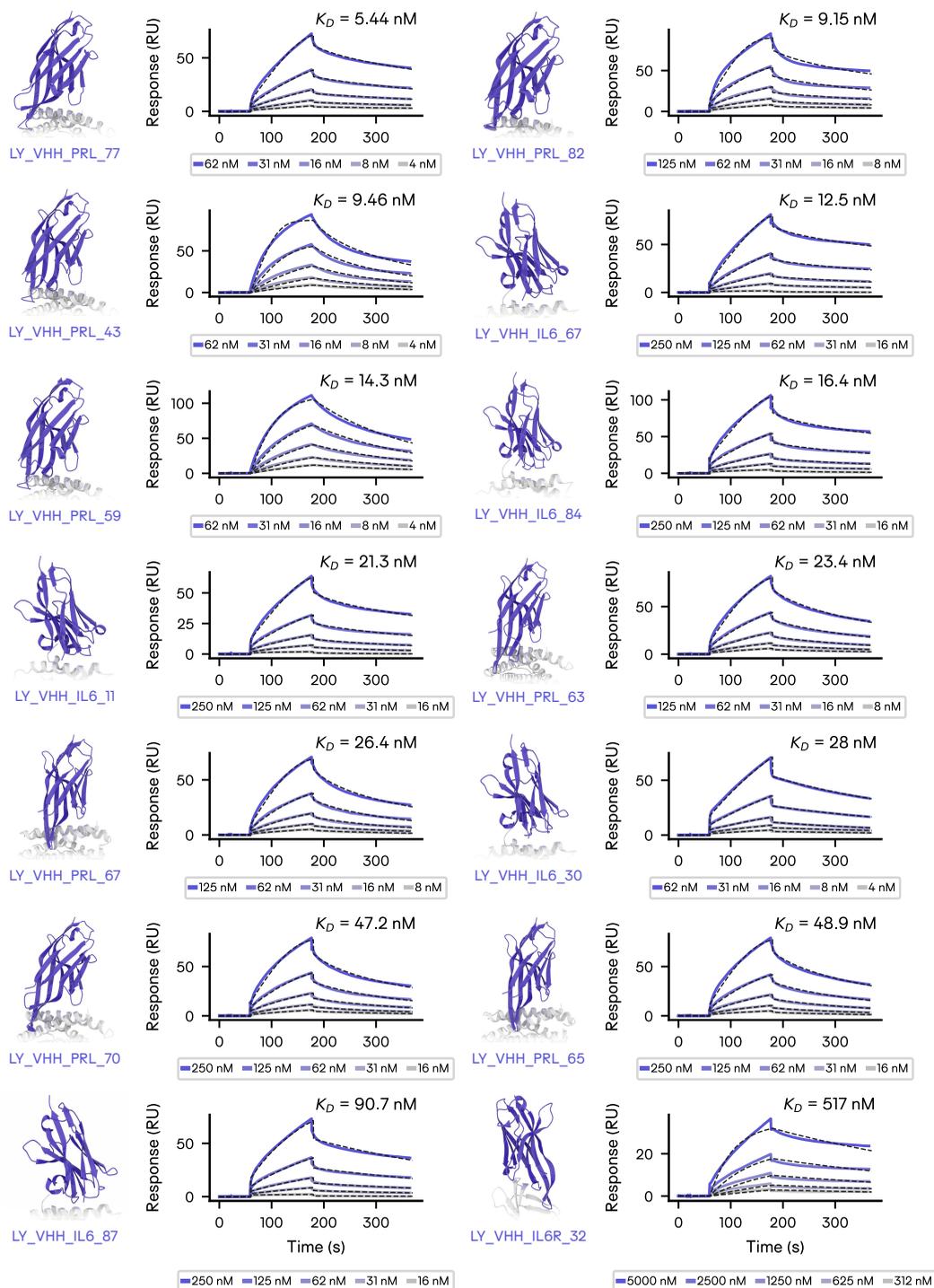
We evaluate Latent-Y across three antibody design settings: epitope discovery campaigns targeting IL-6 [15], IL-6R [16], PRL [17], IL-33 [18], TNF $\alpha$  [19], and SC2RBD [20], with the goal of therapeutic inhibition; cross-species reactivity campaigns against TNFL9 [21]; and literature-inferred design campaigns targeting hTfR1 [22]. Across these campaigns, Latent-Y achieves a target-level success rate of 67%, with per-target hit rates ranging from 1% to 28% of tested sequences, and binding affinities reaching the single-digit nanomolar range.

### 2.1. Latent-Y is a wet-lab-validated agent for drug design at scale

Latent-Y is natively integrated into the Latent Labs Platform and can be accessed entirely through a browser, steered by natural language. A researcher can provide a therapeutic objective as free text, a structured work plan specifying targets and constraints, or a scientific publication from which the agent autonomously extracts target identities, biological context, and design constraints, as shown in Fig. 1a.

Drug design problems are inherently underspecified at the outset: the right epitope, the appropriate structural context, and the relevant constraints often only become clear during the research process itself. Latent-Y is designed to reduce this ambiguity. Upon receiving a design objective, it consults scientific literature and databases to build biological context, identifies appropriate target structures, and characterizes candidate epitopes against functional criteria. Unlike fixed generative pipelines, it accepts and reasons about arbitrary research constraints such as modality, intended function, species requirements, structural considerations, and computational metrics and predictions, translating high-level goals into concrete design decisions. It spawns targeted computational experiments using Latent-X2, reasons about results, and doubles down on the most productive directions, adjusting generation parameters, modifying epitope constraints, or re-routing to alternative modalities as needed. Where standard capabilities are insufficient, the agent can generate custom computational approaches from natural language descriptions, extending its own toolkit to address novel design challenges, as described in Sec. 2.3. Outputs annotated with the agent’s reasoning are produced at each stage for researcher inspection and, where needed, manual override. Before finalizing candidates, the agent performs quality assurance on the designs, including clustering for diversity, sequence similarity searches against external databases, and sequence liability analyses, selecting a final set of lab-ready designs that satisfy the user-defined objectives. Representative condensed reasoning traces illustrating fully autonomous and collaborative human-agent workflows are shown in Fig. 3 and Fig. 5 respectively.

At the level of an individual campaign, Latent-Y compresses the full computational antibody design workflow approximately 56-fold, on average reducing two weeks of expert effort to five hours, as shown in Fig. 1c. When a researcher runs multiple campaigns in parallel, these gains compound further still, multiplying the effective throughput of a single researcher beyond what any individual could achieve sequentially. The largest accelerations arise in the reasoning-intensive stages: literature review and PDB analysis ( $\sim 4,300\times$ ) and structural analysis and epitope selection ( $\sim 350\times$ ), as detailed in Fig. 1d. Baseline expert times were estimated by polling independent, PhD-level protein designers across academia and industry, with a median of 9.8 years of relevant experience, as described in Sec. D. These gains compound further at scale, with a single researcher able to complete in under a day computational work that would previously have taken several months, shifting drug design from a sequential expert workflow to a scalable parallel process limited by scientific ideas rather than time.



**Fig. 2 | Biophysical characterization of the best Latent-Y-designed VHHs against IL-6, IL-6R, and PRL, with affinities reaching the single-digit nanomolar range.** Top-performing *de novo* VHH binders ranked by binding affinity, with corresponding designed bound structures and SPR response curves. Binding affinities were measured using five analyte concentrations with kinetic fitting to determine  $K_D$ , see Sec. A.3. Reported  $K_D$  values span the nanomolar range, with lower  $K_D$  values corresponding to stronger binding.

## 2.2. Latent-Y enables autonomous design of low nM-affinity antibodies

In the epitope discovery setting, we evaluate Latent-Y by running fully autonomous *de novo* VHH design campaigns against six therapeutically relevant targets, experimentally validating the resulting sequences in the wet lab, see Fig. 2 and Tab. S1. All campaigns were specified via natural language prompts framed around inhibiting a natural protein–protein interaction to modulate downstream function, with prompts specifying only the high-level therapeutic objective and leaving all design decisions to the agent. Across all targets, Latent-Y selected epitopes consistent with the intended mechanism of action, indicating that the agent reliably identifies functionally relevant binding sites from high-level goals alone.

Successful campaigns produced experimentally confirmed VHH binders against three therapeutically relevant targets: IL-6 [23], an inflammatory cytokine; IL-6R [24], its receptor; and PRL [25], a hormone for which no antibody-bound structure exists in the PDB. Per-target hit rates ranged from 1% to 17%, as determined by one-point high-throughput surface plasmon resonance (HT-SPR) primary screening, as shown in Fig. 1b and detailed in Sec. A.2. Latent-Y generated VHH binders with binding affinities reaching the single-digit nanomolar range, with the highest affinity binders for PRL, IL-6, and IL-6R measuring 5.44 nM, 12.5 nM, and 517 nM, respectively, validated by five-point SPR, as shown in Fig. 2 and detailed in Sec. A.3. For every target, high-affinity binders were identified from fewer than one plate of designs. Across targets with multiple confirmed binders, successful designs spanned a range of Complementarity-Determining Region (CDR) lengths and antibody frameworks, targeting distinct hotspot residues and reflecting the diversity of solutions explored by the agent. Lab-validated sequences for the best binder per target are provided in Sec. C. We publish the corresponding designed structures on <https://platform.latentlabs.com>, accessible without sign-in.

The IL-6 campaign is illustrated in Fig. 3 as a representative, condensed reasoning trace of a successful campaign. Starting from a prompt specifying only the desired outcome, Latent-Y retrieved biological context from the literature and databases, identified an appropriate target structure, performed spatial reasoning over the target surface, selected candidate epitopes, and generated VHH binders using Latent-X2. Designed binders were iteratively triaged, with feedback from earlier cycles used to prioritize higher-yield strategies and propagate favourable features across subsequent cycles. This included prioritization of productive epitopes such as site II on IL-6 [26] and selection of antibody frameworks associated with high computational success rates. Through a final quality assurance, the agent determined when sufficient diversity and sequence quality had been achieved to satisfy the design objective. To our knowledge, this campaign produced the first experimentally validated fully *de novo* designed antibody binder against IL-6. While this trace illustrates one representative trajectory, the agent dynamically adapts each campaign based on intermediate results and user specifications, with every decision captured in the reasoning trace for full auditability.

## 2.3. Latent-Y designs cross-species antibodies via autonomous capability extension

To evaluate Latent-Y in a complex translational setting, we task the agent with designing cross-reactive VHH binders against the human and cynomolgus variants of TNF ligand superfamily member 9 (TNFL9), a co-stimulatory target under active clinical investigation in immuno-oncology. Cross-species reactivity is a common, but challenging, requirement in drug development: a candidate molecule must engage both the human target and its preclinical homologue to support toxicology studies and clinical dose selection.

This task combines several biological and practical challenges routinely encountered in early-stage drug discovery. The human and cyno orthologues diverge by 11 mutations (~5%). No empirical structure is available for the cyno variant; there are unresolved regions in the available crystal structure for the human variant; and TNFL9 functions as a trimer with a relatively flat binding surface. Beyond these biological complexities, no pre-built cross-species design capability is provided to the agent. Instead, Latent-Y is given privileged but bounded access to the Latent Labs Platform and tasked with developing its own custom generative method guided only by a one-line natural language description, shown in Fig. 5. A human expert works collaboratively with Latent-Y to provide high-level biological steering but does not otherwise intervene, demonstrating that the agent can be effectively directed without coding expertise. The campaign trace, condensed to its key moments for visual brevity, is shown in Fig. 5.

I want you to run a design campaign where you identify the optimal epitope region(s) on human IL-6 that blocks its interaction with IL-6R, and return back 100 passing VHHs targeting those epitope regions. I want you to select the 88 best VHH sequences from these to send to the wet-lab.

Before I begin, I have a few questions to ensure this campaign is set up optimally:

- 1. PDB Structure:** Do you have a structure to use for the IL-6/IL-6R complex?
- 2. Framework preference:** Do you have a VHH framework preference?
- 3. CDR loop length:** Do you have constraints on CDR loop lengths I should be aware of?
- 4. Diversity:** Do you have success criteria beyond the 100 passing / 88 selected target? (e.g. maximize diversity, etc)

1) Pick the most suitable. 2) Whatever works. 3) Stick close to native CDR lengths. 4) No.

Ok, let me plan out an approach. The IL-6/IL-6R interaction is well known so I should find an appropriate structure on the PDB, then

**Tool use**

PDB search and metadata analysis. 8 tool calls.

\*\*\*

Identify an appropriate IL-6/IL-6R complex PDB structure and study relevant metadata.

Run the hotspot researcher to select 3-5 distinct epitope candidates prioritised by competitive blocking potential.

Pilot five VHH frameworks across all epitopes (20 batches, 500 designs each).

Triage: Assess pass rates, drop non-productive configurations, and scale up the best-performing epitope-framework combinations with CDR3 length and hotspot variations until we reach 100 passing binders.

Filter for sequence liabilities, deduplicate, and select the top 88 for synthesis.

\*\*\*

**Tool use**

Interface detection. 2 tool calls.

\*\*\*

PDB search returns 7UWL as top hit — but metadata reveals it is an IL-17/IL-25 structure, not IL-6. IN26 is the receptor alone; 1ALU is IL-6 alone. Selecting 1P9M: the hexameric IL-6/IL-6R $\alpha$ /IL-6R $\beta$ (gp130) signalling complex, which contains both receptor interfaces needed for competitive epitope analysis. Based on the metadata, Chain B = IL-6 (target), Chain C = IL-6R $\alpha$ , Chain A = IL-6R $\beta$

**Tool use**

Interface detection. 2 tool calls.

\*\*\*

Excellent. Now I have a clear picture of both interfaces on IL-6:

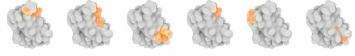
- Site I (IL-6/IL-6R $\alpha$  interface): B67-69, B74-75, B75, B179-180, B183
- Site II (IL-6/IL-6R $\beta$  interface): B24, B27-28, B31, B34, B110-111, B113-114, B117-118

Now let me run the hotspot researcher with the relevant context for a comprehensive analysis.

**Subagent call:** Hotspot researcher 72 min, 45 tool calls.

Structural quality analysis, scanning unresolved regions, scanning glycans, Literature and domain annotation cross-checks, UniProt cross-referencing, surface accessibility and secondary structure scans.

I'll systematically analyze IL-6 (chain B) in the structure for epitopes that would effectively disrupt IL-6R complex formation when occupied by a VHH, starting with the suggested sites.



\*\*\*

Unobserved residues at 1-18 and 47-49. Note the N-glycosylation sequon at B144 (NAS motif) and also the UniProt annotation of N-linked glycosylation at Asn73 (UniProt numbering for IL-6, P05231; B45). As expected, pocket scanning found no deep pockets — IL-6 presents relatively flat epitope surfaces typical of cytokines. — VHH CDR loops can still engage these, but I'll need to carefully examine the interface residues' accessibility, distance and chemistries to build specific hotspot candidates. Let me first focus on the Site I (IL-6R $\alpha$ ) interface residues.

\*\*\*

The C-terminal helix residues (B175-182) of Site I are mostly buried in the bound state; B179 (R) and B74 (F) have large delta\_ssa and can be targeted specifically. Now let me check Site II more carefully.

\*\*\*

The Site II interface is spread across two helices A & B [..] Let me check for a potential site III region [..]. The CD loop region is important for IL-6R $\beta$  signalling complex assembly and therefore also a viable target. Let me check this region.

\*\*\*

**Summary:** I identified 5 viable regions that can disrupt IL-6/IL-6R complex formation on chain B. In order of preference:

- 1. Site I core** — BC loop turn (IL-6R $\alpha$ ). E69/S76/G77. Tight cluster, highly exposed, no glycan risk.
- 2. Site I extension** — BC loop-Helix D bridge (IL-6R $\alpha$ ). N79/E80/Q183. Conformational epitope: 104 residues apart in sequence, ~10 Å in space.
- 3. Site II** — Helix C (IL-6R $\beta$ ). E109/E110/R113. E110 most buried by IL-6R $\beta$  ( $\Delta$ -ASA 0.48) but fully exposed on free IL-6.
- 4. Site II** — Helix A (IL-6R $\beta$ ). E23/K27. Charge-complementary pair on orthogonal face of Site II.
- 5. Site III** — CD loop (hexameric assembly). K131/N132/D134. Highly exposed; moderate glycan risk from B144 sequon.

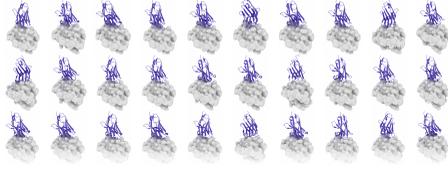
\*\*\*

IL-6 is 163 residues, VHH frameworks are 117-146 — total context ~300 residues, well within model limits. No cropping needed. Launching pilots.

**Tool use**

Submit 20 pilot batches - 5 epitopes x 4 frameworks, 500 designs each.

\*\*\*



**Tool use**

All 20 batches complete. 31 passing from 10,000 designs.

\*\*\*

**Key findings:**

- Site II is the most productive epitope region — both Helix A and Helix C hotspots produce hits.
- 7eov-a is the dominant framework (24/31 passes). Sjds-b is secondary (6/31). Three other frameworks: zero.
- Sites I, III: zero passes — flat helical surface with challenging topology. I need ~69 more, user did not specify an epitope preference, so dropping dead configurations, extending top performers, and exploring hotspot and CDR3 length variations within the productive epitopes.

\*\*\*

**Tool use**

3 batch extensions + 14 new batches with CDR3 +/-2 and alternative hotspot residue subsets.

\*\*\*

Shifting hotspot E23>R24 on Helix A dramatically improved the pass rate — from 2.8% to 15.6% (78 passes from 500 designs).

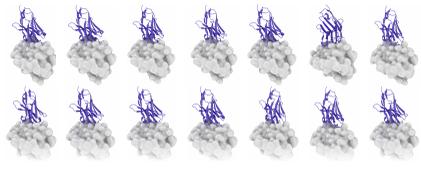
\*\*\*

**Tool use**

QA Assessment on passing binders.

\*\*\*

From 36 batches (~18,500 designs) we find 118 unique candidates across 26 sequence clusters. This is enough to satisfy the user's request: No unpaired cysteines, no N-glycosylation in CDRs, no further sequence liabilities. All novel (>10 edit distance from SABTab). Top 88 selected, spanning both Site II sub-regions, two framework scaffolds, and CDR3 length variation.



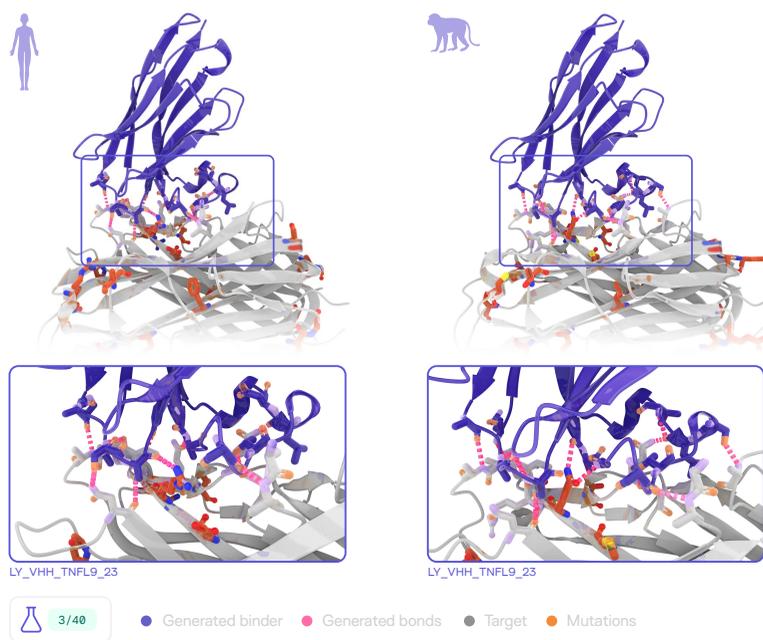
ID	Sequence	Filter	ipTM	DockQ	Complex RMSD
67	EVQLV ... AGIPGDFT ...	PASS	0.86	0.86	1.22
11	EVQLV ... AATVLDPS ...	PASS	0.85	0.87	1.64
87	EVQLV ... AGSYLDPS ...	PASS	0.82	0.80	1.36

ip9m stimulative\_haectet Open ↗

**Fig. 3 | Latent-Y autonomously designs binders to IL-6 to disrupt the IL-6/IL-6R complex formation.** A condensed trace of the full Latent-Y campaign that produced the lab-confirmed IL-6 binders reported in this work, spanning close to 10 000 lines of reasoning in its entirety. The trace shows the agent executing a high-level user prompt (grey box), with tool uses and subagent calls highlighted (purple boxes). Key stages include database search and target identification, hotspot analysis via a dedicated subagent, iterative pilot and scale-up batches, and a final quality assurance to select lab-ready candidates. Ellipses denote reasoning steps omitted for visual brevity.

Starting from a prompt specifying competitive disruption of TNFL9–TNFRSF9 binding, the human co-crystal structure (PDB 6A3V), and the cyno sequence alone, Latent-Y predicts the cyno trimer from sequence and aligns it to the human complex. The agent crops both structures to the relevant context, resolves disordered regions, and maps receptor-contact residues across species. Using dedicated subagents, it characterizes the binding interface and identifies candidate epitope configurations across the trimeric surface that remain minimally affected by the mutations between species. To enable cross-species generation, the agent implements a custom generative method using privileged platform access, translating the high-level request for a joint generation approach into working code without further instruction. The expert reviews the initial outputs to verify the underlying logic and provides targeted biological steering at key points. For example, the agent initially identifies hotspots that are geometrically accessible from the intra-trimeric side *in silico*, due to the crop, but biologically implausible. This is a form of reward hacking that human oversight identifies. Upon prompting, the agent imposes a geometric filter and redirects exploration toward single-chain configurations consistent with the desired binding stoichiometry. Latent-Y integrates each constraint, iterates over hotspot combinations, learns from evidence as more samples come in, and converges on tight three-region epitope configurations that produce dual-passing binders.

Within a set of 40 binders designed and selected for wet lab synthesis by the agent, HT-SPR screening identified three out of 40 binders as hits, exhibiting cross-reactive binding to both human and cyno TNFL9, as shown in Fig. 4 and described in Sec. A.2. These results demonstrate that Latent-Y can autonomously translate a rough conceptual sketch into functional dual-target molecules, providing immediate starting points for affinity maturation.



**Fig. 4 | Latent-Y designs cross-species VHH binders through autonomous capability extension.** Designed structures of the *de novo* VHH complexed with human (left) and cynomolgus (right) TNFL9, with divergent mutations between species highlighted in orange. The VHH is shown bound to the trimeric unit. HT-SPR identified hits cross-reactive binding to both human and cynomolgus targets. To execute this campaign, Latent-Y developed a custom generative method from a one-line natural language description, with expert involvement limited to biological steering and logic verification.

I want to design VHH binders against TNFL9 that bind both the human and cyno proteins and competitively disrupt binding to TNFR9. We have the human co-crystal structure '6a3v', but for cyno I only have the sequence: REGPELSPDN\_SPRSE. Prepare the structures and hotspots, but then wait for me to prototype a cross-reactivity plugin to the code together.

Ok. I'll first predict a cyno structure, align it to the human complex and identify receptor-blocking hotspots that are conserved between species.

Predict the cyno TNFL9 trimer from sequence.

Align the cyno model to the human '6a3v' complex and standardize both into the same residue/chain convention.

Remove unresolved terminal tails and crop to the relevant design context.

Map residues in contact with TNFR9 and remove any hotspots that sit on human/cyno mutation sites.

**Tool use**

Structure prediction, alignment, residue mapping. Multiple code and model calls

The cyno model is now aligned to the human structure and renumbered to match it, which makes cross-structure reasoning much simpler. Across the aligned construct there are 11 residue differences over 184 positions, so cross-reactivity looks feasible if we avoid mutation sites. Let me analyze the interface region.

**Subagent call:** Hotspot researcher

TNFL9 is trimeric, but the relevant receptor-facing context is much smaller. I've removed the flexible tails that are unresolved in the crystal structure as well, and cropped both species to chains A and E so we keep the receptor-binding surface while avoiding unnecessary context.

The TNFR9 interface spans several spatial clusters. After excluding mutated residues, I see two broad options:

- A-chain hotspots that directly overlap the receptor-binding face.
- Cross-chain A+E hotspot sets that span the groove between protomers.

The A+E sets are geometrically attractive, but because TNFL9 is a trimer I need to be careful not to optimise toward the trimer interface instead of the receptor-exposed face.

Ready to prototype a cross-reactivity plugin to the code together.

Can you write a custom generation routine that combines the model outputs at generation time in a smart way to achieve cross-reactivity? Use the Latent Labs Platform primitives and tools where helpful. Prototype the idea in a notebook first so I can take a quick look at an example.

Yes. I'll write a small analysis notebook using Latent Labs primitives to jointly design each candidate against the human and cyno structures simultaneously.

**Advanced tool use**

Notebook generation, code synthesis, custom generation setup.

Ok this approach looks good, please integrate this into the generative loop and start a few pilot batches.

**Advanced tool use**

Code use, use of Latent Labs Platform primitives.

I started with a small hotspot scan across A-only, E-heavy, and balanced A+E sets. The first informative wave pointed to one clear winner: a balanced A+E set around A100,151,194 and E189,210. E-heavy sets underperformed, while the balanced set produced the first dual-pass sample. Long CDR3s (23) look promising for the balanced sets.

One failure mode I'm worried about with the crop and hotspots that you use is that the binder can go on the wrong side of the trimer, toward the trimer interface rather than the receptor-facing side. Can we align sampled targets to the reference complex and measure where the COM of the binder sits relative to the natural ligand?

**Advanced tool use**

Alignment snippet generation, structural superposition, binder-location analysis.

Yes. I'll write a notebook snippet that aligns the sampled target to the reference complex, then measures binder placement relative to the natural ligand position.

**Advanced tool use**

Alignment snippet generation, structural superposition, binder-location analysis.

Yes, the second one is good. Let's use a COM distance threshold of 20Å with a simultaneous target alignment RMSD of 2Å or better as a 'binder is in the right place' filter. Can you add that to the validation checks?

Understood. I'll add that filter and cache the reference structure so we can apply it efficiently during validation.

**Tool use**

Validation update, cached reference loading, result aggregation.

With that location filter in place, the next exploratory wave still shows a strong cross-chain signal. v3\_AE\_e210' with A115,161-E1210) on '7eow' is the best config by far: average min-IPFM 0.407; max 0.757 after 94 samples.

**Key findings:**

- Cross-chain A+E is still the strongest signal.
- The average signal is about 2x higher than the A-only sets.
- I should keep the cross-chain branch alive, but not over-commit to it yet.

I want to try a few more A-only ones, because TNFL9 is mostly bound in a 1:1 stoichiometry rather than cross interfaces, so let's not do too many A-E combos only. Let's relax the span a bit to say '22A' or so - that should still be fine.

Got it. Let me re-examine A-only combos with the relaxed 22A threshold.

**Tool use**

Distance matrix calculation, hotspot recombination, config generation.

I found several viable cross-region A-only combos at 22A. The key bridging residues are A100, A101, A102, and A149 at the N-term/mid boundary.

Best cross-region A-only sets:

- A101,152,227 spans all 3 epitope regions with only 10.3Å max distance.
- A100,01,149 is a tight bridge cluster.
- A101,149,152 bridges cleanly into the mid region.
- A149 is a standout bridge residue and is conserved between species.

Launch all of those - but let's not add more A+E.

Understood. I'll launch the cross-region A-only configs and keep the already-running A+E jobs only as a comparison set.

**Tool use**

Wave redesign, config generation, further launches.

A-only is now starting to catch up. v3\_w2\_bridg' with A100,151,194) on '5jids' reached 0.777, which is the best A-only result so far, and '5jids' clearly outperforms '7eow' on the mid-region hotspot set.

A clearer pattern is emerging:

- Tight three-region A-only sets work better than wider spreads.
- '5jids' is consistently the strongest framework on this epitope.
- 'Seak' is clearly underperforming here.

v3\_w3\_tigh' with A100,149,232) on '5jids' just reached 0.788 and became the first A-only config to beat the earlier cross-chain A+E designs.

**Advanced tool use**

Continued autonomous iteration as in other campaigns to reach design target.

That result is now reinforced by an independent neighbour: A196,151,234) on '5jids' reached 0.787. I'm going to scale the best A-only motifs and test neighbouring hotspot variants, especially A149 versus A151.

**Advanced tool use**

Neighbour exploration paid off. v3\_w3\_tigh\_ver2' with A100,151,232) on '5jids' produced dual-pass binders. Let me scale up this configuration.

**Fig. 5 | Latent-Y collaboratively designs cross-species TNFL9 binders through autonomous capability extension.** A condensed trace of the full Latent-Y campaign that produced the lab-validated cross-species binders reported in this work, spanning over 30 000 lines in its entirety. User prompts and interventions are shown in grey boxes, with tool uses and subagent calls highlighted in purple. Key stages include cyno structure prediction and alignment, hotspot analysis via a dedicated subagent, development of a custom generative method, and iterative wave-based exploration with intermediate findings summaries. Ellipses denote reasoning steps omitted for visual brevity.

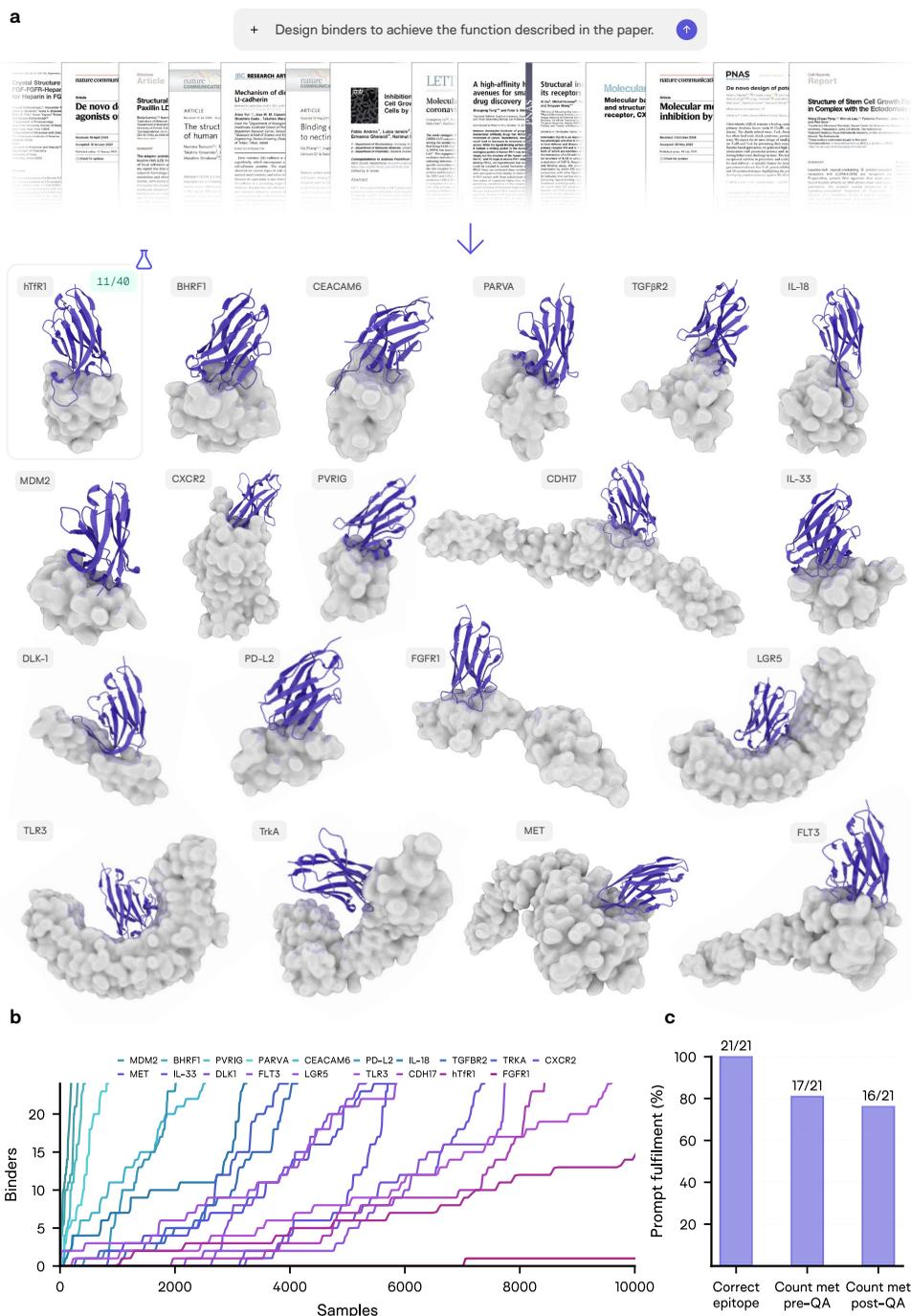
## 2.4. Latent-Y translates scientific publications into antibody binders targeting the reported epitope

To evaluate Latent-Y’s ability to reason from scientific context alone, we benchmark the agent across 21 peer-reviewed publications, each describing a therapeutically relevant protein–protein interaction, as detailed in Tab. S2. In this setting, all design information must be derived solely from the publication: the agent infers the biological context, mechanism of action, and the identity and structural location of the functionally relevant site to target directly from the text. The prompt only states the target protein name to avoid ambiguity for cases where multiple proteins discussed in a paper can be viable therapeutic targets. Each task uses a peer-reviewed publication as its sole input, mimicking the kind of scientific starting point that researchers encounter in practice, whether developing a work plan, drafting a grant proposal, or building on prior literature. Using published papers ensures that the scenarios are not hypothetical but reflect targets and interactions of genuine interest to the scientific community, drawn from leading journals including *Nature*, *Science*, *Cell*, and *Proceedings of the National Academy of Sciences (PNAS)*.

As highlighted in Tab. S2, the selected publications span diverse disease areas including oncology, immunology, inflammation, neuroscience, and metabolic disorders, and encompass a range of binding interactions to disrupt, including natural protein partners, natural peptides, non-antibody biologics, and computationally designed proteins. To prevent information leakage, we selected targets for which no Fab-, scFv-, or VHH-bound structure of the epitope specified in the source publication was available in the PDB as of 16 March 2026. This ensures that no antibody-specific structural information is accessible to the agent. Each campaign is evaluated under a uniform prompt, asking the agent to generate 24 computationally passing VHH binders within a sampling budget of 10 000 designs, with the full prompt provided in Sec. B.

Across 21 campaigns, Latent-Y successfully identified the correct target epitope in 21 of 21 cases (100%), generated 24 or more passing binders before quality assurance (QA), as detailed in Sec. B, in 17 of 21 cases (81%), and delivered 24 or more passing binders after QA in 16 of 21 cases (76%), as shown in Fig. 6c. Further details are provided in Sec. B. The number of computationally passing binders accumulates non-linearly with total samples generated, frequently accelerating as campaigns progress and the agent refines its generative strategy, as shown in Fig. 6b. This pattern reflects the agent’s iterative explore-then-exploit reasoning: early batches probe diverse epitope, framework, and CDR length configurations, and once productive combinations are identified, the agent doubles down, rapidly accumulating passing designs. The agent completed the majority of campaigns well within the 10 000-sample budget, with a median budget consumption of approximately 4,600 samples across completed runs to reach 24 passing binders. Two out of 21 campaigns did not yield any passing binders within the specified budget. For the 19 campaigns that produced binders passing our computational filters, we show the design with the highest ipTM in Fig. 6a and the cumulative number of binders over the sample budget in Fig. 6b.

The hTfR1 campaign, targeting human transferrin receptor for blood–brain barrier crossing applications [27], was selected for experimental validation. Starting from a *Science Translational Medicine* publication describing Fc-mediated brain delivery [27], the agent inferred the relevant epitope from the published mechanism. To maximize the pool of candidates for experimental testing, this campaign was extended beyond the standard 10,000-sample budget, allowing the selection of 40 high-quality binders after QA. HT-SPR screening identified 11 of 40 tested designs as binder hits, as described in Sec. A.2, yielding a hit rate of 28%, shown in Fig. 6a and Fig. 1b.



**Fig. 6 | Benchmarking Latent-Y on autonomous VHH design from peer-reviewed scientific publications.**

(a) Designed complexes of *de novo* VHH binders against 19 therapeutic targets for which Latent-Y found computational passes within a strict 10 000-sample budget, evaluated under a uniform prompt. Each VHH was designed based on context derived entirely from the text of a provided, peer-reviewed publication. The hTfR1 campaign (top left) was experimentally validated, yielding a binding hit rate of 11/40 as determined by HT-SPR. The remaining campaigns were not sent for experimental testing. (b) Number of computationally passing binders as a function of total samples generated across all 21 campaigns. (c) Agent reasoning and prompt fulfilment metrics across campaigns, showing success rates for correct epitope identification, generation of 24 or more passing binders before QA, and after QA, as described in Sec. B.

## 3. Methods

### 3.1. Latent-Y

Latent-Y is an agentic AI system [28] built on frontier large language models as its reasoning engine, integrated natively with the Latent Labs Platform via a model context protocol (MCP, [29]) server that manages access to Latent-X2 inference, computational scoring, platform tooling, and the Latent Labs compute infrastructure. The agent harness is provider-agnostic; we evaluated several leading frontier LLMs and found performance to be robust across this variation, consistent with recent findings on frontier reasoning models [30–32]. The same harness can operate in tandem with general-purpose coding agent capabilities, further extending the agent’s action space, as demonstrated in Sec. 2.3.

Effective context management is a central design principle of the harness, following recent findings on long-running agent systems [33, 34]. Specialized subagents handle distinct analytical tasks, for instance, a dedicated hotspot researcher subagent for target characterization and epitope analysis, and a quality assurance subagent for candidate evaluation, diversity filtering, and sequence liability analysis, with structured context hand-off between the orchestrator and subagents. The agent draws on standard bioinformatics utilities [35–38], external APIs for biomolecular databases [39–44] and scientific literature [45], and purpose-built tools for the specific demands of protein binder design campaigns. Users typically specify the desired modality — nanobodies, macrocyclic peptides, or miniprotein binders — directly in their prompt, though the agent can also evaluate and route across modalities autonomously based on epitope geometry, pharmacological requirements, and empirical model performance. System prompts encode accumulated drug design reasoning, guiding the agent towards productive strategies while preserving flexibility to adapt dynamically, reflecting findings that soft guidance outperforms rigid workflow specification for complex open-ended tasks [33, 46, 47].

Latent-Y’s reasoning follows an explore-then-exploit pattern, spawning targeted computational experiments, reasoning over intermediate results, and concentrating resources on productive directions as campaigns progress, connecting to recent work on autonomous research agents [48–51]. The agent can extend its own capabilities by generating custom computational methods from natural language descriptions when standard tools are insufficient, as demonstrated in Sec. 2.3. Each campaign produces a complete reasoning trace capturing all decisions, tool calls, and strategy updates. Researchers can interrupt the agent at any point to inject biological context, override decisions, or redirect strategy, with the agent integrating these inputs and adapting its subsequent reasoning and operation accordingly.

### 3.2. Wet-lab methods

Binder screening and characterization followed a tiered workflow. A primary screen was performed using a one-point HT-SPR assay to identify candidates with measurable target binding, as in Sec. A.2. Designs exceeding the binding response threshold specified in Sec. A.2 were designated as hits. Hits from our primary screen that were advanced to determine  $K_D$  were evaluated by five-point SPR, with affinities reported for binders meeting predefined criteria described in Sec. A.3.

## 4. Discussion

This work presents a significant milestone in computational drug discovery: the first autonomous agent for *de novo* biologics design with lab-validated results. Using text prompts expressing goals and constraints, Latent-Y delivers novel antibody sequences confirmed in the laboratory, demonstrating that the full workflow of an expert drug design campaign can be executed autonomously and without manual intervention. We regard this as a meaningful step towards the broader goal of an AI scientist for biology.

Across nine targets that span three qualitatively different campaign types, Latent-Y successfully produced lab-confirmed binders against six, achieving a 67% target-level success rate with binding affinities reaching the single-digit nanomolar range. A particularly notable demonstration is Latent-Y’s ability to reason from scientific literature: given existing publications as input, the agent autonomously identified targets and epitopes, reasoned about published mechanisms of action, and designed binders accordingly. One such campaign was confirmed in the laboratory. The literature-inferred design benchmark further illustrates this at scale: 21 campaigns run in parallel, each seeded from a peer-reviewed publication describing a therapeutically relevant interaction, represent a volume of simultaneous scientific exploration that would be infeasible for any individual expert working alone. In user studies, experts working with Latent-Y completed design campaigns 56-fold faster than independent expert time estimates, compressing weeks of computational work into hours, with further gains achievable by running campaigns in parallel across multiple programs simultaneously.

Crucially, Latent-Y’s behaviour is not predetermined. Rather than executing a fixed workflow, the agent navigates each campaign adaptively, a property that connects to efforts toward autonomous AI scientists across scientific domains [48, 49] and the emerging paradigm of autoresearch [50]. Latent-Y contributes to this landscape with lab-validated results in therapeutic antibody design. Beyond adaptive reasoning within campaigns, Latent-Y can extend its own generative capabilities in response to novel design challenges. This is demonstrated by the cross-species campaign in which the agent implemented a custom generative method from a brief natural language description, yielding nanobodies that simultaneously bound human and cynomolgus homologues, confirmed in the laboratory. Reasoning traces are fully observable at every step, capturing each decision, tool call, and strategy update, providing the transparency and auditability that responsible deployment of autonomous agents requires. Current limitations reflect the performance of the underlying frontier LLMs, which we have evaluated across several leading models, the generative capabilities of Latent-X2 [1], and the tools available to the agent. Laboratory and clinical validation of designed molecules remains essential. Latent-Y accelerates the computational stages of drug discovery but does not replace the experimental, confirmatory stages that must follow.

A question of central importance for AI scientists is whether they will produce truly novel scientific discoveries. Within the context of drug design, Latent-Y already delivers on this promise: every confirmed binder it produces is a novel molecule, designed *de novo* from a text prompt, that did not previously exist. The broader question of whether agents will uncover unexpected mechanisms or entirely new target biology remains an open and compelling frontier. Closing the loop with experimental feedback, expanding the agent’s action space, and ultimately integrating with fully robotic laboratories are promising directions. Extending laboratory validation to macrocyclic peptides and mini-binders, modalities the agent already supports, is a further natural direction.

Latent-Y and future versions stand to turn drug design into an increasingly computational discipline, making world-class molecular design expertise available to any researcher with a well-posed scientific question, and enabling drug-discovery organisations to operate at a scale and speed not previously possible. Latent-Y is available to selected partners at [platform.latentlabs.com](https://platform.latentlabs.com).

## Contributors

Sebastian M. Schmon, Daniella Pretorius, Simon Mathis, Rebecca Bartke-Croughan, Aishaini Puvanendran, James Vuckovic, Henry Kenlay, Mária Vlachynská, Alex Bridgland, Ivan Grishin, Sven Over, David Li, Bridget Li, Jonathan Crabbé, Agrin Hilmkil\*\*, Alexander W. R. Nelson\*, David Yuan\*\*, Annette Obika, Simon A. A. Kohl\*\*\*

\*\*\* Corresponding author. E-mail: [simon@latentlabs.com](mailto:simon@latentlabs.com).

\*\* Work performed while at Latent Labs.

\* Work performed as an advisor to Latent Labs.

### Author contributions

**Conceptualization and team leadership:** S.K. conceived the research direction and priorities for applications, S.S., S.K. led the team and research. D.P. led the experimental design with contributions from R.B.C. and H.K. A.P. conceptualized and led user research studies, with contributions from D.P. and A.O. A.P., A.O. and A.N. contributed to project delivery and narrative.

**Machine learning development:** S.S., S.K. developed the agent, with contributions from H.K., S.M., A.B., J.C. and A.H.; J.V., A.B., S.O., A.H. and I.G. deployed and maintained computational infrastructure for model inference.

**Platform infrastructure:** S.S. developed the agentic platform integration with contributions from A.H. S.O., I.G., D.L. and S.S. developed and maintained the agent platform infrastructure, including front end and back end.

**Computational design and evaluation:** D.P., S.S., S.M. led computational protein design workflows with contributions from H.K., B.L. and A.P. D.P., R.B.C. and S.M. analysed experimental data. S.S., S.M., D.P. performed computational benchmarking.

**Experimental validation:** D.Y., R.B.C. and D.P. oversaw internal and external validation. R.B.C. conducted internal experiments and contributed to experimental design. D.Y., D.P. and R.B.C. managed external laboratory partnerships.

**Writing and figures:** S.K. oversaw manuscript delivery. S.S., S.K., D.P., R.B.C., A.P. and S.M. wrote the manuscript. M.V., A.B., S.M., S.S., D.L. and A.P. made figures.

All authors contributed to the work and approved the final manuscript.

### Competing interests

All authors have contributed as employees, contractors or advisors of Latent Labs Technologies Inc. or Latent Labs Limited.

### References

- [1] Latent Labs Team, Henry Kenlay, et al. “Drug-like antibodies with low immunogenicity in human panels designed with Latent-X2”. In: *arXiv* (Dec. 2025). arXiv:2512.20263 [q-bio]. doi: [10.48550/arXiv.2512.20263](https://doi.org/10.48550/arXiv.2512.20263). URL: <http://arxiv.org/abs/2512.20263>.
- [2] Chai Discovery Team, Jacques Boitreaud, et al. “Drug-like antibody design against challenging targets with atomic precision”. In: *bioRxiv* (2025).
- [3] Nabla Bio and Surojit Biswas. “De novo design of epitope-specific antibodies against soluble and multipass membrane proteins with high specificity, developability, and function”. In: *bioRxiv* (2025), pp. 2025–01.
- [4] John Jumper, Richard Evans, et al. “Highly accurate protein structure prediction with AlphaFold”. In: *nature* 596.7873 (2021), pp. 583–589.
- [5] Josh Abramson, Jonas Adler, et al. “Accurate structure prediction of biomolecular interactions with AlphaFold 3”. In: *Nature* 630.8016 (2024), pp. 493–500.
- [6] Justas Dauparas, Ivan Anishchenko, et al. “Robust deep learning–based protein sequence design using ProteinMPNN”. In: *Science* 378.6615 (2022), pp. 49–56.
- [7] Joseph L Watson, David Juergens, et al. “De novo design of protein structure and function with RFdiffusion”. In: *Nature* 620.7976 (2023), pp. 1089–1100.
- [8] Vinicius Zambaldi, David La, et al. “De novo design of high-affinity protein binders with AlphaProteo”. In: *arXiv* (2024). arXiv:2409.08022.
- [9] Martin Pacesa, Lennart Nickel, et al. “BindCraft: one-shot design of functional protein binders”. In: *bioRxiv* (2024). bioRxiv preprint.

- [10] Latent Labs Team, Alex Bridgland, et al. “Latent-X: An Atom-level Frontier Model for De Novo Protein Binder Design”. In: *arXiv* (2025). arXiv:2507.19375.
- [11] Thomas Hayes, Roshan Rao, et al. “Simulating 500 million years of evolution with a language model”. In: *Science* 387.6736 (2025), pp. 850–858.
- [12] Tomas Geffner, Kieran Didi, et al. “Proteina: Scaling flow-based protein structure generative models”. In: *arXiv preprint arXiv:2503.00710* (2025).
- [13] Hannes Stark, Felix Faltings, et al. “BoltzGen: Toward Universal Binder Design”. In: *bioRxiv* (2025). bioRxiv preprint.
- [14] Protenix Team, Milong Ren, et al. “PXDesign: Fast, modular, and accurate de novo design of protein binders”. In: *bioRxiv* (2025), pp. 2025–08.
- [15] Toshio Tanaka, Masashi Narazaki, and Tadimitsu Kishimoto. “IL-6 in inflammation, immunity, and disease”. In: *Cold Spring Harbor Perspectives in Biology* 6.10 (2014), a016295. doi: [10.1101/cshperspect.a016295](https://doi.org/10.1101/cshperspect.a016295).
- [16] S Barillé, R Bataille, and M Amiot. “The role of interleukin-6 and interleukin-6/interleukin-6 receptor-alpha complex in the pathogenesis of multiple myeloma”. In: *European Cytokine Network* 11.4 (2000), pp. 546–551.
- [17] Violaine Bernard, Jacques Young, and Nadine Binart. “Prolactin — a pleiotropic factor in health and disease”. In: *Nature Reviews Endocrinology* 15.6 (2019), pp. 356–365. doi: [10.1038/s41574-019-0194-6](https://doi.org/10.1038/s41574-019-0194-6).
- [18] Foo Yew Liew, Jean-Philippe Girard, and Heth Roderick Turnquist. “Interleukin-33 in health and disease”. In: *Nature Reviews Immunology* 16.11 (2016), pp. 676–689. doi: [10.1038/nri.2016.95](https://doi.org/10.1038/nri.2016.95).
- [19] Dominic Brenner, Henning Blaser, and Tak W Mak. “Regulation of tumour necrosis factor signalling: live or let die”. In: *Nature Reviews Immunology* 15.6 (2015), pp. 362–374. doi: [10.1038/nri3834](https://doi.org/10.1038/nri3834).
- [20] Jun Lan, Jiwan Ge, et al. “Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor”. In: *Nature* 581.7807 (2020), pp. 215–220. doi: [10.1038/s41586-020-2180-5](https://doi.org/10.1038/s41586-020-2180-5).
- [21] Chao Wang, Gloria H Y Lin, et al. “Immune regulation by 4-1BB and 4-1BBL: complexities and challenges”. In: *Immunological Reviews* 229.1 (2009), pp. 192–215.
- [22] Elena Gammella, Paolo Buratti, et al. “The transferrin receptor: the cellular iron gate”. In: *Metallomics* 9.10 (2017), pp. 1367–1375. doi: [10.1039/c7mt00143f](https://doi.org/10.1039/c7mt00143f).
- [23] Frits van Rhee, Luis Fayad, et al. “Siltuximab, a Novel Anti-Interleukin-6 Monoclonal Antibody, for Castleman’s Disease”. en. In: *Journal of Clinical Oncology* 28.23 (Aug. 2010), pp. 3701–3708. issn: 0732-183X, 1527-7755. doi: [10.1200/JCO.2009.27.2377](https://doi.org/10.1200/JCO.2009.27.2377). url: <https://ascopubs.org/doi/10.1200/JCO.2009.27.2377> (visited on 03/22/2026).
- [24] Toshio Tanaka, Masashi Narazaki, and Tadimitsu Kishimoto. “Tocilizumab, a humanized anti-interleukin-6 receptor antibody, for treatment of rheumatoid arthritis”. In: *Biomark Med* 8.5 (2014), pp. 745–758. doi: [10.2217/bmm.14.37](https://doi.org/10.2217/bmm.14.37).
- [25] Stephanie Maciuba, Gregory D. Bowden, et al. “Discovery and characterization of prolactin neutralizing monoclonal antibodies for the treatment of female-prevalent pain disorders”. en. In: *mAbs* 15.1 (Sept. 2023), p. 2254676. issn: 1942-0862, 1942-0870. doi: [10.1080/19420862.2023.2254676](https://doi.org/10.1080/19420862.2023.2254676). url: <https://www.tandfonline.com/doi/full/10.1080/19420862.2023.2254676> (visited on 03/22/2026).
- [26] Just P Brakenhoff, FD De Hon, et al. “Development of a human interleukin-6 receptor antagonist.” In: *Journal of Biological Chemistry* 269.1 (1994), pp. 86–93.
- [27] Mihalis S Kariolis, Robert C Wells, et al. “Brain delivery of therapeutic proteins using an Fc fragment blood-brain barrier transport vehicle in mice and monkeys”. en. In: *Sci. Transl. Med.* (2020).
- [28] Shunyu Yao, Jeffrey Zhao, et al. “React: Synergizing reasoning and acting in language models”. In: *The eleventh international conference on learning representations*. 2022.
- [29] Anthropic. *Model Context Protocol*. <https://modelcontextprotocol.io>. 2024.
- [30] Anthropic. *System Card: Claude Opus 4 & Claude Sonnet 4*. Tech. rep. Anthropic, May 2025. url: <https://www-cdn.anthropic.com/4263b940cabb546aa0e3283f35b686f4f3b2ff47.pdf>.
- [31] OpenAI, Josh Achiam, et al. *GPT-4 Technical Report*. 2024. arXiv: 2303.08774 [cs.CL]. url: <https://arxiv.org/abs/2303.08774>.
- [32] Gemini Team, Rohan Anil, et al. *Gemini: A Family of Highly Capable Multimodal Models*. 2025. arXiv: 2312.11805 [cs.CL]. url: <https://arxiv.org/abs/2312.11805>.
- [33] Anthropic. *Effective harnesses for long-running agents*. Nov. 26, 2025. url: <https://www.anthropic.com/engineering/effective-harnesses-for-long-running-agents>.
- [34] Anthropic. *Effective context engineering for AI agents*. Sept. 29, 2025. url: <https://www.anthropic.com/engineering/effective-context-engineering-for-ai-agents>.

- [35] James Dunbar and Charlotte M Deane. “ANARCI: antigen receptor numbering and receptor classification”. In: *Bioinformatics* 32.2 (2016), pp. 298–300.
- [36] Martin Steinegger and Johannes Söding. “MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets”. In: *Nature Biotechnology* 35.11 (2017), pp. 1026–1028.
- [37] Stephen F Altschul, Warren Gish, et al. “Basic local alignment search tool”. In: *Journal of Molecular Biology* 215.3 (1990), pp. 403–410.
- [38] Patrick Kunzmann and Kay Hamacher. “Biotite: a unifying open source computational biology framework in Python”. In: *BMC bioinformatics* 19.1 (2018), p. 346.
- [39] UniProt Consortium. “UniProt: a worldwide hub of protein knowledge”. In: *Nucleic acids research* 47.D1 (2019), pp. D506–D515.
- [40] Helen M Berman, John Westbrook, et al. “The protein data bank”. In: *Nucleic Acids Research* 28.1 (2000), pp. 235–242.
- [41] National Center for Biotechnology Information (NCBI). *PubMed [Internet]*. Bethesda (MD), 1996. URL: <https://www.ncbi.nlm.nih.gov/Web/Search/entrezfs.html>.
- [42] James Dunbar, Konrad Krawczyk, et al. “SAbDab: the structural antibody database”. In: *Nucleic Acids Research* 42.D1 (2014), pp. D1140–D1146.
- [43] Anna Gaulton, Louisa J Bellis, et al. “ChEMBL: a large-scale bioactivity database for drug discovery”. In: *Nucleic acids research* 40.D1 (2012), pp. D1100–D1107.
- [44] Mihaly Varadi, Stephen Anyango, et al. “AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models”. In: *Nucleic acids research* 50.D1 (2022), pp. D439–D444.
- [45] National Center for Biotechnology Information (NCBI). *PubMed [Internet]*. Bethesda (MD), 1996. URL: <https://pubmed.ncbi.nlm.nih.gov/>.
- [46] OpenAI Codex Team. *Harness engineering: leveraging Codex in an agent-first world*. Feb. 11, 2026. URL: <https://openai.com/index/harness-engineering/>.
- [47] Amp Team. *Context Management in Amp*. URL: <https://ampcode.com/guides/context-management>.
- [48] Juraj Gottweis, Wei-Hung Weng, et al. “Towards an AI co-scientist”. In: *arXiv preprint arXiv:2502.18864* (2025).
- [49] Ludovico Mitchener, Angela Yiu, et al. “Kosmos: An ai scientist for autonomous discovery”. In: *arXiv preprint arXiv:2511.02824* (2025).
- [50] Andrej Karpathy. *autoresearch: AI agents running research automatically*. <https://github.com/karpathy/autoresearch>. 2026.
- [51] Peter Steinberger and OpenClaw Contributors. *OpenClaw: Your own personal AI assistant*. <https://github.com/openclaw/openclaw>. 2025.
- [52] Chloe S. Adams, Hyojin Kim, et al. “De novo design of protein minibinder agonists of TLR3”. In: *Nature Communications* 16.1 (Jan. 2025), p. 1234. ISSN: 2041-1723. DOI: [10.1038/s41467-025-56369-w](https://doi.org/10.1038/s41467-025-56369-w). URL: <https://www.nature.com/articles/s41467-025-56369-w> (visited on 03/16/2026).
- [53] Erik Procko, Geoffrey Y. Berguig, et al. “A Computationally Designed Inhibitor of an Epstein-Barr Viral Bcl-2 Protein Induces Apoptosis in Infected Cells”. en. In: *Cell* 157.7 (June 2014), pp. 1644–1656. ISSN: 00928674. DOI: [10.1016/j.cell.2014.04.034](https://doi.org/10.1016/j.cell.2014.04.034). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0092867414006138> (visited on 03/18/2026).
- [54] Carin C. Stamper, Yan Zhang, et al. “Crystal structure of the B7-1/CTLA-4 complex that inhibits human immune responses”. en. In: *Nature* 410.6828 (Mar. 2001), pp. 608–611. ISSN: 0028-0836, 1476-4687. DOI: [10.1038/35069118](https://doi.org/10.1038/35069118). URL: <https://www.nature.com/articles/35069118> (visited on 03/18/2026).
- [55] Anna Yui, Jose M.M. Caaveiro, et al. “Mechanism of dimerization and structural features of human L1-cadherin”. en. In: *Journal of Biological Chemistry* 297.3 (Sept. 2021), p. 101054. ISSN: 00219258. DOI: [10.1016/j.jbc.2021.101054](https://doi.org/10.1016/j.jbc.2021.101054). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0021925821008565> (visited on 03/18/2026).
- [56] Daniel A. Bonsor, Sebastian Günther, et al. “Diverse oligomeric states of CEACAM IgV domains”. en. In: *Proceedings of the National Academy of Sciences* 112.44 (Nov. 2015), pp. 13561–13566. ISSN: 0027-8424, 1091-6490. DOI: [10.1073/pnas.1509511112](https://doi.org/10.1073/pnas.1509511112). URL: <https://pnas.org/doi/full/10.1073/pnas.1509511112> (visited on 03/18/2026).
- [57] Shirsha Saha, Fumiya K. Sano, et al. “Molecular basis of promiscuous chemokine binding and structural mimicry at the C-X-C chemokine receptor, CXCR2”. en. In: *Molecular Cell* 85.5 (Mar. 2025), 976–988.e9.

- ISSN: 10972765. DOI: [10.1016/j.molcel.2025.01.024](https://doi.org/10.1016/j.molcel.2025.01.024). URL: <https://linkinghub.elsevier.com/retrieve/pii/S1097276525000589> (visited on 03/18/2026).
- [58] Daniel Antfolk, Qianqian Ming, et al. “Molecular mechanism of Activin receptor inhibition by DLK1”. en. In: *Nature Communications* 16.1 (July 2025), p. 5976. ISSN: 2041-1723. DOI: [10.1038/s41467-025-60634-3](https://doi.org/10.1038/s41467-025-60634-3). URL: <https://www.nature.com/articles/s41467-025-60634-3> (visited on 03/18/2026).
- [59] Joseph Schlessinger, Alexander N. Plotnikov, et al. “Crystal Structure of a Ternary FGF-FGFR-Heparin Complex Reveals a Dual Role for Heparin in FGFR Binding and Dimerization”. en. In: *Molecular Cell* 6.3 (Sept. 2000), pp. 743–750. ISSN: 10972765. DOI: [10.1016/S1097-2765\(00\)00073-3](https://doi.org/10.1016/S1097-2765(00)00073-3). URL: <https://linkinghub.elsevier.com/retrieve/pii/S1097276500000733> (visited on 03/17/2026).
- [60] Kenneth Verstraete, Gonzalez Vandriessche, et al. “Structural insights into the extracellular assembly of the hematopoietic Flt3 signaling complex”. en. In: *Blood* 118.1 (July 2011), pp. 60–68. ISSN: 0006-4971, 1528-0020. DOI: [10.1182/blood-2011-01-329532](https://doi.org/10.1182/blood-2011-01-329532). URL: <https://ashpublications.org/blood/article/118/1/60/28436/Structural-insights-into-the-extracellular> (visited on 03/18/2026).
- [61] Naotaka Tsutsumi, Takeshi Kimura, et al. “The structural basis for receptor recognition of human interleukin-18”. en. In: *Nature Communications* 5.1 (Dec. 2014), p. 5340. ISSN: 2041-1723. DOI: [10.1038/ncomms6340](https://doi.org/10.1038/ncomms6340). URL: <https://www.nature.com/articles/ncomms6340> (visited on 03/18/2026).
- [62] Guy P. A. Vigers, Lana J. Anderson, et al. “Crystal structure of the type-I interleukin-1 receptor complexed with interleukin-1 $\beta$ ”. en. In: *Nature* 386.6621 (Mar. 1997), pp. 190–194. ISSN: 0028-0836, 1476-4687. DOI: [10.1038/386190a0](https://doi.org/10.1038/386190a0). URL: <https://www.nature.com/articles/386190a0> (visited on 03/18/2026).
- [63] Xi Liu, Michal Hammel, et al. “Structural insights into the interaction of IL-33 with its receptors”. en. In: *Proceedings of the National Academy of Sciences* 110.37 (Sept. 2013), pp. 14918–14923. ISSN: 0027-8424, 1091-6490. DOI: [10.1073/pnas.1308651110](https://doi.org/10.1073/pnas.1308651110). URL: <https://pnas.org/doi/full/10.1073/pnas.1308651110> (visited on 03/18/2026).
- [64] Weng Chuan Peng, Wim de Lau, et al. “Structure of Stem Cell Growth Factor R-spondin 1 in Complex with the Ectodomain of Its Receptor LGR5”. en. In: *Cell Reports* 3.6 (June 2013), pp. 1885–1892. ISSN: 22111247. DOI: [10.1016/j.celrep.2013.06.009](https://doi.org/10.1016/j.celrep.2013.06.009). URL: <https://linkinghub.elsevier.com/retrieve/pii/S2211124713002878> (visited on 03/18/2026).
- [65] Paul H. Kussie, Svetlana Gorina, et al. “Structure of the MDM2 Oncoprotein Bound to the p53 Tumor Suppressor Transactivation Domain”. en. In: *Science* 274.5289 (Nov. 1996), pp. 948–953. ISSN: 0036-8075, 1095-9203. DOI: [10.1126/science.274.5289.948](https://doi.org/10.1126/science.274.5289.948). URL: <https://www.science.org/doi/10.1126/science.274.5289.948> (visited on 03/16/2026).
- [66] Fabio Andres, Luisa Iamele, et al. “Inhibition of the MET Kinase Activity and Cell Growth in MET-Addicted Cancer Cells by Bi-Paratopic Linking”. en. In: *Journal of Molecular Biology* 431.10 (May 2019), pp. 2020–2039. ISSN: 00222836. DOI: [10.1016/j.jmb.2019.03.024](https://doi.org/10.1016/j.jmb.2019.03.024). URL: <https://linkinghub.elsevier.com/retrieve/pii/S002228361930155X> (visited on 03/16/2026).
- [67] Sonja Lorenz, Ioannis Vakonakis, et al. “Structural Analysis of the Interactions Between Paxillin LD Motifs and  $\alpha$ -Parvin”. en. In: *Structure* 16.10 (Oct. 2008), pp. 1521–1531. ISSN: 09692126. DOI: [10.1016/j.str.2008.08.007](https://doi.org/10.1016/j.str.2008.08.007). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0969212608003341> (visited on 03/18/2026).
- [68] Shaogeng Tang and Peter S. Kim. “A high-affinity human PD-1/PD-L2 complex informs avenues for small-molecule immune checkpoint drug discovery”. en. In: *Proceedings of the National Academy of Sciences* 116.49 (Dec. 2019), pp. 24500–24506. ISSN: 0027-8424, 1091-6490. DOI: [10.1073/pnas.1916916116](https://doi.org/10.1073/pnas.1916916116). URL: <https://pnas.org/doi/full/10.1073/pnas.1916916116> (visited on 03/17/2026).
- [69] Songtao Hu, Pu Han, et al. “Structural basis for the immune recognition and selectivity of the immune receptor PVRIG for ligand Nectin-2”. en. In: *Structure* 32.7 (July 2024), 918–929.e4. ISSN: 09692126. DOI: [10.1016/j.str.2024.03.012](https://doi.org/10.1016/j.str.2024.03.012). URL: <https://linkinghub.elsevier.com/retrieve/pii/S0969212624000947> (visited on 03/18/2026).
- [70] Wei Yang, Derrick R. Hicks, et al. “Design of high-affinity binders to immune modulating receptors for cancer immunotherapy”. en. In: *Nature Communications* 16.1 (Feb. 2025), p. 2001. ISSN: 2041-1723. DOI: [10.1038/s41467-025-57192-z](https://doi.org/10.1038/s41467-025-57192-z). URL: <https://www.nature.com/articles/s41467-025-57192-z> (visited on 03/18/2026).
- [71] Longxing Cao, Brian Coventry, et al. “Design of protein-binding proteins from the target structure alone”. en. In: *Nature* 605.7910 (May 2022), pp. 551–560. ISSN: 0028-0836, 1476-4687. DOI: [10.1038/s41586-022-04654-9](https://doi.org/10.1038/s41586-022-04654-9). URL: <https://www.nature.com/articles/s41586-022-04654-9> (visited on 03/22/2026).
- [72] Carol A Rohl, Charlie EM Strauss, et al. “Protein structure prediction using Rosetta”. In: *Methods in enzymology*. Vol. 383. Elsevier, 2004, pp. 66–93.

# Supplementary information

## A. Wet-lab methods

### A.1. Cell-free protein production and purification

VHH constructs were expressed using an *E. coli*-based cell-free protein synthesis system. DNA sequences encoding each VHH were codon-optimized, synthesized, and subcloned into the pIVEX expression vector containing a C-terminal His-tag to facilitate affinity purification. Cell-free protein synthesis reactions were assembled by combining S30 cell lysate, synthesis buffer, required enzymes, and plasmid DNA in a 24-deep-well plate. Each reaction had a final volume of 5 mL and was incubated at 30°C for 6 hours. The reaction mixture was collected for purification. Proteins were purified using Ni-charged magnetic beads and dialysed into the desired buffer. The purified protein was filter-sterilised before storage. The concentration was determined by A280 protein assay, using a BSA standard. The protein purity was determined by standard SDS-PAGE confirmation. For this, samples were mixed with reducing loading buffer before running.

All target proteins and positive control proteins used in this study were commercially acquired as described in Tab. S1.

### A.2. Primary screening through high-throughput one-point SPR using Carterra LSA

Purified VHHs against all targets listed in Tab. S1 (IL-6[15], IL-6R[16], PRL[17], cTNFL9, hTNFL9[21], hTfR1[22], IL-33[18], TNF $\alpha$ [19], and SC2RBD[20]) were evaluated for binding using a high-throughput screen on the Carterra LSA platform. The assay was conducted at 25°C, with a Carterra LSA chip HC 200M and His capture kit. The His-tagged ligand was then injected. The analyte was diluted and injected over the surface for interaction analysis. All data were processed using Kinetics Evaluation Software. A reference channel and blank injections of running buffer were included in each cycle, serving as double references for the subtraction of resonance units (RU). For all targets, samples with  $R_{\max} > 100$  were considered “hits”. For each target, the corresponding natural ligand was included as a target-level positive control to verify assay performance, as described in Tab. S1.

### A.3. Affinity determination by five-point SPR via Biacore 8K

For targets IL-6, IL-6R, and PRL, further biophysical characterization of hits identified by one-point SPR was performed with five-point SPR. The assay was performed at 25°C using Biacore 8K, with chip Series S Sensor Chip CM5 and an anti-histidine antibody kit for capture of the ligand. The His-tagged ligand was captured on the chip surface, and diluted analyte samples were injected as five-point concentration series, at a flow rate of 30  $\mu$ L/min. Association and dissociation times were 120 s and 180 s.

The data were processed using Biacore 8K Evaluation Software (version 5.0). Sensorgrams were double-referenced using a blank reference surface and buffer-only injections to correct for nonspecific binding and bulk refractive index effects. Kinetic fitting was used. Fits were accepted based on a  $\chi^2$  value of less than 10% of  $R_{\max}$ . Global fitting of data to a 1:1 model across the whole concentration series was used to determine  $k_{\text{on}}$ ,  $k_{\text{off}}$ , and  $K_D$ . For each target, the corresponding natural ligand was included as a target-level positive control to verify assay performance, as shown in Tab. S1.

**Tab. S1 | Reagents used for experimental validation.** Target proteins and positive controls for each design campaign. Vendors: Acro = Acro Biosystems; Sino = Sino Biological. Oligomeric state: M = monomer; D = dimer; fD = Fc-dimer; T = trimer. Species: *H. s.* = *Homo sapiens*; Cyno = *Macaca fascicularis*, cynomolgus macaque.

Target Protein					Positive Control		
Name	Vendor	Cat. No.	Species	Oligomer	Name	Vendor	Cat. No.
TNFL9	Acro	41L-C5254	Cyno	T	TNFRSF9	Acro	41B-H53H3
TNFL9	Acro	41L-H5269	<i>H.s.</i>	T	TNFRSF9	Acro	41B-H53H3
TNF $\alpha$	Sino	10602-HNAE	<i>H.s.</i>	T	TNFR1	Sino	10872-H08H
hTfR1	Acro	TFR-H5213	<i>H.s.</i>	D	Transferrin	Acro	TRN-H52H3
IL-33	Sino	10368-HNAE	<i>H.s.</i>	M	ST2/IL-1 RL1	Sino	10105-H08H
IL-6	Sino	10395-HNAE	<i>H.s.</i>	M	IL-6R	Acro	ILR-H4223
IL-6R	Sino	10398-H02H	<i>H.s.</i>	fD	IL-6	Acro	IL6-H5243
SC2RBD	Acro	SPD-C5255	SARS-CoV-2	fD	ACE2	Sino	10108-H08H
PRL	Acro	PRN-H5257	<i>H.s.</i>	fD	PRLR	Acro	PRR-H52Ha

## B. Literature-inferred design benchmark

For benchmarking prompt fulfilment when inferring targets and epitopes from literature context we employed a single standardized prompt that was accompanied by a PDF version of each publication listed in Tab. S2:

I want to design 24 VHH binders to [TARGET\_NAME], that target the function described in the technical paper. You may use a sampling budget of up to 10 000 samples. When you submit any batches for this campaign, please use the project\_id=camp:xxx.

[TARGET\_NAME] was substituted with the corresponding target name entry in Tab. S2. Because most of the selected publications describe multiple potential targets, specifying a single target name allowed us to evaluate the agent’s ability to follow user-defined constraints, a prerequisite for reliable autonomous operation in a drug-discovery setting. The project\_id=camp:xxx called for the agent to use consistent project identifiers on the Latent Labs Platform, simplifying the benchmark analysis across the campaigns.

For hTfR1, the sampling budget was increased beyond the standard 10 000 samples to allow for a larger number of computationally passing binders for wet-lab validation. All remaining 20 targets used the standard budget of 10 000 samples as specified in the prompt.

**Prompt fulfilment metrics.** To quantify task completion, we evaluated each campaign against three prompt fulfilment criteria, summarized in Fig. 6.

*Correct epitope* was assessed by comparing the epitope engaged by the designed binders against the functional epitope described in the source publication. A campaign was considered successful on this criterion if the agent identified and targeted an epitope consistent with the published mechanism of action.

*Count met pre-QA* and *Count met post-QA* measure whether the agent fulfilled the prompt’s request for 24 computationally passing VHH designs, before and after quality assurance, respectively. QA comprises automated sequence-level analysis applied to all candidate designs, including removal of duplicate sequences, sequence similarity search against The Structural Antibody Database (SAbDab) to confirm novelty, and detection of common liability motifs including N-linked glycosylation sequons and unpaired cysteines within CDR regions. A campaign was considered to have met the count criterion if at least 24 designs remained after each respective filtering stage.

Together, these three metrics characterize the agent’s ability to follow scientific instructions, reason about biological context, and deliver a sufficient quantity of high-quality, diverse candidates within a fixed computational budget—the core requirements for reliable autonomous operation in a drug-discovery setting. The full list of publications and targets is provided in Tab. [S2](#).

**Tab. S2 | Structural data and literature for design campaigns.** Target names, PDB IDs, source publications, therapeutic areas, target classes, and interaction types.

Target	PDB ID	Publication Name	Therapeutic Area	Target Type	Type of PPI
TLR3	8YHT	De novo design of protein mimibinder-agonists of TLR3 [52]	Immunology	Toll-like receptor	Receptor: <i>de novo</i> mimibinder, agonist
BHRF1	40YD	A computationally designed inhibitor of an Epstein-Barr viral Bcl-2 protein induces apoptosis in infected cells [53]	Oncology / virology	Viral anti-apoptotic protein	Protein: <i>de novo</i> mimibinder, antagonist
CD80	1I8L	Crystal structure of the B7-1/CTLA-4 complex that inhibits human immune responses [54]	Immunology	Immune checkpoint	Receptor: Ligand
CDH17	7CYW	Mechanism of dimerization and structural features of human L1-cadherin [55]	Oncology / gastroenterology	Cell adhesion molecule	Self-dimerization
CEACAM6	4YIQ	Diverse oligomeric states of CEACAM IgV domains [56]	Oncology / immunology	Cell adhesion molecule	Protein oligomerization
CXCR2	8XVU	Molecular basis of promiscuous chemokine binding and structural mimicry at the C-X-C chemokine receptor, CXCR2[57]	Immunology / inflammation	GPCR	Receptor: Ligand
DLK1	9DZ0	Molecular mechanism of Activin receptor inhibition by DLK1 [58]	Metabolism / myology	Secreted signaling protein	Receptor: Ligand
FGFR1	1FQ9	Crystal structure of a ternary FGF-FGFR-heparin complex reveals a dual role for heparin in FGFR binding and dimerization [59]	Oncology	Growth factor receptor	Receptor: Ligand
FLT3	3Q57	Structural insights into the extracellular assembly of the hematopoietic Flt3 signaling complex [60]	Oncology / hematology	Receptor tyrosine kinase	Receptor: Ligand
hTTR	6W3H	Brain delivery of therapeutic proteins using an Fc fragment blood-brain barrier transport vehicle in mice and monkeys [27]	Neurology / drug delivery	Transport receptor	Receptor: Fc fragment
IL-18	3W04	The structural basis for receptor recognition of human interleukin-18 [61]	Immunology / inflammation	Cytokine	Receptor: Ligand
IL1R1	1ITB	Crystal structure of the type-I interleukin-1 receptor complexed with interleukin-1 beta [62]	Inflammation	Cytokine receptor	Receptor: Ligand
IL-33	4KC3	Structural insights into the interaction of IL-33 with its receptors [63]	Immunology	Cytokine	Receptor: Ligand
LGR5	4BSR	Structure of Stem Cell Growth Factor R-Spondin 1 in Complex with the Ectodomain of its Receptor Lgr5 [64]	Oncology / stem cell	Wnt pathway receptor	Receptor: Ligand
MDM2	1YCR	Structure of the MDM2 Oncoprotein Bound to the p53 Tumor Suppressor Transactivation Domain [65]	Oncology	Oncoprotein	Signaling protein: peptide
MET	6GCU	Inhibition of the MET Kinase Activity and Cell Growth in MET-Addicted Cancer Cells by Bi-Paratopic Linking[66]	Oncology	Receptor tyrosine kinase	Receptor: bi-paratopic binder
PARVA	2VZD	Structural Analysis of the Interactions between Paxillin Ld Motifs and Alpha-Parvin[67]	Oncology / metastasis	Adhesion adaptor protein	Protein: protein (intracellular)
PD-L2	6UWT	A high-affinity human PD-1/PPD-L2 complex informs avenues for small-molecule immune checkpoint drug discovery [68]	Oncology / immuno-oncology	Immune checkpoint	Receptor: Ligand
PVRIG	8X6B	Structural basis for the immune recognition and selectivity of the immune receptor PVRIG for ligand Nectin-2 [69]	Immunology / immuno-oncology	Immune checkpoint receptor	Receptor: Ligand
TGFβR2	8G4K	Design of high-affinity binders to immune modulating receptors for cancer immunotherapy [70]	Oncology / immuno-oncology	Cytokine receptor	Receptor: <i>de novo</i> mimibinder
ThkA	7N3T	Design of protein-binding proteins from the target structure alone [71]	Neurology / oncology	Receptor tyrosine kinase	Receptor: <i>de novo</i> mimibinder

## C. Binder sequences

**Tab. S3 | Sequences and binding affinities of best binders from epitope discovery campaigns.** Selected binders per target.  $K_D$  values were obtained by SPR, see Sec. A.3.

Binder	$K_D$ (M)	Amino Acid Sequence
LY_VHH_PRL_77	$5.44 \times 10^{-9}$	QVQLVESGGGLVQPGGSLRLSCAASAPSGYDLYFLDLGWFRQAPGGGL EAVAAINDFTGKTYADSVKGRFTISRDNKNTLYLQMNSLRAEDTAV YYCHADVLLVGGKSDPDDIKKSSAWGQGLTVTVSS
LY_VHH_IL6_67	$1.25 \times 10^{-8}$	EVQLVESGGGLVQPGGSLRLSCAASQPFVSGMVMGWFRQAPGKGRELV AAIRTSDGSTYYPDSVEGRFTISRDNKRMVYLLQMNSLRAEDTAVYYC AGTILPSSIPLSELTSDDFAYWGQGTQVTVSS
LY_VHH_IL6R_32	$5.17 \times 10^{-7}$	QVQLVESGGGLVQPGGSLRLSCAASLSSDFTFYDLLGWFRQAPGGGL EAVAAIDPVSGATYYADSVKGRFTISRDNKNTLYLQMNSLRAEDTAV YYCMMRGGDGITGGSTITYSDYWGQGLTVTVSS

## D. Expert baseline estimation and agent comparison for computational protein design

To establish baseline timelines for computational protein design workflows, we conducted structured interviews with ten PhD-level computational protein designers across academia and industry, recruited based on direct experience with binder design campaigns using current computational methods including AlphaFold [4], RFdiffusion [7], Rosetta [72], and commercial platforms. Participants had a median of 9.8 years of relevant experience since starting their PhD, spanning a range of 5–15 years.

Interviews followed a task-decomposed protocol to mitigate planning fallacy, whereby individuals systematically underestimate time requirements for complex tasks. Rather than requesting a single aggregate estimate, participants first described their typical workflow in an open-ended fashion, then provided time estimates for each major stage of an epitope discovery campaign: literature review and PDB structure selection, structural analysis and epitope selection, computational binder generation, and quality assessment and candidate selection. Participants then validated their aggregate estimate against the sum of individual stage estimates. The reference scenario specified designing a binder against a therapeutic target from initial target specification through to validated candidates ready for experimental testing, based on each participant’s typical workflows, tools, and computational resources. Estimates reflected total elapsed time from the expert’s perspective, including time waiting for computational jobs to complete. Where participants indicated team-based workflows, estimates were adjusted to reflect individual person-hours.

For each workflow stage, participants provided minimum, typical, and maximum time estimates. Final reported values represent the mean typical estimate across participants, with error bars indicating the range from mean minimum to mean maximum. All participants provided informed consent prior to interview, including consent for audio recording and use of anonymized estimates in published research. Participant identities and affiliations were kept confidential.

Agent-assisted timelines were obtained from five fully autonomous design campaigns targeting IL-33, IL-6, MCL-1, PRL, and SC2RBD. For the PRL campaign, the agent immediately identified the correct epitope and PDB structure, resulting in near-zero time for literature review, structural analysis, and epitope selection. To avoid distorting the stage-level comparison, this campaign was excluded from those two stages but retained in the overall workflow timing and all other stages.