Efficient generation of epitope-targeted *de novo* antibodies with Germinal

Luis S. Mille-Fragoso[1,2,3,†,*], John N. Wang[4,5,†], Claudia L. Driscoll[5,6,†], Haoyu Dai[6,†], Talal Widatalla[5,7], Xiaowei Zhang[1,2], Brian L. Hie[2,5,6,7,8,*], and Xiaojing J. Gao[2,3,6,7*]

[1]Department of Bioengineering, Stanford University, Stanford, CA, USA.
[2]Sarafan ChEM-H, Stanford University, Stanford, CA, USA.
[3]Stanford Bio-X, Stanford University, Stanford, CA, USA.
[4]Department of Computer Science, Stanford University, Stanford, CA, USA.
[5]Arc Institute, Palo Alto, CA, USA.
[6]Department of Chemical Engineering, Stanford University, Stanford, CA, USA.
[7]Stanford Biophysics, Stanford University, Stanford, CA, USA.
[8]Stanford Data Science, Stanford University, Stanford, CA, USA.

Abstract

Obtaining novel antibodies against specific protein targets is a widely important yet experimentally laborious process. Meanwhile, computational methods for antibody design have been limited by low success rates that currently require resource-intensive screening. Here, we introduce Germinal, a broadly enabling generative framework that designs antibodies against specific epitopes with nanomolar binding affinities while requiring only low-throughput experimental testing. Our method co-optimizes antibody structure and sequence by integrating a structure predictor with an antibody-specific protein language model to perform *de novo* design of functional complementarity-determining regions (CDRs) onto a user-specified structural framework. When tested against four diverse protein targets, Germinal achieved an experimental success rate of 4-22% across all targets, testing only 43-101 designs for each antigen. Validated nanobodies also exhibited robust expression in mammalian cells and nanomolar binding affinities. We provide open-source code and full computational and experimental protocols to facilitate wide adoption. Germinal represents a milestone in low-throughput, epitope-targeted *de novo* antibody design, with notable implications for the development of molecular tools and therapeutics.

1. Introduction

Antibodies play a central role in adaptive immunity by binding a remarkable diversity of molecular epitopes on target antigens, often with high specificity. Together with well-characterized biochemical properties and favorable therapeutic profiles, these capabilities have made antibodies heavily exploited as general-purpose and high specificity binders in biomedicine, biotechnology, and basic research (Bailly et al., 2020).

Antibody generation against arbitrary antigens currently requires animal immunization or large library screening campaigns (Fridy et al., 2024; Alexander and Leong, 2024). However, these experimental approaches are constrained by fundamental limitations. First, these processes are laborious and expensive, and do not always yield successful binders (Wilson and Andrews, 2012). Second, once binders are identified, their molecular and structural characterization can be difficult (Cheng et al., 2024; Fekete et al., 2013). Finally, these methods offer no control over which specific region of the antigen, known as the epitope, the antibody recognizes, making it difficult to direct binding to functionally important sites.

*Correspondence: L.S.M.-F. (lsmille@stanford.edu), B.L.H. (brianhie@stanford.edu), X.J.G. (xjgao@stanford.edu). †Equal contribution.

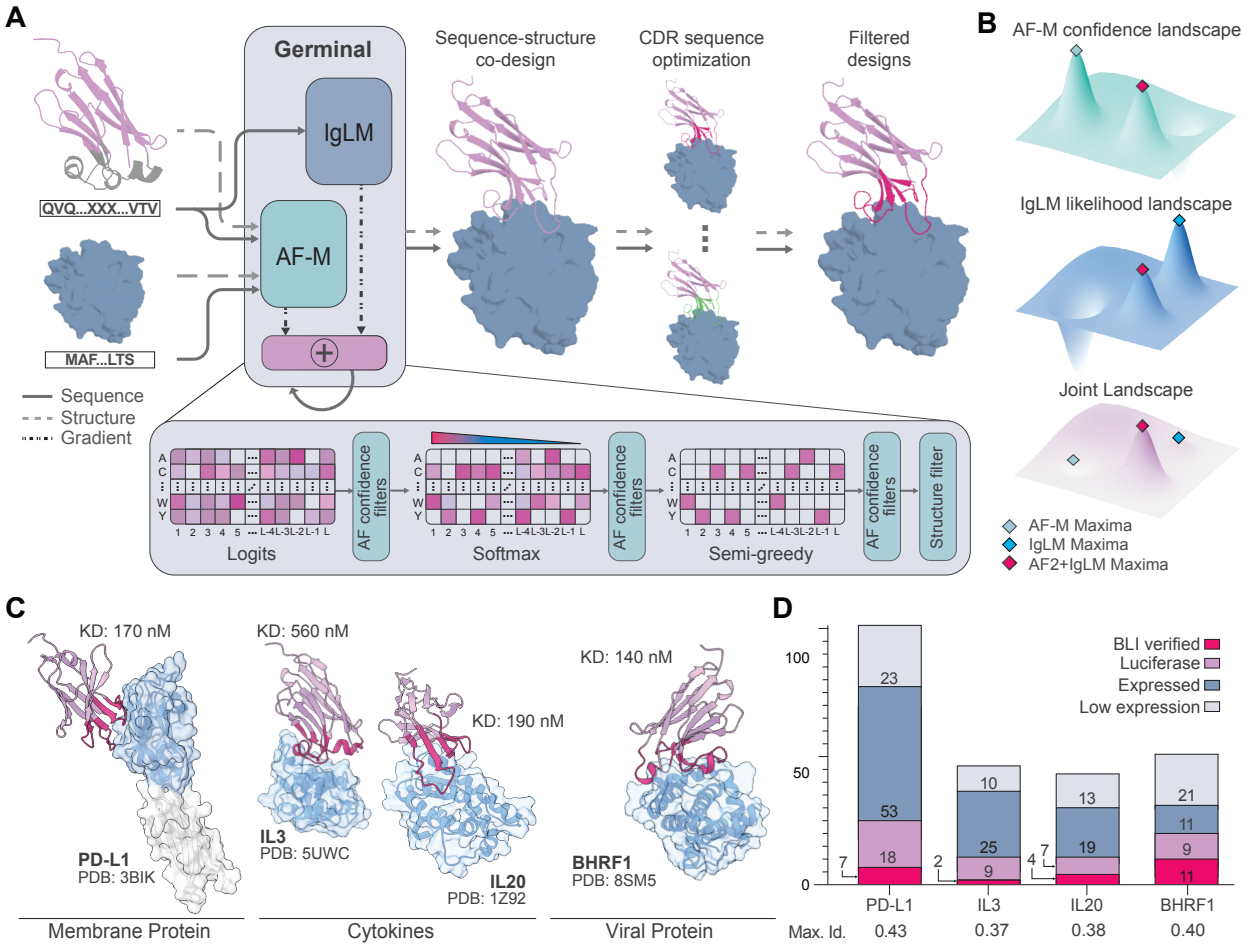

Figure 1 | **Efficient antibody generation via joint optimization of AlphaFold-Multimer and IgLM with Germinal.** (**A**) Overview of the Germinal pipeline. A target and antibody framework are provided as structural templates while CDRs are freely designed. Gradients from AlphaFold-Multimer (AF-M) and an antibody-specific language model (IgLM) are iteratively combined throughout the design stage to optimize sequences during three phases (Logits → Softmax (temperature annealing) → Semi-greedy). (**B**) Structure and sequence joint optimization landscape. AF-M (structure/complex-compatibility) and IgLM (antibody sequence prior) objectives define distinct optima. Their joint objective steers optimization toward sequences that are simultaneously structurally confident and antibody-like. (**C**) Representative predicted complexes across the four targets. Designed single-domain antibodies, or nanobodies, bound to specified epitopes on PD-L1, IL3, IL20, and BHRF1 are shown in pink (CDRs are highlighted). Domains used for design are colored in blue. White domains were excluded from the target structure used during design. Best dissociation constants (K_D) measured by BLI for each target are annotated. PDB entries shown beneath each target correspond to the structural examples used for visualization. (**D**) Screening results by target. Stacked bars summarize the per-library results (43 to 101 designs per target). Designs fall into four categories depending on their expression and binding profiles. Low-expression (gray), expressed but filtered as non-binder in split-luciferase assay (blue), binders exhibiting significant signals in the split-luciferase assay (pink), and BLI-verified binders (magenta). Numbers indicate counts within each category. BLI-verified success rates range from 4 to 22%. “Max. Id.” denotes the highest pairwise CDR sequence identity of any design in the library to the PDB. AF-M: AlphaFold-Multimer. IgLM: antibody-specific language model. CDR: complementarity-determining region. BLI: bio-layer interferometry. K_D: dissociation constant.

Advances in machine learning have led to atomically accurate structure prediction of protein monomers and multimeric complexes with methods such as AlphaFold (Jumper et al., 2021; Evans et al., 2022). Subsequent efforts have also leveraged these structure predictors to design protein sequences that have desirable structural properties, including binding interactions. For example, methods such as RFdiffusion have fine-tuned structure prediction models with diffusion objectives to generate the backbone coordinates of potential protein binders (Watson et al., 2023; Baek et al., 2021). Alternatively, several studies have leveraged gradient-based methods that invert structure prediction to generate proteins and protein binders (Frank et al., 2024; Jendrusch et al., 2025; Goverde et al., 2023; Wicky et al., 2022). Recently, BindCraft inverted AF-M to obtain high experimental success rates for *de novo* miniproteins that bind arbitrary targets (Pacesa et al., 2025; Evans et al., 2022).

However, despite numerous preliminary efforts (Bennett et al., 2024; Shanehsazzadeh et al., 2023; Nabla Bio and Biswas, 2025), robust *de novo* antibody design against specific target epitopes with high success rates remains challenging, especially due to the inherent flexibility of their complementarity determining regions (CDRs) and the highly constrained nature of antibody sequence space. For example, efforts to design novel CDRs have leveraged large-scale library screens containing thousands of designs to ultimately yield a handful of binders, often with activity in the micromolar range (Bennett et al., 2024). These low experimental success rates require resource-intensive screening campaigns that are inaccessible to most laboratories. In contrast, advances in generative modeling that reduce experimental validation to tens rather than thousands of designs would democratize access to these technologies across molecular biology.

Here, we present Germinal, a broadly enabling generative pipeline for *de novo* antibody design (**Figure 1A**). In particular, Germinal achieves epitope-targeted, *de novo* CDR design with unprecedented experimental success rates that enable low-throughput experimental testing. Germinal leverages a computational pipeline that biases protein binders to a specified antibody framework while freely designing the CDRs, a strategy that not only maintains robust expression in human cells, but also preserves the therapeutic developability profiles conferred by known framework regions (FR). To achieve this, we combine backpropagation of AlphaFold-Multimer (AF-M) (Evans et al., 2022) with an antibody-specific protein language model (Shuai et al., 2023), allowing us to produce designs with realistic CDR sequences (**Figure 1B**). Similarly, Germinal uses custom loss functions that ensure correct spatial positioning of the designs to favor CDR over framework binding contacts. We experimentally validated designs for mammalian cell expression and specific binding using a split-luciferase assay followed by bio-layer interferometry (BLI), enabling efficient identification of high-affinity binders.

Across epitopes from four distinct antigens (**Figure 1C**), we observed experimentally validated binding with 4-22% success rates after testing libraries consisting of only 43 to 101 *de novo* nanobodies, for each antigen (**Figure 1D**). BLI characterization of a subset of the validated binders further revealed 2-11 binders per target with sub-micromolar dissociation constants (K_D).

To promote broad accessibility, we make our full computational pipeline open-source and freely available to the scientific community, alongside our efficient experimental protocol for initial binder screening. We anticipate that Germinal will greatly reduce the need for the extensive experimental infrastructure developed specifically for antibody discovery. By supporting precise epitope targeting while maintaining favorable antibody properties, Germinal opens new possibilities for the development of antibody molecules for molecular biology and therapeutic design.

2. Results

2.1. Design of antibodies via dual-objective structure and sequence optimization

Computational design of *de novo* antibodies remains challenging due to their specific structural constraints, the high variability of their binding regions (CDRs), and a lack of antibody training data for structure prediction tools (Ruffolo et al., 2023; Hitawala and Gray, 2024). Binder design models based on these structure predictors are biased toward protein–protein interfaces dominated by secondary structures rather than the more flexible conformations characteristic of CDRs (Watson et al., 2023; Pacesa et al., 2025). Recent work has also shown that general structure-based models alone struggle to redesign functional CDRs (Alamo et al., 2025). Meanwhile, vast amounts of antibody sequence data have led to antibody-specific language models that have learned the distribution of natural antibody sequences (Shuai et al., 2023; Olsen et al., 2022a,b). We hypothesized that integrating structure prediction models with antibody-specific language models could overcome challenges of data scarcity and structural bias, unlocking efficient *de novo* antibody design.

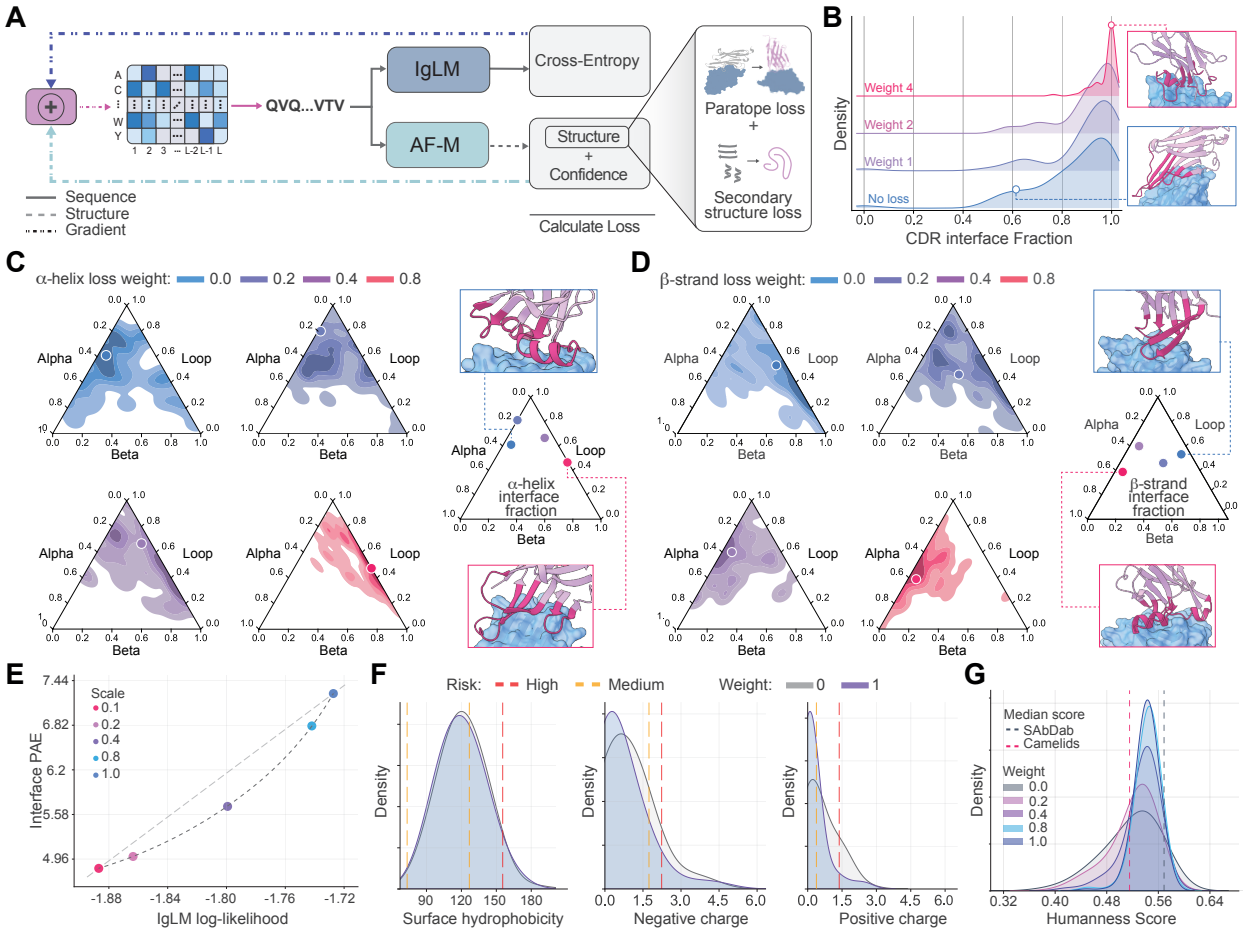

Figure 2 | **Germinal steers generated designs toward natural antibody-like properties through structural and sequence guidance.** (**A**) Schematic of Germinal's design stage. Structural losses and IgLM sequence guidance are integrated via gradient combination. (**B**) Paratope loss ensures binding occurs primarily through CDR regions rather than framework residues. (**C-D**) Secondary structure losses prevent CDRs from being dominated by α-helices and β-strands. Ternary kernel density plots display the secondary structure composition of the interface for designed nanobodies. Selected points represent the median interface fraction for designs generated at each weight. (**E**) Pareto frontier between AF-M structural confidence, shown as iPAE, and IgLM sequence likelihood reveals competing optimization objectives. (**F**) Comparison of nanobody developability properties from the Therapeutic Nanobody Profiler (TNP) between sequences generated with (weight of 1.0) and without (weight of 0) IgLM guidance. Designs within the amber region are defined as low risk, between amber and red, medium risk, and above red, high risk. (**G**) Comparison of the OASis humanness score amongst sequences generated with varying weighting of IgLM guidance. Shown in gray and pink are the median humanness scores of all nanobodies, and all camelid nanobodies, respectively, in the SAbDab-nano database.

Germinal integrates a protein structure predictor, AF-M (Evans et al., 2022), with an antibody-specific protein language model, IgLM (Shuai et al., 2023) (**Figure 2A**), by merging their gradients to co-optimize both fold geometry and antibody naturalness in a joint optimization landscape (**Figure 1B**). Germinal's generation pipeline consists of three main stages: design, sequence optimization, and filtering.

In the design stage, Germinal leverages the gradients of AF-M to sample sequences that, with high confidence, bind desired epitopes in their predicted structures. Germinal also leverages the gradients of IgLM to bias sampling toward sequences with high probability under the language model, reflecting their likelihood of resembling naturally occurring antibodies. During this stage, Germinal also allows users to explicitly bias samples toward a desired antibody framework sequence (e.g., a framework with a favorable developability profile) (**Figure S1**), as well as augmenting structure guidance by providing the framework's structure as a template to AF-M (**Methods**).

Sampled sequences with high-quality predicted structures are then carried forward to a sequence optimization stage. In this stage, we use AbMPNN (Dreyer et al., 2023; Dauparas et al., 2022), a structure-conditioned sequence design model fine-tuned on antibodies, to redesign CDR residues that are not in direct contact with the antigen. This approach aims to improve binder stability while preserving the binding interface (**Methods**).

Finally, designs are passed through a filtering stage, where we use a different structure predictor from the one used during the design stage to provide a separate assessment of design quality; in particular, we used AlphaFold 3 (AF3) due to its superior accuracy on antibody–antigen complexes (Abramson et al., 2024; Hitawala and Gray, 2024). We filter and rank designs based on AF3 confidence scores alongside PyRosetta-derived biophysical and biochemical scores (Chaudhury et al., 2010) (**Table S1**), yielding a final set of candidates for experimental testing (**Methods**).

We initially found that naive application of AF-M guidance towards antibody binder design resulted in generations with paratopes (the part of the antibody that contacts the target) composed of framework residues or enriched with secondary-structures (**Figure 2C,D**). To address this, Germinal incorporates three custom loss functions that effectively guide generations toward natural antibody binding conformations in contrast to those favored by naive AF-M. A paratope-specific loss ensures that binding occurs primarily through the designed CDRs (**Figure 2B**). Additionally, Germinal contains two secondary-structure losses, an α-helix loss and β-strand loss, that encourage the generated CDRs to adopt flexible loop conformations (**Figure 2C,D**). Together, these losses encourage designs toward binding interfaces dominated by CDRs (Fernández-Quintero et al., 2018; Kunik et al., 2012; Tsuchiya and Mizuguchi, 2016) while promoting their conformational flexibility, a feature that has been linked to enhanced antigen specificity and the capacity to generate broadly neutralizing binders in natural antibodies (Li et al., 2015; Prigent et al., 2018).

Combining AF-M structure guidance with IgLM sequence guidance is central to our approach. Notably, we found that the preference an antibody language model has for a given sequence (IgLM log-likelihood) and the predicted binding confidence (AF-M interface predicted aligned error; iPAE) are competing objectives, exhibiting a trade-off that yields a Pareto frontier and underscoring the need for joint optimization (**Figure 2E** and **S3**). Upon increasing the strength of IgLM guidance, we observed improvements in therapeutic developability and safety profiles for generated sequences as predicted by the Therapeutic Nanobody Profiler (TNP) (Gordon et al., 2025) (**Figure 2F**). Additionally, we found that IgLM guidance improves the "humanness" and reduces the immunogenicity risk of generated sequences, as predicted by the OASis tool (Prihoda et al., 2022), which calculates the similarity of a given sequence against a large database of human antibodies (Olsen et al., 2022a) (**Figure 2G**). In total, Germinal enables controllable, multi-objective generation of antibody-like sequences.

2.2. Using Germinal to target diverse antigens

We next sought to apply Germinal to four diverse targets: Protein Death Ligand 1 (PD-L1), an immune checkpoint ligand expressed on both tumor and immune cells, which is both a clinically validated therapeutic target and a frequent target of *de novo* binder design efforts; interleukin-3 (IL3) and interleukin-20 (IL20), cytokines involved in immune signaling for which no *de novo* binders have been reported to date; and BHRF1, a viral Bcl-2–like anti-apoptotic protein from the Epstein–Barr virus, as a representative example of a non-human, pathogen-derived target.

After sampling, we observed that 1,584 nanobody trajectories for PD-L1, 1,379 trajectories for IL3, 739 trajectories for IL20, and 1,456 trajectories for BHRF1 had passed initial filters. Notably, many of these designs

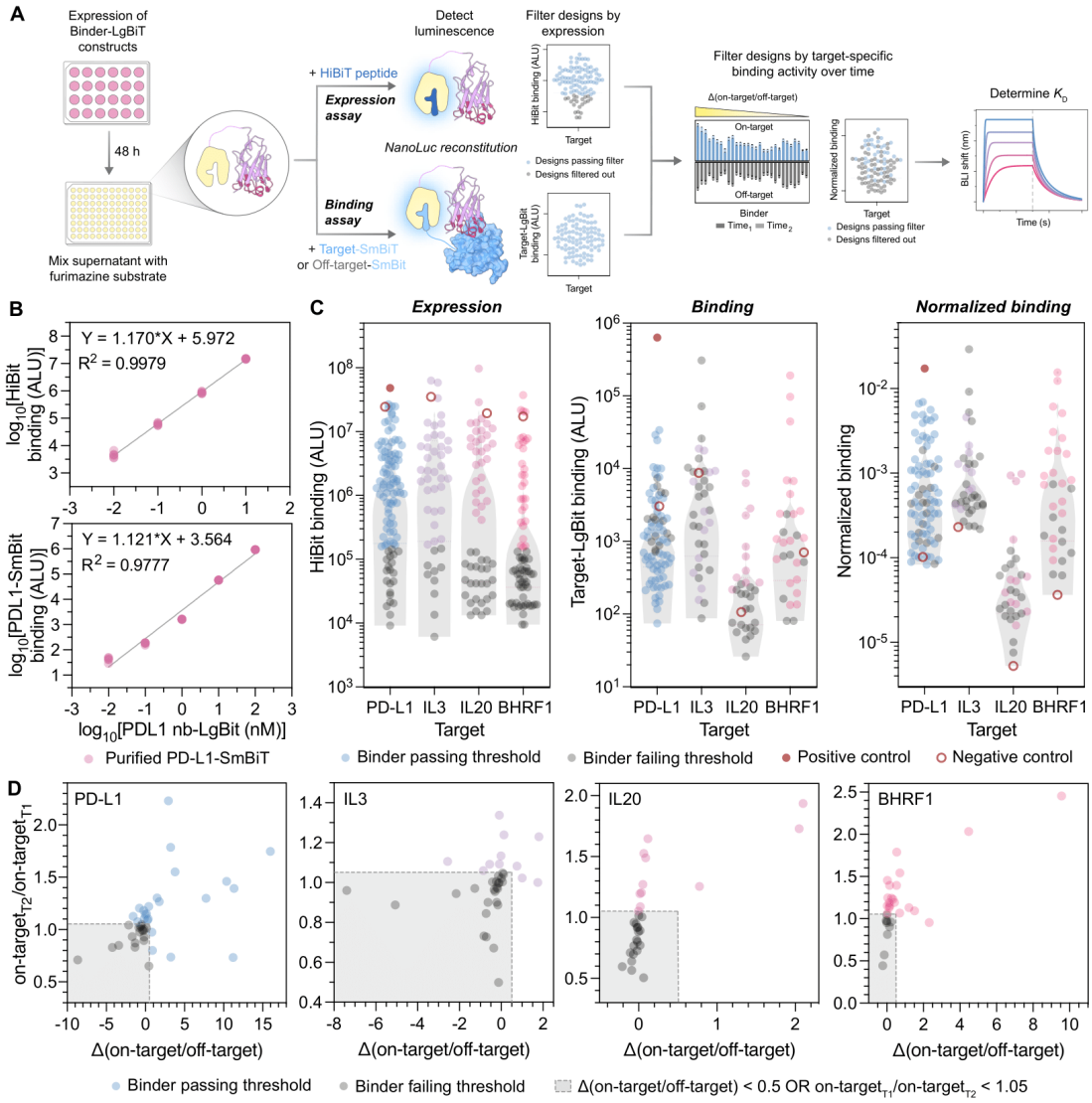

Figure 3 | **Screening expression and binding of designed binders using a split-luciferase assay.** (**A**) Experimental workflow used to identify lead binders. Binder-LgBiT fusions were assayed via luminescence measurements for (i) expression by adding HiBiT peptide and (ii) binding by adding either desired target fused to SmBiT protein ('on-target' binding), or an off-target protein-SmBiT fusion ('off-target' binding). Designs with favorable expression and binding activity in this assay were further characterized with BLI. (**B**) *Top:* expression standard curve. $\log_{10}$–$\log_{10}$ linear relationship between HiBiT-LgBit produced luminescence reads and binder concentration (fit and R^2 shown). *Bottom:* binding standard curve. $\log_{10}$–$\log_{10}$ linear relationship between PD-L1–SmBiT luminescence reads and concentration of a known PD-L1 nanobody–LgBiT (reported $K_D \approx 3$ nM) (fit and R^2 shown). (**C**) *Left:* designed binder expression, shown as raw luminescence (arbitrary luminescence units, ALU). Designs shown in gray fall below the expression threshold (raw HiBiT signal > 150,000 ALU). *Middle:* Binding activity for all designs that passed the expression threshold, shown as raw luminescence. *Right:* expression-normalized binding estimates used for identification of lead binder candidates. The positive control nanobody against PD-L1 is shown as a solid red circle, while negative control nanobody against TNFα (for PD-L1 and IL3 screens) or against PD-L1 (for IL20 and BHRF1 screens) is indicated by a hollow red circle. Designs shown in gray failed to meet the binding threshold. (**D**) A custom binding filter was lastly applied to design candidates that passed the expression threshold. Here, binders were selected if the difference in the normalized on-target/normalized off-target binding ratio between T_1 and T_2 was greater than 0.5, or if the ratio of normalized on-target binding (on-target T_2/on-target T_1) was greater than 1.05 (**Methods**). Passing candidates are shown as colored circles, while candidates that fall below the binding threshold (shaded gray region) are shown as gray circles.

had promising computational metrics (**Figure S2**) and low CDR sequence identity (<50%) (**Figure S4**) to any existing sequence in SAbDab (Dunbar et al., 2014) or OAS (Olsen et al., 2022a) . We then ranked these designs for experimental screening based on a diverse set of metrics (**Table S2**), ultimately selecting the top 101 designs for PD-L1, 46 designs for IL3, 43 designs for IL20, and 52 designs for BHRF1 for downstream validation. Notably, all input target structures to the pipeline are AF3 predictions and not the respective PDB structure, highlighting Germinal's potential to generate binders even in the absence of an experimentally determined target structure.

We also used Germinal to design single-chain variable fragments (scFv) against these targets. The resulting sequences passed our filtering criteria at rates comparable to nanobody design runs and achieved similarly favorable values across key evaluation metrics (**Figure S2**, **S5**, and **S6**). Notably, several of the metrics identified as strong predictors of nanobody binding success are based on interface contact area (for example, LIA, number of hydrogen bonds, and number of interface residues) (**Figure S7**), and thus are naturally improved in scFv designs due to the presence of two binding chains (**Figure S5** and **S6**). These *in silico* results suggest that the design quality of Germinal-generated scFvs is similar to that of Germinal-generated nanobodies.

2.3. Experimental characterization of generated nanobodies

We next sought to experimentally validate the designs produced by Germinal. To provide a rapid and accessible readout of both binder expression and antigen binding, we employed an initial NanoBiT screen based on a split-luciferase assay (**Figure 3A**) (Dixon et al., 2016). The system is based on the large luciferase subunit (LgBiT), which we fuse to our binders and can be complemented by either a high-affinity peptide (HiBiT, K_D = 0.7 nM) or a very low-affinity peptide (SmBiT, K_D = 190 μM). Thus, the addition of HiBiT allows for direct quantification of LgBiT-tagged binder expression, while the addition of an antigen-SmBiT fusion protein enables the detection of potential binders.

To validate that the luminescence signal provides a reliable and quantitative measurement, we leveraged a known PD-L1 binder, KN035 (Zhang et al., 2017) (PDB: 5JDS, K_D = 3 nM), at varying concentrations. We confirmed a strong linear relationship between luminescence and KN035 concentration upon addition of 20 nM HiBiT to quantify expression (**Figure 3B** *top*) and upon addition of 20 nM of PD-L1-SmBit fusion protein to test antigen binding (**Figure 3B** *bottom*).

For all four targets, designed nanobodies were fused to LgBiT and encoded in plasmids that were transiently transfected into HEK293 cells. Alongside our designs, we used a validated nanobody against the same antigen when available as positive control, and a nanobody targeting an unrelated antigen as negative control (**Methods**). We measured luminescence directly from the supernatants after addition of HiBiT (expression) or of purified recombinant antigen-SmBiT fusions (binding). Encouragingly, measured expression levels from the HiBiT assay (**Figure 3C** *left*) correlated well with protein yields obtained from independent recombinant expression of nanobodies in Expi293F cells (**Figure S8**).

To identify promising candidates, binding measurements (SmBiT–LgBiT) were normalized to the corresponding expression measurements (HiBiT–LgBiT), providing a relative measure of binding potential. We selected candidates for further characterization if they met two criteria: (i) expression luminescence was above 150,000 RLU (**Figure 3C** *middle*) and (ii) we observed an increase in normalized binding luminescence between two time points (**Figure 3C** *right*, **3D**, and **S9**; **Methods**). To obtain more quantitative measurements of binding affinity, all nanobody designs that passed these expression and binding filters were expressed without LgBiT fusion, purified, and characterized by BLI (**Methods**). As a result, we tested 25 out of 101 PD-L1 nanobodies, 11 out of 43 IL20 nanobodies, 11 out of 46 IL3 nanobodies, and 20 out of 52 BHRF1 nanobodies for binding to their respective antigens.

We observed detectable binding for 7 of 25 luciferase-passing PD-L1 designs (**Figure 4A-C**), 2 of 11 IL3 designs, 4 of 11 IL20 designs, and 11 of 20 BHRF1 designs (**Figure S10**), with best binders having a K_D of 170 nM for PD-L1 (**Figure 4C**), 560 nM for IL3 (**Figure 4D**), 190 nM for IL20 (**Figure 4E**), and 140 nM for BHRF1 (**Figure 4F**). Notably, we obtained nanomolar binders for all four antigens tested.

To further assess whether binders targeted the intended epitope, we measured the effects of alanine mutations to the putative epitopes of PD-L1 and IL3. For PD-L1, we introduced alanine substitutions at three predicted interface residues: I54, Y56, and E58. Each substitution reduced binding signal and weakened affinity by >10-fold, consistent with these residues contributing directly to the nanobody–PD-L1 interface. We also

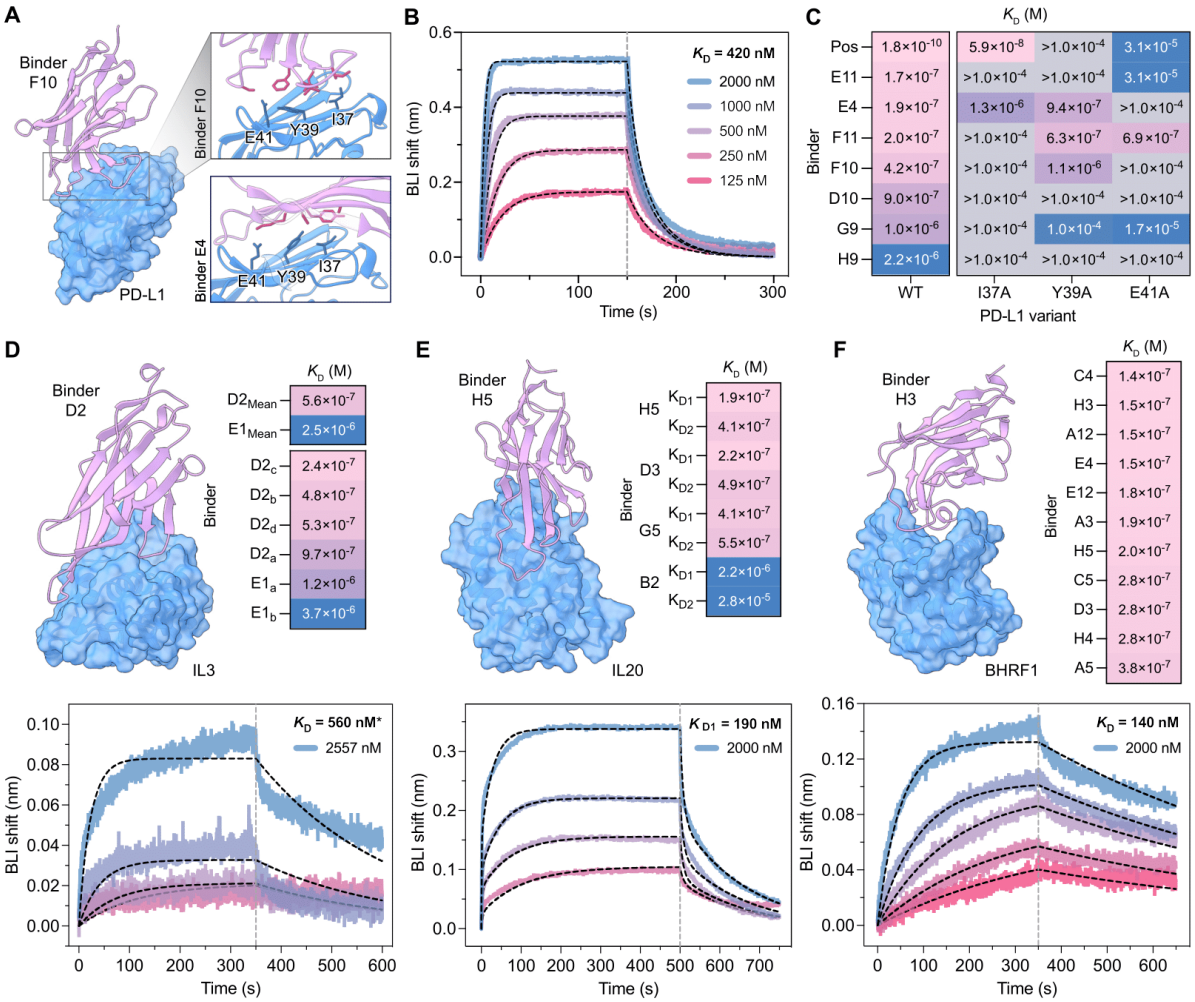

Figure 4 | **Biophysical characterization of hit nanobodies and affinity determination.** (**A**) A representative structure predicted by AF3 of an anti-PD-L1 designed nanobody, F10, is shown in complex with PD-L1 target (*left*). Designed binders are shown in pink as cartoon representations, while targets are shown as surfaces in blue. The target-binder interfaces are visualized in the zoom-boxes (*right*), revealing key PD-L1 surface residues that were mutated to alanine for epitope investigation. (**B**) Representative BLI traces for PD-L1 binder F10 at multiple analyte concentrations. The estimated K_D is shown in bold. The vertical dashed line indicates the time at which the association phase ends and the dissociation phase begins. Binding curves are shown as colored lines, and kinetic fits are shown as black dashed lines. (**C**) Heat-map of estimated affinities for PD-L1 binders when screened against PD-L1 variants. Designs showing clear binding kinetics and high-quality fits ($R^2 > 0.95$) on BLI when tested against wild-type PD-L1 were tested against different three PD-L1 variants with an alanine substitution at each of the key interface residues indicated in (**A**), consistently resulting in reduced affinities. Binding events that fell below the limit of detection for BLI were assigned a K_D of $>1.0\times10^{-4}$ M, indicating weak or negligible binding interactions. (**D**) Representative structure predicted by AF3 of the anti-IL3 binder, D2, in complex with IL3 (*top left*). Affinities for IL3 designs identified to be hits on BLI are shown in the heat map (*upper top right*), and a representative BLI sensogram for D2 titrated against IL3 is shown (*bottom*). For both IL3 binders, due to poorer BLI curve fits ($R^2 = 0.93$), we calculated their K_Ds as the mean of at least two replicate BLI experiments per binder (*lower top right*; reported with an asterisk on the BLI plot). (**E**) Representative structure predicted by AF3 of the anti-IL20 binder H5 in complex with IL20 (*top left*), and heat-map of IL20 designs identified as hits on BLI (*top right*). IL20 kinetic curves were fitted to second-order binding kinetics with both K_{D1} and K_{D2} indicated on the heat-map. A representative BLI sensogram is shown for binder H5 (*bottom*). (**F**) Representative AF-M structure of anti-BHRF1 binder H3 in complex with BHRF1 (*top left*). The affinities for each BHRF1 binder characterized as a hit by BLI are indicated in the heat-map (*top right*), with a representative BLI sensogram shown below for binder H3 (*bottom*).

confirmed that these mutations did not broadly disrupt antigen folding, as nanobodies with distinct binding modes were differentially affected, with no single mutation abolishing binding across all designs (**Figure 4C** and **S11**).

We note that because the main screening protocol was based on the split-luciferase assay—which primarily reflects apparent association and is influenced by transfection efficiency, cell debris, and other assay artifacts—the reported BLI success rates are likely conservative. Additionally, while no strong cross-reactivity was observed in our luciferase (**Figure S9**) or BLI (**Figure S12**) experiments, normalized luminescence from the SmBiT assay is still dependent on the target antigen and more comprehensive poly-specificity measurements will be required to rigorously confirm antigen specificity.

3. Discussion

We present an end-to-end pipeline for *de novo* nanobody design that achieves nanomolar binders with low-throughput experimental validation. A high-throughput split-luciferase assay was devised to initially screen and filter computational binder designs, efficiently identifying lead binders for subsequent validation by BLI. This pipeline yielded nanobody binders against four diverse soluble protein targets, with a success rate of 4-22%. Importantly, we make our full methodological details and code publicly available to the scientific community.

Our method differs from previous studies that have leveraged protein language models to guide directed evolution of antibodies toward improved binding affinity, antibody stability, or other therapeutically relevant properties (Hie et al., 2024; Shanker et al., 2024; Widatalla et al., 2024). These efforts have assumed at least weak initial function (for example, affinity maturation of a weak antibody binder), whereas Germinal designs antibody CDRs completely *de novo* without any starting binder, and we note that protein language models could enable efficient affinity maturation of our generated antibodies as well. Concurrent with our work, Chai Discovery Team et al. (2025) describe nanobody and scFv design methods that identify binders based on BLI signal above background noise after screening ~20 designs per antigen, achieving this binding criterion in approximately half of tested antigens, though with unreported methods and code.

Despite unprecedented capabilities, we note several limitations with Germinal that present clear opportunities for future development. First, the method's performance is fundamentally constrained by the confidence and accuracy of AF-M predictions. Additionally, the need for structure prediction and network backpropagation at each iteration requires extensive inference-time sampling to generate successful designs, rendering it computationally intensive; future improvements in optimization algorithms and integration with other generative approaches could significantly reduce this computational burden. Memory constraints also limit the method to smaller protein targets, though most epitopes can be successfully targeted by truncating large proteins to the regions of interest. Furthermore, while our luciferase-based screening assay efficiently narrows down pools of candidate designs to true binders, it currently does not provide quantitative binding affinities, necessitating secondary characterization to determine true dissociation constants. Finally, Germinal is currently restricted to protein epitopes, leaving important classes of targets such as glycans, small molecules, and other non-protein antigens beyond its scope.

The ability to computationally design epitope-targeted antibodies with high success rates opens transformative possibilities across biotechnology and medicine. Germinal could unlock targeting of previously inaccessible epitopes, including intracellular domains, specific conformational states, and highly conserved regions that fail to elicit natural immune responses. Rapid antibody design could compress therapeutic antibody identification from weeks to days, improving our response to novel pathogens or rapidly evolving disease targets. As our generative pipeline becomes more efficient, systematic design campaigns could produce antibodies against entire proteomes, vastly expanding our repertoire of affinity reagents and therapeutic molecules.

4. Acknowledgments

We thank Connor C. Call, Julia Kazaks, and Brian Plosky for helpful discussions and support with the manuscript. B.L.H. acknowledges funding support from Arc Institute, the Gates Foundation, Stanford Institute for Human-Centered Artificial Intelligence (HAI) Hoffman-Yee Research Grants, V. Gupta, and R. Tonsing. T.W. acknowledges support by the NSF Graduate Research Fellowship and the Stanford Graduate Fellowship. L.S.M acknowledges support by the Stanford Bio-X fellowship, the Stanford Graduate Fellowship, the Stanford University Sarafan ChEM-H Chemistry–Biology Interface training program and the Stanford Interdisciplinary Graduate Fellowship. X.Z. is supported by the Stanford Interdisciplinary Graduate Fellowship affiliated with ChEM-H. X.Z. is supported by the Stanford Interdisciplinary Graduate Fellowship affiliated with ChEM-H This work was funded by the National Institutes of Health (NIH DP2EB035891 to X.J.G.), Longevity Impetus grants (to X.J.G.), and Stanford Bio-X Interdisciplinary Initiatives seed grant program (to X.J.G.).

5. Author Contributions

L.S.M., J.N.W., T.W., B.L.H., and X.J.G. conceived the study. L.S.M., J.N.W., C.L.D., H.D., X.Z. designed the study. B.L.H and X.J.G supervised the project. L.S.M. and J.N.W. developed the algorithms and pipeline. L.S.M., J.N.W., and T.W. performed computational analysis of designs. L.S.M. and J.N.W. ran design campaigns and selected designs for experiments. L.S.M, J.N.W, C.L.D, and H.D. cloned plasmids used in this work. L.S.M, C.L.D, H.D. and X.Z. ran luciferase assays. C.L.D, and H.D. performed protein purification. L.S.M, C.L.D, and H.D. ran BLI validation. H.D. performed differential analysis on computational and experimental data. L.S.M, J.N.W., C.L.D., H.D., T.W., B.L.H, and X.J.G wrote the manuscript.

6. Competing Interests

B.L.H. acknowledges outside interest in Arpelos Biosciences and Genyro as a scientific co-founder. X.J.G. is a cofounder and serves on the scientific advisory board of Radar Tx. L.S.M, J.N.W., C.L.D., H.D., T.W., X.Z., B.L.H, and X.J.G are named on a provisional patent application applied for by Stanford University and Arc Institute related to this manuscript.

7. Code and Data Availability

We make code for running the pipeline available at `http://www.github.com/SantiagoMille/germinal`.

We make the following files available as supplementary data:

- **Data S1:** Template plasmid maps used in this work.
  `supp_file1_plasmidmaps.xlsx`
- **Data S2:** Designed sequences per target and their corresponding names.
  `supp_file2_all_seq_for_binders.xlsx`

References

J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, 630(8016):493–500, June 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07487-w. URL https://www.nature.com/articles/s41586-024-07487-w. Publisher: Nature Publishing Group.

D. d. Alamo, R. Frick, D. Truan, and J. Karpiak. Adapting ProteinMPNN for antibody design without retraining, May 2025. URL https://www.biorxiv.org/content/10.1101/2025.05.09.653228v1. Pages: 2025.05.09.653228 Section: New Results.

E. Alexander and K. W. Leong. Discovery of nanobodies: a comprehensive review of their applications and potential over the past five years. *J Nanobiotechnology*, 22(1):661, oct 2024. doi: 10.1186/s12951-024-02900-y. URL http://dx.doi.org/10.1186/s12951-024-02900-y.

M. Baek, F. Dimaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer, and et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, 2021. ISSN 0036-8075. doi: 10.1126/science.abj8754.

M. Bailly, C. Mieczkowski, V. Juan, E. Metwally, D. Tomazela, J. Baker, M. Uchida, E. Kofman, F. Raoufi, S. Motlagh, and et al. Predicting antibody developability profiles through early stage discovery screening. *mAbs*, 12(1):1743053, 2020. ISSN 1942-0862. doi: 10.1080/19420862.2020.1743053.

N. R. Bennett, J. L. Watson, R. J. Ragotte, A. J. Borst, D. L. See, et al. Atomically accurate de novo design of single-domain antibodies, Mar. 2024. URL https://www.biorxiv.org/content/10.1101/2024.03.14.585103v1. Pages: 2024.03.14.585103 Section: New Results.

L. Borras, T. Gunde, J. Tietz, U. Bauer, V. Hulmann-Cottier, J. P. A. Grimshaw, and D. M. Urech. Generic approach for the generation of stable humanized single-chain fv fragments from rabbit monoclonal antibodies. *J Biol Chem*, 285(12):9054–9066, mar 2010. doi: 10.1074/jbc.M109.072876. URL http://dx.doi.org/10.1074/jbc.M109.072876.

K. T. Byun, B. Kim, J. Cho, I. Lee, M. G. Lee, D. Park, T.-B. Kang, H.-S. Won, and C. G. Kim. Development of an anti-HER2 single-chain variable antibody fragment construct for high-yield soluble expression in escherichia coli and one-step chromatographic purification. *Biomolecules*, 13(10), oct 2023. doi: 10.3390/biom13101508. URL http://dx.doi.org/10.3390/biom13101508.

Chai Discovery Team, J. Boitreaud, J. Dent, D. Geisz, M. McPartlon, J. Meier, Z. Qiao, A. Rogozhnikov, N. Rollins, P. Wollenhaupt, and K. Wu. Zero-shot antibody design in a 24-well plate. *bioRxiv*, 2025. doi: 10.1101/2025.07.05.663018. URL https://www.biorxiv.org/content/early/2025/07/06/2025.07.05.663018.

S. Chaudhury, S. Lyskov, and J. J. Gray. PyRosetta: a script-based interface for implementing molecular modeling algorithms using rosetta. *Bioinformatics*, 26(5):689–691, mar 2010. doi: 10.1093/bioinformatics/btq007. URL http://dx.doi.org/10.1093/bioinformatics/btq007.

J. Cheng, T. Liang, X.-Q. Xie, Z. Feng, and L. Meng. A new era of antibody discovery: an in-depth review of AI-driven approaches. *Drug Discov Today*, 29(6):103984, jun 2024. doi: 10.1016/j.drudis.2024.103984. URL http://dx.doi.org/10.1016/j.drudis.2024.103984.

J. Dauparas, I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, A. Courbet, R. J. de Haas, N. Bethel, P. J. Y. Leung, T. F. Huddy, S. Pellock, D. Tischer, F. Chan, B. Koepnick, H. Nguyen, A. Kang, B. Sankaran, A. K. Bera, N. P. King, and D. Baker. Robust deep learning–based protein sequence design using proteinmpnn. *Science*, 378(6615):49–56, 2022. doi: 10.1126/science.add2187. URL https://www.science.org/doi/abs/10.1126/science.add2187.

A. S. Dixon, M. K. Schwinn, M. P. Hall, K. Zimmerman, P. Otto, T. H. Lubben, B. L. Butler, B. F. Binkowski, T. Machleidt, T. A. Kirkland, M. G. Wood, C. T. Eggers, L. P. Encell, and K. V. Wood. Nanoluc complementation reporter optimized for accurate measurement of protein interactions in cells. *ACS Chem Biol*, 11(2):400–408, feb 2016. doi: 10.1021/acschembio.5b00753. URL http://dx.doi.org/10.1021/acschembio.5b00753.

F. A. Dreyer, D. Cutting, C. Schneider, H. Kenlay, and C. M. Deane. Inverse folding for antibody sequence design using deep learning. In *The 2023 ICML Workshop on Computational Biology*, Honolulu, Hawai'i, USA, 2023. Exscientia, Oxford Science Park, Oxford, OX4 4GE, UK. URL https://icml-compbio.github.io/2023/papers/WCBICML2023_paper61.pdf. Equal contribution: Frédéric A. Dreyer and Daniel Cutting.

J. Dunbar and C. M. Deane. ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics*, 32(2):298–300, jan 2016. doi: 10.1093/bioinformatics/btv552. URL http://dx.doi.org/10.1093/bioinformatics/btv552.

J. Dunbar, K. Krawczyk, J. Leem, T. Baker, A. Fuchs, G. Georges, J. Shi, and C. M. Deane. SAbDab: the structural antibody database. *Nucleic Acids Res*, 42(Database issue):D1140–6, jan 2014. doi: 10.1093/nar/gkt1043. URL http://dx.doi.org/10.1093/nar/gkt1043.

R. Evans, M. O'Neill, A. Pritzel, N. Antropova, A. Senior, et al. Protein complex prediction with AlphaFold-Multimer, Mar. 2022. URL https://www.biorxiv.org/content/10.1101/2021.10.04.463034v2. Pages: 2021.10.04.463034 Section: New Results.

S. Fekete, A.-L. Gassner, S. Rudaz, J. Schappler, and D. Guillarme. Analytical strategies for the characterization of therapeutic monoclonal antibodies. *TrAC Trends in Analytical Chemistry*, 42:74–83, jan 2013. ISSN 01659936. doi: 10.1016/j.trac.2012.09.012. URL https://linkinghub.elsevier.com/retrieve/pii/S0165993612002841.

M. L. Fernández-Quintero, J. R. Loeffler, J. Kraml, U. Kahler, A. S. Kamenik, and K. R. Liedl. Characterizing the diversity of the CDR-h3 loop conformational ensembles in relationship to antibody binding properties. *Front Immunol*, 9:3065, 2018. ISSN 1664-3224. doi: 10.3389/fimmu.2018.03065. URL https://www.frontiersin.org/article/10.3389/fimmu.2018.03065/full.

C. Frank, A. Khoshouei, L. Fuss, D. Schiwietz, D. Putz, et al. Scalable protein design using optimization in a relaxed sequence space. *Science*, 386(6720):439–445, 2024. doi: 10.1126/science.adq1741. URL https://www.science.org/doi/abs/10.1126/science.adq1741.

P. C. Fridy, M. P. Rout, and N. E. Ketaren. Nanobodies: From high-throughput identification to therapeutic development. *Mol Cell Proteomics*, 23(12):100865, dec 2024. doi: 10.1016/j.mcpro.2024.100865. URL http://dx.doi.org/10.1016/j.mcpro.2024.100865.

G. L. Gordon, J. Gervasio, C. Souders, and C. M. Deane. The therapeutic nanobody profiler: characterising and predicting nanobody developability to improve therapeutic design. *bioRxiv*, 2025. doi: 10.1101/2025.08.11.669635. URL https://www.biorxiv.org/content/early/2025/08/14/2025.08.11.669635.

C. A. Goverde, B. Wolf, H. Khakzad, S. Rosset, and B. E. Correia. De novo protein design by inversion of the alphafold structure prediction network. *Protein Science*, 32(6):e4653, 2023. doi: https://doi.org/10.1002/pro.4653. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.4653.

B. L. Hie, V. R. Shanker, D. Xu, T. U. J. Bruun, P. A. Weidenbacher, S. Tang, W. Wu, J. E. Pak, and P. S. Kim. Efficient evolution of human antibodies from general protein language models. *Nature Biotechnology*, 42(2):275–283, 2 2024. ISSN 1546-1696. doi: 10.1038/s41587-023-01763-2. URL https://doi.org/10.1038/s41587-023-01763-2.

F. N. Hitawala and J. J. Gray. What has AlphaFold3 learned about antibody and nanobody docking, and what remains unsolved? *BioRxiv*, sep 2024. doi: 10.1101/2024.09.21.614257. URL http://biorxiv.org/lookup/doi/10.1101/2024.09.21.614257.

M. A. Jendrusch, A. L. J. Yang, E. Cacace, J. Bobonis, C. G. P. Voogdt, S. Kaspar, K. Schweimer, C. Perez-Borrajero, K. Lapouge, J. Scheurich, K. Remans, J. Hennig, A. Typas, J. O. Korbel, and S. K. Sadiq. Alphadesign: a de novo protein design framework based on alphafold. *Molecular Systems Biology*, 21(9):1166–1189, 2025. doi: https://doi.org/10.1038/s44320-025-00119-z. URL https://www.embopress.org/doi/abs/10.1038/s44320-025-00119-z.

J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, and et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021. ISSN 0028-0836. doi: 10.1038/s41586-021-03819-2.

V. Kunik, B. Peters, and Y. Ofran. Structural consensus among antibodies defines the antigen binding site. *PLoS Computational Biology*, 8(2):e1002388, 2012. doi: 10.1371/journal.pcbi.1002388. URL https://doi.org/10.1371/journal.pcbi.1002388.

M.-P. Lefranc, C. Pommié, M. Ruiz, V. Giudicelli, E. Foulquier, L. Truong, V. Thouvenin-Contet, and G. Lefranc. IMGT unique numbering for immunoglobulin and t cell receptor variable domains and ig superfamily v-like domains. *Dev Comp Immunol*, 27(1):55–77, jan 2003. doi: 10.1016/S0145-{305X}(02)00039-3. URL http://dx.doi.org/10.1016/S0145-{305X}(02)00039-3.

T. Li, M. B. Tracka, S. Uddin, J. Casas-Finet, D. J. Jacobs, and D. R. Livesay. Rigidity emerges during antibody evolution in three distinct antibody systems: Evidence from QSFR analysis of fab fragments. *PLoS Comput Biol*, 11(7):e1004327, jul 2015. doi: 10.1371/journal.pcbi.1004327. URL http://dx.doi.org/10.1371/journal.pcbi.1004327.

Nabla Bio and S. Biswas. De novo design of epitope-specific antibodies against soluble and multipass membrane proteins with high specificity, developability, and function, Jan. 2025. URL http://biorxiv.org/lookup/doi/10.1101/2025.01.21.633066.

T. H. Olsen, F. Boyles, and C. M. Deane. Observed antibody space: A diverse database of cleaned, annotated, and translated unpaired and paired antibody sequences. *Protein Science*, 31(1):141–146, 2022a. doi: https://doi.org/10.1002/pro.4205. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/pro.4205.

T. H. Olsen, I. H. Moal, and C. M. Deane. AbLang: An antibody language model for completing antibody sequences, Jan. 2022b. URL https://www.biorxiv.org/content/10.1101/2022.01.20.477061v1. Pages: 2022.01.20.477061 Section: New Results.

M. Pacesa, L. Nickel, C. Schellhaas, J. Schmidt, E. Pyatova, L. Kissling, P. Barendse, J. Choudhury, S. Kapoor, A. Alcaraz-Serna, Y. Cho, K. H. Ghamary, L. Vinué, B. J. Yachnin, A. M. Wollacott, S. Buckley, A. H. Westphal, S. Lindhoud, S. Georgeon, C. A. Goverde, G. N. Hatzopoulos, P. Gönczy, Y. D. Muller, G. Schwank, D. C. Swarts, A. J. Vecchio, B. L. Schneider, S. Ovchinnikov, and B. E. Correia. One-shot design of functional protein binders with BindCraft. *Nature*, aug 2025. ISSN 0028-0836. doi: 10.1038/s41586-025-09429-6. URL https://www.nature.com/articles/s41586-025-09429-6.

J. Prigent, A. Jarossay, C. Planchais, C. Eden, J. Dufloo, A. Kök, V. Lorin, O. Vratskikh, T. Couderc, T. Bruel, O. Schwartz, M. S. Seaman, O. Ohlenschläger, J. D. Dimitrov, and H. Mouquet. Conformational plasticity in broadly neutralizing HIV-1 antibodies triggers polyreactivity. *Cell Rep*, 23(9):2568–2581, may 2018. doi: 10.1016/j.celrep.2018.04.101. URL http://dx.doi.org/10.1016/j.celrep.2018.04.101.

D. Prihoda, J. Maamary, A. Waight, V. Juan, L. Fayadat-Dilman, D. Svozil, and D. A. Bitton. Biophi: A platform for antibody design, humanization, and humanness evaluation based on natural antibody repertoires and deep learning. *mAbs*, 14(1):2020203, 2022. doi: 10.1080/19420862.2021.2020203. URL https://doi.org/10.1080/19420862.2021.2020203. PMID: 35133949.

J. A. Ruffolo, L.-S. Chu, S. P. Mahajan, and J. J. Gray. Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies. *Nature Communications*, 14(1), 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-38063-x.

O. Sener and V. Koltun. Multi-Task Learning as Multi-Objective Optimization, Jan. 2019. URL http://arxiv.org/abs/1810.04650. arXiv:1810.04650 [cs].

A. Shanehsazzadeh, J. Alverio, G. Kasun, S. Levine, J. A. Khan, C. Chung, N. Diaz, B. K. Luton, Y. Tarter, C. McCloskey, K. B. Bateman, H. Carter, D. Chapman, R. Consbruck, A. Jaeger, C. Kohnert, G. Kopec-Belliveau, J. M. Sutton, Z. Guo, G. Canales, K. Ejan, E. Marsh, A. Ruelos, R. Ripley, B. Stoddard, R. Caguiat, K. Chapman, M. Saunders, J. Sharp, D. G. da Silva, A. Feltner, J. Ripley, M. E. Bryant, D. Castillo, J. Meier, C. M. Stegmann, K. Moran, C. Lemke, S. Abdulhaqq, L. R. Klug, S. Bachas, and A. Corporation. In vitro validated antibody design against multiple therapeutic antigens using generative inverse folding. *BioRxiv*, dec 2023. doi: 10.1101/2023.12.08.570889. URL http://biorxiv.org/lookup/doi/10.1101/2023.12.08.570889.

V. R. Shanker, T. U. J. Bruun, B. L. Hie, and P. S. Kim. Unsupervised evolution of protein and antibody complexes with a structure-informed language model. *Science*, 385(6704):46–53, 2024. doi: 10.1126/science.adk8946. URL https://www.science.org/doi/abs/10.1126/science.adk8946.

R. W. Shuai, J. A. Ruffolo, and J. J. Gray. IgLM: Infilling language modeling for antibody sequence design. *Cell Systems*, 14(11):979–989.e4, Nov. 2023. ISSN 24054712. doi: 10.1016/j.cels.2023.10.001. URL https://linkinghub.elsevier.com/retrieve/pii/S2405471223002715.

E. Trotta, P. H. Bessette, S. L. Silveria, L. K. Ely, K. M. Jude, D. T. Le, C. R. Holst, A. Coyle, M. Potempa, L. L. Lanier, K. C. Garcia, N. K. Crellin, I. J. Rondon, and J. A. Bluestone. A human anti-IL-2 antibody that potentiates regulatory t cells by a structure-based mechanism. *Nat Med*, 24(7):1005–1014, jul 2018. doi: 10.1038/s41591-018-0070-2. URL http://dx.doi.org/10.1038/s41591-018-0070-2.

Y. Tsuchiya and K. Mizuguchi. The diversity of h3 loops determines the antigen-binding tendencies of antibody CDR loops. *Protein Sci*, 25(4):815–825, apr 2016. doi: 10.1002/pro.2874. URL http://dx.doi.org/10.1002/pro.2874.

C. Vincke, R. Loris, D. Saerens, S. Martinez-Rodriguez, S. Muyldermans, and K. Conrath. General strategy to humanize a camelid single-domain antibody and identification of a universal humanized nanobody scaffold. *J Biol Chem*, 284(5):3273–3284, jan 2009. doi: 10.1074/jbc.M806889200. URL http://dx.doi.org/10.1074/jbc.M806889200.

J. L. Watson, D. Juergens, N. R. Bennett, B. L. Trippe, J. Yim, H. E. Eisenach, W. Ahern, A. J. Borst, R. J. Ragotte, L. F. Milles, and et al. De novo design of protein structure and function with RFdiffusion. *Nature*, 620(7976): 1089–1100, 2023. ISSN 0028-0836. doi: 10.1038/s41586-023-06415-8.

B. I. M. Wicky, L. F. Milles, A. Courbet, R. J. Ragotte, J. Dauparas, E. Kinfu, S. Tipps, R. D. Kibler, M. Baek, F. DiMaio, X. Li, L. Carter, A. Kang, H. Nguyen, A. K. Bera, and D. Baker. Hallucinating symmetric protein assemblies. *Science*, 378(6615):56–61, 2022. doi: 10.1126/science.add1964. URL https://www.science.org/doi/abs/10.1126/science.add1964.

T. Widatalla, R. Rafailov, and B. Hie. Aligning protein generative models with experimental fitness via direct preference optimization. *bioRxiv*, pages 2024–05, 2024.

P. C. Wilson and S. F. Andrews. Tools to therapeutically harness the human antibody response. *Nat Rev Immunol*, 12(10):709–719, oct 2012. doi: 10.1038/nri3285. URL http://dx.doi.org/10.1038/nri3285.

T. Yu, S. Kumar, A. Gupta, S. Levine, K. Hausman, and C. Finn. Gradient Surgery for Multi-Task Learning, Dec. 2020. URL http://arxiv.org/abs/2001.06782. arXiv:2001.06782 [cs].

V. Zambaldi, D. La, A. E. Chu, H. Patani, A. E. Danson, et al. De novo design of high-affinity protein binders with AlphaProteo, Sept. 2024. URL http://arxiv.org/abs/2409.08022. arXiv:2409.08022 [q-bio].

F. Zhang, H. Wei, X. Wang, Y. Bai, P. Wang, J. Wu, X. Jiang, Y. Wang, H. Cai, T. Xu, and A. Zhou. Structural basis of a novel PD-l1 nanobody for immune checkpoint blockade. *Cell Discov*, 3:17004, mar 2017. doi: 10.1038/celldisc.2017.4. URL http://dx.doi.org/10.1038/celldisc.2017.4.

A. Methods

A.1. Germinal algorithm

Germinal performs gradient-based optimization on the joint landscape defined by AF-M and IgLM objectives, merging their gradients into a single update direction. Following previous work (Pacesa et al., 2025), design proceeds through three phases (logit, softmax, and semi-greedy optimization) that progressively anneal continuous logits into one-hot sequence representations (**Figure 2A**). Non-interface CDR residues are then redesigned during sequence optimization using an antibody-finetuned sequence design model. Finally, designs undergo filtering and ranking to select top candidates for experimental testing.

A.1.1. Input preparation

To run the design pipeline, we first defined both the target epitope and the structure of the antigen to be used. To reduce computational cost, the antigen structure was restricted to independently folding domains whenever possible. For PD-L1, we used the V-like domain (residues 19–127), while for IL3, IL20, and BHRF1, we used the full mature protein sequence excluding signal peptides. Antigens were predicted using AF3 (Abramson et al., 2024) and validated to fold with high confidence and accuracy compared to the experimental structure (RMSD < 1 Å; **Figure 1C**), if available. This procedure was also applied to new antibody frameworks.

Antibody PDB filenames were standardized to encode the lengths of each CDR (IMGT (Lefranc et al., 2003) numbering, identified using AbNumber (Dunbar and Deane, 2016)) separated by underscores, followed by the binder type ("nb" or "scFv"). These CDR lengths were provided to the pipeline using command line arguments or through an input settings file. These lengths, along with user-defined start points, were used to calculate CDR positions. For antigens with known binders (e.g., natural receptors), we targeted the same epitope as the reference, biologically relevant interaction. Epitope positions were specified using a command line argument or through an input settings file.

A.1.2. Initialization

Before the design stage, the antibody framework and antigen structures are merged into a unified PDB containing two chains, which serves as the starting complex. The position-specific scoring matrix (PSSM) is initialized by setting framework positions to their one-hot encodings and CDR positions to one of three different initializations depending on user choice: (i) uniform random noise, (ii) Gumbel noise centered around the template framework sequence, or (iii) zeros. This flexibility allows users to control the trade-off between broad sampling and guided starting points. Gumbel initialization provided the most stable optimization and was adopted as the default.

During AF-M initialization, the antigen and antibody framework structures are supplied as templates to AF-M. Importantly, only the framework regions (FRs) are used as templates, with all CDR residues masked in the input feature matrices (i.e., atom positions, side chains, sequence). This ensures that AF-M provides structural guidance based on stable framework features without biasing the optimization of CDR loops to the natural CDRs of the framework structure.

A.1.3. Sequence bias

To bias sequences toward a chosen framework, we use a bias weight to modulate between framework preservation and exploration. This term is applied to framework region (FR) residues only (i.e., all residues other than CDRs). Specifically, at every iteration, this bias is added to the logits corresponding to the original framework sequence before the current binder is used as input to AF-M and IgLM, essentially shifting the probability distribution to favor the original framework sequence. Unlike hard-fixing, this approach still permits mutations in FR positions if gradients strongly favor them. Based on parameter sweeps (**Figure S13**), we selected a value of 10, which minimized framework mutations while allowing enough flexibility for confidence optimization, and hence successful trajectories.

A.1.4. Design stages

- **Logits**: In this stage, the sequence is represented as continuous logits for each position and amino acid type. Gradient-based optimization is performed directly on these logits using the combined gradients from AF-M (structural confidence) and IgLM (sequence naturalness). The logits allow for smooth optimization in continuous space while maintaining differentiability.
- **Softmax**: In this stage, the sequence logits of the binder are normalized to sequence probabilities using the softmax function, and gradients are similarly applied to update this representation of the sequence.
- **Semi-greedy optimization**: In this final stage, the softmax probabilities from the previous stage are used to sample mutations, and mutations with best loss are fixed.

A.1.5. Optimization strategy

Gradients from AF-M and IgLM are merged at each iteration of the logit and softmax phase using either a weighted sum, Multiple Gradient Descent Algorithm (MGDA) (Sener and Koltun, 2019) or a weighted version of PCGrad (Yu et al., 2020) (see ***Gradient merging***). In the semi-greedy phase, optimization shifts from continuous gradients to discrete sequence updates. At each of the 10 semi-greedy iterations, five candidate sequences are sampled, and the one that best balances AF-M structural confidence with IgLM sequence likelihood—calculated as a weighted sum of the AF-M loss and IgLM log-likelihood—is accepted. Notably, for all three stages, the IgLM weight is adjusted according to a four-step schedule. During the logits phase, the IgLM weight is linearly annealed from v_1 to v_2. In the softmax phase, the weight is set to v_3. Finally, in the semi-greedy phase, the weight is set to v_4. These values are supplied as a command line argument or through an input settings file. This annealing schedule empirically provided more robust performance than fixed IgLM weights, and is the default setting in Germinal.

Additionally, because Germinal does not include a discrete one-hot optimization phase, we check for sequence convergence at the end of the softmax phase. Specifically, for each CDR position we compute the maximum probability across all the 20 amino acids, and then average these values across all CDR positions, representing the average positional certainty in the most likely amino acid. To ensure that the model is confident enough in the sequence to justify discretization, only sequences exceeding an empirically determined threshold (0.1) for this metric are advanced to the semi-greedy stage.

At the end of each phase, the current best design is evaluated against AF-M confidence metrics (pLDDT $>$ 0.8, ipTM $>$ 0.68 and iPAE $<$ 0.35) and designs advance to the next phase only if they satisfy these thresholds. At the end of the design stage, remaining sequences are co-folded with AF3 for independent evaluation prior to ***Sequence optimization***. Designs must satisfy strict AF3 confidence thresholds (ipTM $>$ 0.75, PAE $<$ 7, pTM $>$ 0.85, and pLDDT $>$ 0.85), exhibit a CDR interface percentage greater than 60%, and be predicted to bind the specified epitope to proceed.

Importantly, the AF-M loss used in Germinal is computed as a weighted sum of multiple structural and interface metrics. A full list of the loss terms and their weightings, as well as biases added during optimization are shown below (see ***Loss and bias terms***). Additionally, during each iteration, AF-M is run using three recycles, with the final recycle taken as the prediction for each iteration.

A.1.6. Gradient merging

Central to Germinal is the gradient merging algorithm, which reconciles updates from AF-M and IgLM objectives. Before merging is applied, gradients from both models are normalized to their respective ℓ_2-norms, and scaled by the effective sequence length (the number of non-zero positions in the AF-M gradient). This ensures optimization of the two objectives is not biased due to magnitude differences between the two gradients. Three strategies were explored used for gradient merging:

- **Weighted Sum**: gradients are merged by a direct weighted sum where the AF-M weight is fixed to 1, and the IgLM weight λ is set by a user-defined parameter.

$$\nabla\text{Germinal} \;=\; \nabla\text{AF-M} \;+\; \lambda \cdot \nabla\text{IgLM}$$

- **Weighted PCGrad**: As described in Yu et al. (2020), PCGrad mitigates destructive interference between gradients by resolving conflicts in direction. Specifically, one gradient is projected onto the orthogonal

complement of the other if the dot product is negative, thereby preventing cancellation of progress along critical dimensions. In our implementation, the previously normalized vectors are scaled by a user-defined weight.

- **Multiple Gradient Descent Algorithm**: MGDA (Sener and Koltun, 2019) formulates gradient merging as a multi-objective optimization problem, where the goal is to compute a Pareto-optimal descent direction. Concretely, MGDA solves for the convex combination of AF-M and IgLM gradients that minimizes the squared norm of the merged update

$$\nabla\text{Germinal} \;=\; \alpha \cdot \nabla\text{AFM} \;+\; (1-a) \cdot \nabla\text{IgLM}$$

with $\alpha \in [0, 1]$ chosen to balance objectives without privileging one a priori. This is achieved by solving a small quadratic program at each step. In practice, MGDA is more computationally intensive than PCGrad or weighted sums, but it provides a principled way to balance objectives when their contributions vary dynamically during optimization. After computing the MGDA update, the resulting gradient is re-normalized to have unit ℓ_2-norm and scaled by the effective sequence length, as described above.

A.1.7. Loss and bias terms

Below, we list all biases, weights, and custom loss terms utilized in Germinal.

We use the following biases:

- **Sequence bias:** Higher values enforce greater similarity to the framework sequence.
- **IgLM scale:** Scaling factor applied to normalized IgLM gradients relative to AF-M gradients.

We use the following ColabDesign loss functions (weights):

- **pLDDT loss:** Emphasizes high per-residue confidence in AF-M predictions.
- **ipLDDT loss:** Emphasizes high per-residue confidence at the predicted interface.
- **iPAE loss:** Penalizes high alignment error between predicted and target structures at the interface.
- **PAE loss:** Penalizes high alignment error of the predicted complex.
- **ipTM loss:** Favors predictions with high predicted TM-score, encouraging correct geometry at the interface.
- **pTM loss:** Favors predictions with high predicted TM-score, encouraging correct overall geometry.
- **Inter-protein contacts:** Ensures that the designed antibody has enough contacts to the antigen chain within a certain distance. The number of contacts and distance can be specified by the user. If a hotspot is provided, the loss favors contacts with the specified positions.
- **Intra-protein contacts:** Encourages each residue in the designed sequence to form a sufficient number of contacts with other residues within the same protein. The user can specify both the minimum number of contacts required and the distance threshold used to define a contact.

We use the following custom loss functions:

- **α-helical loss:** This loss is an inverted variant of the helix loss from Pacesa et al. (2025), restricted to CDR positions. We discourage α-helical geometry within the CDRs by penalizing $\{i, i+3\}$ pairs whose predicted distances fall in the 2-6.2 Å range characteristic of α-helices. Using AlphaFold's distogram, we compute a binary cross-entropy penalty per residue pair that is large when probability mass lies inside this window and small otherwise. We also apply a positional mask to restrict evaluation to CDR residues, averaging the resulting penalties. Minimizing this loss reduces helix-like contacts in the CDRs and biases designs toward flexible, loop-like conformations.
- **β-strand loss:** Analogous to the helix term, this loss suppresses β-strand–like geometry within the selected CDR positions by penalizing $\{i, i+3\}$ pairs whose predicted distances fall in a β-like window (9.75–11.5 Å). Similar to the helix loss, we compute a binary cross-entropy penalty per pair with respect to this window using the distogram. To restrict this loss to CDR residues, we apply a positional mask that includes all CDR including the position directly before and directly after each CDR. We sort the probability of each pair constituting a β-strand and the loss from the top three most likely pairs is averaged as the final β-strand loss. Minimizing this loss reduces β-strands in the targeted regions, complementing the α-helix loss to favor loop-like conformations.

- **Paratope loss:** We develop a paratope loss to achieve two goals: (1) encourage the CDR residue to bind to the specified epitope and (2) discourage the framework to bind to the target. For the first goal, we define a loss $\mathcal{L}_{\text{CDR}}$, which operates on two sets of residues: `target_id`, which includes all specified hotspot residues on the target, and `pos_id`, which include all CDR residues on the antibody. For each pair of residues between `target_id` and `pos_id`, we calculate a binary cross-entropy (BCE) loss using the distogram distance probabilities against a user-defined distance threshold. We then average the top k closest predicted contacts for each epitope residue, across all residues in `target_id` to yield a final scalar loss $\mathcal{L}_{\text{epitope}}$. For the second goal, we calculate the same loss with different sets of residues. Here, we set `target_id` to all target residues and define `fw_id` as the set of framework residues on the antibody. Following the same loss procedure yields $\mathcal{L}_{\text{framework}}$ (**Figure 2B**). We then define the final paratope loss as

$$\mathcal{L}_{\text{paratope}} = \frac{\mathcal{L}_{\text{CDR}}}{\mathcal{L}_{\text{framework}} - \lambda}$$

  where λ is a user-defined weight and is referred to as an "offset" value in the codebase. Note that both $\mathcal{L}_{\text{CDR}}$ and $\mathcal{L}_{\text{framework}}$ will decrease when the corresponding region gains more contacts with its target. Thus, the combined term $\mathcal{L}_{\text{paratope}}$, which is proportional to $\mathcal{L}_{\text{CDR}}$ and inversely proportional to $\mathcal{L}_{\text{framework}}$, increases when framework contacts are abundant and decreases when CDR residues are predicted to bind to the specified epitope, steering contacts toward the intended CDR pose via minimization of the loss.

A.1.8. Sequence optimization

Design candidates that passed initial filters are passed through AbMPNN (Dreyer et al., 2023; Dauparas et al., 2022), a variant of ProteinMPNN fine-tuned on OAS (Olsen et al., 2022a) and SAbDab (Dunbar et al., 2014). AbMPNN was used to generate 40 sequence variants per design, of which the top 4 were retained for re-folding with AF3 and final filtering. To enhance stability while ensuring that residues making key paratope–epitope contacts remain unchanged, optimization was restricted to CDR residues with all heavy atoms > 3 Å from the antigen.

A.1.9. IgLM Implementation

Slight modifications were made to the original implementation of IgLM (Shuai et al., 2023) in order to incorporate the PSSM sequence representation during design. All inputs to IgLM are one-hot amino acid sequences; at each iteration, we convert the raw PSSM to a probability distribution using a softmax operation with a temperature of 0.6, then use a straight-through estimator to obtain discrete sequence representations. We condition each sequence with a `<HEAVY>` chain token and `<CAMEL>` species token, given a camelid framework was used for nanobody designs (Vincke et al., 2009). Loss is calculated by taking the mean negative log-likelihood over the entire sequence, which is then backpropagated to the raw PSSM to yield the update gradient for the iteration.

A.1.10. Design filtering

For filtering, designs were evaluated against a set of stringent thresholds derived from structural prediction confidence scores and PyRosetta-derived metrics (Chaudhury et al., 2010). Only sequences passing all thresholds advanced to the second stage. Seeking to utilize an independent structure prediction model from AF-M and given its superior accuracy in modeling antibody-antigen complexes (Hitawala and Gray, 2024), we selected AlphaFold 3 (AF3) (Abramson et al., 2024) as the final co-folding model. All AF3 predictions were performed without structural templates and only using MSAs for the target. The metrics and thresholds applied for filtering are detailed in **Table S1**.

A.1.11. Design ranking

After filtering, we proceed to an optional, yet recommended, ranking stage, to reduce the number of designs to be experimentally validated. Ranking is conducted using a Borda count–based ranking algorithm, which integrates multiple structural and sequence-based metrics, each weighted by relative importance. Remaining designs were scored across metrics detailed in **Table S2**.

The combination of hard thresholds followed by Borda ranking ensured that only high-confidence, well-structured, and framework-preserving candidates were advanced. Retrospective analysis further highlighted the importance of several metrics (**Figure S7**).

A.1.12. Computational resources

All sampling was performed on NVIDIA H100 GPUs. Although exact runtimes vary heavily by antigen, typical sampling runs require 200–500 H100-GPU hours to yield 200–400 designs that pass pipeline filters.

A.2. Experimental characterization

A.2.1. Framework selection

To identify a robust scaffold for CDR grafting, we evaluated both nanobody and scFv frameworks for their expression and tolerance to CDR replacement. For nanobodies, we screened three frameworks: the humanized nanobody hNbBCII-FGLA, its camelid predecessor cAbBCII, and an alpaca nanobody that has been used in prior in-house experiments. These scaffolds were chosen due to their known stability and successful use as universal grafting frameworks (Vincke et al., 2009). For scFvs, we tested an anti-IL2 scFv from our collection (FAB-F5111) (Trotta et al., 2018), an scFv version of trastuzumab (Byun et al., 2023), and a previously reported scFv framework (FW1.4) (Borras et al., 2010). All nanobody scaffolds were grafted with PD-L1-nb (Zhang et al., 2017) CDRs (PDB: 5JDS), and all scFvs were grafted with F5111 CDRs to assess the effect of framework on binding and expression activities. Among the nanobody grafts, only hNbBCII and cAbBCII expressed robustly and retained binding activities (**Figure S1**). We selected hNbBCII for downstream work based on its prior successful use (Bennett et al., 2024). Among the scFv grafts, only the trastuzumab scaffold yielded appreciable expression, but none of the scFv grafts showed a significant binding signal. We retained trastuzumab given its previous successful use in other computational antibody design pipelines (Bennett et al., 2024).

A.2.2. Plasmids and cloning

Constructs were cloned by PCR methods using either NEBNext High-Fidelity Polymerase (New England Biolabs; #M0544) followed by Gibson assembly of fragments (New England Biolabs; #E2621S), or Phusion Flash High-Fidelity Master Mix (Thermo Fisher; #F548L) followed by In-Fusion fragment assembly (Takara Bio). In all cases, constructs were assembled in a three-piece In-Fusion strategy, in which the mammalian expression vector backbone was split into two fragments by PCR with pUC-ori primers, and homology regions were designed to overlap with the synthetic DNA fragments obtained from Twist Biosciences. DNA sequences of designed binders and targets (PD-L1-d1, PD-L1-d1d2, IL3, IL20, and BHRF1 (**Figure 1C**) were ordered from Twist Biosciences with Gibson cloning adaptors for insertion into a pCMV mammalian expression vector. Target proteins were cloned with their native N-terminal signal peptide, with the exception of BHRF1 which was cloned with the IL2 signal peptide (MYRMQLLSCIALSLALVTN). All designed binders were cloned with the IL2 signal peptide. For split-luciferase assays, all target proteins were fused to a C-terminal SmBit peptide and His tag, while all designed binders were fused to a C-terminal LgBit protein. For BLI experiments, target proteins were cloned with a C-terminal AviTag and His tag, except for IL3, which was cloned in both C-terminal AviTag and N-terminal AviTag formats due to the proximity of its termini to the target epitope. Designed binders were fused to a C-terminal His tag only. Plasmids were propagated in Turbo Competent Cells (New England Biolabs) for high-yield transformation and outgrowth, and purified using standard plasmid extraction protocols. All constructs were validated by whole plasmid Nanopore sequencing. A full list of tested binder sequences is provided in **Data S2**.

A.2.3. Expi293 protein expression

All target proteins for split-luciferase assays and lead designed binders were expressed in Expi293F cells (Thermo Fisher; #A14635). For BLI experiments, target proteins possessing an AviTag were expressed in Expi293F cells which constitutively express the BirA enzyme, enabling biotinylation of targets in vivo. Cells were maintained in a 2:1 mixture of FreeStyle 293 Expression Medium (Thermo Fisher; #12338018):Expi293 Expression Medium (Thermo Fisher; #A1435102) and grown at 37 °C with 8% (v/v) CO2 and shaking at 125 rpm. After ≥ 3 passages, cells were diluted to 2 × 106 cells/mL and grown for 20-24 h, when they were diluted to 3 × 106 cells/mL for transfection in a final volume of 30-50 mL (designed binders) or 100-250 mL (targets).

Transfection mixtures were prepared with the following components per mL of expression culture: 0.5 μg midi-prepped DNA, 1.3 μL FectoPRO transfection reagent (Polyplus; #101000007), and 100 μL expression media. After incubating for 10 min, the transfection cocktails were added to expression cultures. Cultures were immediately supplemented with glucose (0.45% (w/v) final concentration) and valproic acid (2 mM final concentration), and grown for 4-5 days post-transfection, when cell viability fell below 50%. Secreted protein was harvested from cells by centrifugation at 4000 × g at 4 °C, followed by filtration through a 0.22 μM syringe filter (Thermo Fisher; 723-2520). 10× phosphate buffered saline (PBS; 1.37 M NaCl, 27 mM KCl, 100 mM Na2HPO4, 18 mM KH2PO4, pH 7.4) and 1.63 M NaCl were added to filtered supernatant each to a final dilution of 1:10 and, if necessary, supernatants were pH-adjusted to pH 7.6 using 1 M Tris-HCl pH 9.0.

A.2.4. Ni-NTA purification of His-tagged proteins

All His-tagged proteins were purified using HisPur Cobalt resin (Thermo Scientific; #89965). Cobalt resin was washed twice in binding buffer (high-salt phosphate buffered saline [HS-PBS; 300 mM NaCl, 2.7 mM KCl, 10 mM Na2HPO4, 1.8 mM KH2PO4, pH 7.4], 10 mM imidazole) before adding to clarified supernatants and incubating for 1 h at 4 °C with end-over-end rotation. Purification mixtures were transferred to chromatography columns for gravity flow purification. Resin was first washed with binding buffer, followed by wash buffer (HS-PBS, 30 mM imidazole), before proteins were eluted with elution buffer (HS-PBS, 200 mM imidazole). Eluants were immediately buffer exchanged into PBS pH 7.4 and concentrated using centrifugal concentrators with 5 or 10 kDa molecular-weight cut-off (Millipore; #UFC901024), by four rounds of centrifugation (4,000-15,000 × g, 10 min, 4 °C) followed by resuspension in PBS pH 7.4. Protein concentrations were determined from A280 using their extinction coefficients predicted from ExPASy ProtParam.

A.2.5. HEK293 protein expression

Transient transfections were performed on all of the binder plasmids using jetOPTIMUS® DNA transfection Reagent (Avantor) in 24-well plates, with each plate incorporating internal controls. When available, validated binders (PD-L1 binder, KN035 [PDB: 5JDS]; BHRF1 binder, GDM_BHRF1_35 [Zambaldi et al. (2024)]) were included as positive controls. Binders not associated with a given target (a PD-L1 binder, KN035, for IL20 and BHRF1; a TNFα binder [PDB: 5M2J] for PD-L1 and IL3) were included as negative controls. This layout ensured that every plate contained both experimental and control conditions for consistent data normalization and comparison. Each binder plasmid was transfected in two biological replicates, with each replicate consisting of 450 ng of binder plasmid and 5 ng of CMV-driven mCherry plasmid (as co-transfection marker). A total of 990 ng binder plasmid DNA was added to 110 μL of transfection buffer with co-transfection marker mixed. 0.88 μL of transfection reagent was then added, well mixed, and incubated for 9 minutes. Following incubation, 50 μL of the transfection complex was added dropwise to each well of Human Embryonic Kidney (HEK) 293 cells (Thermo Scientific; #R78007) at 80-90% confluency. The cells were cultured in Dulbecco's Modified Eagle Medium (DMEM), supplemented with 10% fetal bovine serum (Fisher Scientific; #FB1299910), 1 mM sodium pyruvate (EMD Millipore; #TMS-005-C), 1× penicillin–streptomycin (Genesee catalog, #25-512), 2 mM l-glutamine (Genesee; #25-509) and 1× MEM non-essential amino acids (Genesee; #25-536), under standard conditions (37 °C, 5% CO_2). 6 to 12 hours after transfection, 300 μL of DMEM was removed from each well and replaced with 200 μL of fresh DMEM media to reduce reagent effects and concentrate final protein concentration. All split-luciferase assays were carried out at least 36 hours after changing the media.

A.2.6. HiBit-LgBit NanoLuc complementation reporter assay for binder expression

Binder expression was quantified by luminescence of the Hibit-LgBit complementation system (Nano-Glo® HiBiT Lytic Detection System; Promega) measured with plate reader (Tecan Infinite® M Plex, multimode microplate reader; #30190085). 200 μL of media were first collected from each well in 24-well plates and split equally between 2 wells in 96-well plates. During all experiments, media from the first biological replicate well were distributed to the first two technical replicates of split-luciferase assay, and media from the second biological replicate added to the third technical replicate. Each split-luciferase assay reaction was prepared in a total volume of 100 μL, consisting of 49 μL Nano-Glo® buffer, 44.9 μL phosphate-buffered saline (PBS, pH 7.4), 1 μL Nano-Glo® substrate, 0.1 μL HiBit (20 nM) and 5 μL conditioned media. First luminescence reading was recorded immediately after addition of samples and second measurement was performed after 4 minutes of incubation. Longer incubation time leads to signal saturation, causing the plate reader to end the assay.

A.2.7. SmBit-LgBit NanoLuc complementation reporter assay

Binder-target interaction and specificity was quantified by luminescence of the SmBiT–LgBiT complementation system (Nano-Glo® Live Cell Assay System, Promega) on the plate reader (Tecan Infinite® M Plex, multimode microplate reader; #30190085). Media from the same well used in the Hibit-LgBit assay were used to maintain consistency. Each reaction was prepared in a total volume of 100 μL, consisting of 49 μL Nano-Glo® buffer, 39 μL phosphate-buffered saline (PBS, pH 7.4), 1 μL Nano-Glo® substrate, 20 nM target protein (from 2 μM stock solution), and 10 μL conditioned media. To assess specificity, parallel reactions were assembled with the same volume of conditioned media but substituting the target protein with 20 nM off-target protein (IL2 for PD-L1 binders; PD-L1 for BHRF1, IL20, and IL3 binders; all from 2 μM stock solutions). First luminescence was recorded right after all the binders were added and second measurement was performed after 8 minutes of incubation.

To verify that binding activity could be represented by SmBiT–LgBiT signals normalized by HiBiT–LgBiT (expression) signals, purified PD-L1 KN035 binder (K_D of ~3 nM) of different concentrations (0.01 nM, 0.1 nM, 1 nM, 10 nM, 100 nM) was tested using the same experimental setup as the split-luciferase assays used for all binders with the exception of the 5 μL of media, which were replaced by direct addition of purified protein at the specified dilutions. HiBiT–LgBiT measurements were not collected at 100 nM due to signal saturation. Both the HiBiT–LgBiT and SmBiT–LgBiT assays exhibited log-linear responses across the tested concentration range (**Figure 3B**), supporting the use of SmBiT–LgBiT normalized by HiBiT–LgBiT as a valid measure of binding activity.

A.2.8. Identification of binder leads for further characterization

For each split-luciferase assay condition, three technical replicates were averaged to obtain a representative luminescence value for both expression (HiBiT–LgBiT assay) and binding (SmBiT–LgBiT assay). In the initial screens of PD-L1 and IL3, binders were first defined based on binding activity, using a threshold of at least a five-fold increase in normalized binding signal relative to the negative control. A set of PD-L1 nanobodies with different expected activities was also tested on BLI to gain insights into both their expression and binding behaviors. Based on these initial tests, only samples with HiBiT–LgBiT luminescence values above 150,000 arbitrary luminescence units (ALU) were considered expressed. Binding activity for downstream validation was subsequently determined using more refined criteria following our initial PD-L1 experiments: we defined a binder as either (i) an increase in the binding ratio (on-target/off-target) between T1 and T2 greater than 0.5, or (ii) an increase in the on-target binding ratio (on-target T2 / on-target T1) greater than 1.05.

Additionally, we identified two nanobodies (one with a high normalized on-target/off-target binding luminescence ratio and one with a low ratio) each for PD-L1, IL20, and BHRF1 to be tested on BLI against their target antigen and an off-target antigen (**Figure S8**). We observed highly target-specific binding for each higher, regardless of its on-target/off-target normalized binding luminescence ratio (**Figure S12**). Thus, owing to the sensitivity of the split-luciferase binding assay to the target-SmBiT construct, on-target/off-target luminescence ratio alone cannot reliably serve as a filter for specificity. For PD-L1 binders, which were part of an initial assay evaluation, off-target screens were performed only on a subset of candidates; those lacking off-target measurements were classified as either non-binders in the split-luciferase assay or as not expressed.

A.2.9. Bio-layer interferometry (BLI)

Target proteins with AviTag and SmBiT removed and binder proteins with LgBiT removed, were used for BLI analysis. All reactions were run on an Octet RED96 at 30 °C, and samples were run in PBS pH 7.4 supplemented with 0.2% BSA and 0.05% Tween 20 (assay buffer). For each of the binder proteins, a titration of five concentrations was prepared by 2-fold serial dilution from the highest concentration tested. Lead nanobody designs were assayed for binding to biotinylated target proteins immobilized onto Octet Streptavidin Biosensors (Sartorius; 18-5019). Sensors were pre-incubated in an assay buffer for 10 minutes, before incubation in wells containing the assay buffer to determine the baseline signal. Sensors were then immersed in wells containing biotinylated target protein (200 nM in assay buffer) for 300 s or until a 1 nm BLI shift was reached. Loaded sensors were again washed and baselined in wells containing assay buffer alone, before immersion in nanobody protein solutions for association. Dissociation was monitored by transferring the sensors back into assay buffer alone. For each binder titration, a reference channel was prepared by loading sensors with target protein, but associating in assay buffer alone, to account for any change in signal due to dissociation of target protein

from the biosensor tips. Association and dissociation binding curves were fit in Octet System Data Analysis Software version 9.0.0.15 using a 1:1 monovalent model for IgGs to determine apparent KD, with signals baseline-corrected by subtraction of the reference channel. For BLI experiments with IL20, sensorgrams fit best to a 2:1 heterogeneous ligand model, consistent with two kinetic populations of immobilized ligand. When sensograms were fit with 1:1 binding model, systematic residuals were observed, particularly during dissociation, while residuals were minimized and random when fit to a 2:1 model. BLI hits were identified if the design displayed clear association kinetics (BLI shift > 0.05 nm for the highest tested concentration), high-quality fits ($R^2 > 0.95$). For the IL3 binders (**Figure 4D**), the fit curves failed to achieve an $R^2 > 0.95$. To address this, binding assays were repeated against versions of IL3 possessing either a C-terminal or an N-terminal AviTag (**Figure S8B, C**). The averaged results from all technical replicates were used to represent the binding kinetics.

A.2.10. Interface mutation screen

Interface mutation screening was carried out by substituting individual interface hot-spot residues with alanine (PD-L1: I54, Y56, E58). Binding affinities were measured by BLI using the same procedure as for wild-type PD-L1 and IL3, except that the target proteins were replaced with the alanine mutant variants. All BLI data were collected and analyzed using Octet System Data Analysis Software (version 9.0.0.15), with signals baseline-corrected by subtraction of the reference channel.

B. Supplementary Tables

Table S1 | **Set of confidence metrics and PyRosetta-derived scores used as pipeline filters.**

Metric	**Condition**	**Details**
Binder near the epitope	True	Intersection of predicted interface residues and epitope residues. Should be at least 3 residues.
CDR3-epitope contacts	> 1	Number of CDR3 residues that are part of the interface residues.
Clashes after relaxation	$= 0$	Rosetta metrics
Surface hydrophobicity	≤ 0.4	Rosetta metrics
Interface hydrophobicity	≥ 0.45	Rosetta metrics
pDockQ2	> 0.23	Calculated according to the author's implementation.
Interface shape complementarity	≥ 0.5	Rosetta metrics
Binder RMSD	< 6	RMSD between the final predicted structure (AF3) and the initial design output by the pipeline (AF-M).
Interface hydrogen bonds	> 4	Rosetta metrics
CDR Interface Percentage	$> 60\%$	Number of residues at the interface that are CDR residues over total interface residues.
pLDDT	> 85	AF3 metric
pTM	≥ 0.85	AF3 metric
ipTM	≥ 0.75	AF3 metric
PAE	≤ 7	AF3 metric

Table S2 | **Set of metrics used to rank designs.**

Metric	**Filter threshold / ranking weight**	**Details and use**
IgLM log-likelihood	≥ -1.85 / 0.10	IgLM's confidence in the sequence. Filter and ranking.
AF3 PAE	≤ 6	Used as a filter.
AF3 pTM	≥ 0.83	Used as a filter.
AF3 ipTM	≥ 0.75	Used as a Filter.
Binder score	≤ -390 / 0.10	PyRosetta energy score. Used as filter and a ranking criteria.
Interface shape complementarity	≥ 0.6 / 0.10	Used as a filter and ranking criteria.
Interface hydrophobicity	50–60% / 0.10	Used as a filter and ranking criteria.
Interface hydrogen bonds	≥ 4 / 0.15	Used as a filter and ranking criteria.
# of framework mutations	< 3	Used as a filter.
CDR-specific SAP score	< 36 / 0.10	Used as a filter and ranking criteria.
pDockQ2	≥ 0.45 / 0.10	Used as a filter and ranking criteria.
CDR3-epitope contacts	≥ 4 / 0.10	Used as a filter and ranking criteria.
CDR Interface Percentage	≥ 0.5	Used as a filter and ranking criteria.
Interface dG	< -35	PyRosetta energy score. Used as a filter and ranking criteria.
AbMPNN score	0.10	Negative log-likelihood of AbMPNN-designed residues (lower is better). Used as a filter.

C. Supplementary Figures

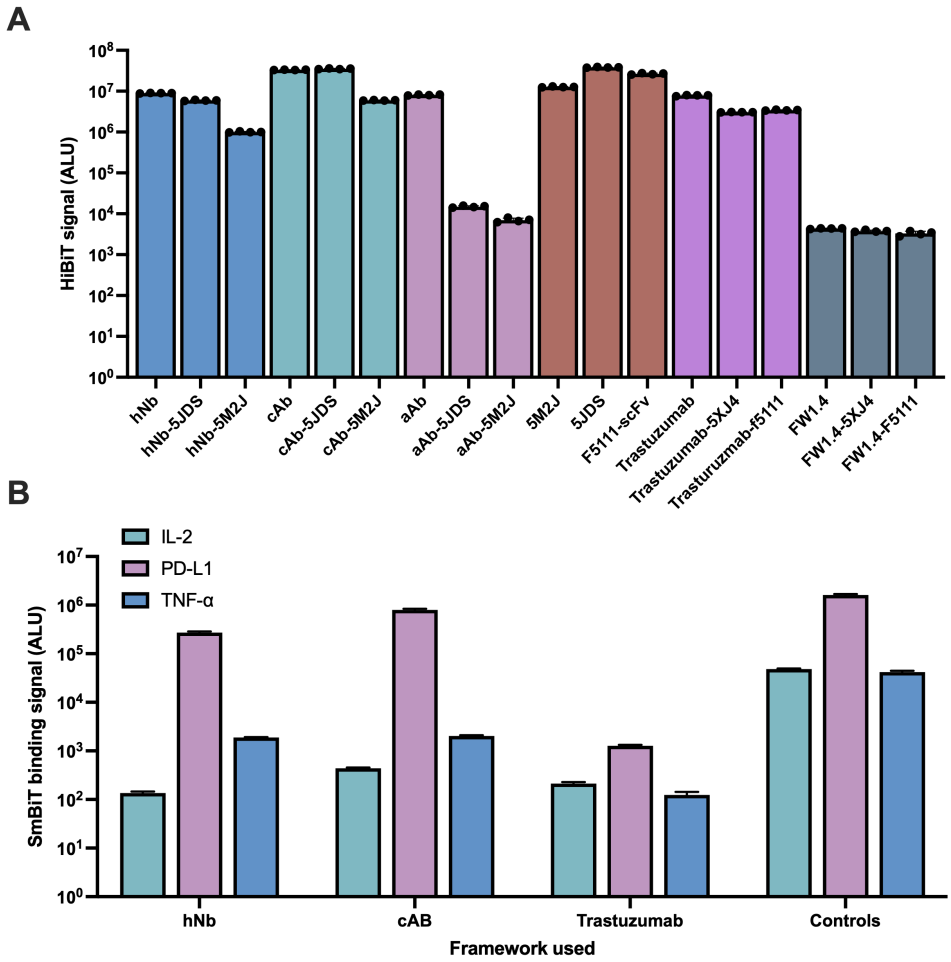

Figure S1 | **Experimental validation of different nanobody and scFv frameworks.** (**A,B**) Different combinations of engrafted CDRs were tested on different antibody frameworks to select the one with highest expression (**A**) and capacity to withstand changes to their CDRs while maintaining binding (**B**). Three nanobody frameworks: hNbBCII-FGLA (PDB: 3EAK), cAbBCII (PDB: 3DWT) and aAb), and two scFv frameworks: trastuzumab (PDB: 6ZQK) and FW1.4 were first tested for expression. For nanobodies, CDRs from a PD-L1 (PDB: 5JDS) and a TNF-α (PDB: 5M2J) nanobody were used. Similarly, for scFvs, CDRs from F5111, an IL2 FAB, and a PD-L1 scFv (PDB: 5XJ4) were used. Designs were only tested if they showed substantial expression levels ($> 10^6$). Bar plots in (**B**) show each of the two engrafted design per framework against its intended antigen, plus a negative control (i.e., hNb-5JDS exclusively against PD-L1, hNb-5M2J exclusively against TNF-α, and a negative control wild-type hNb against an unrelated antigen, IL2). For nanobodies, IL2 was used as control while for scFv TNF-α was used as negative control.

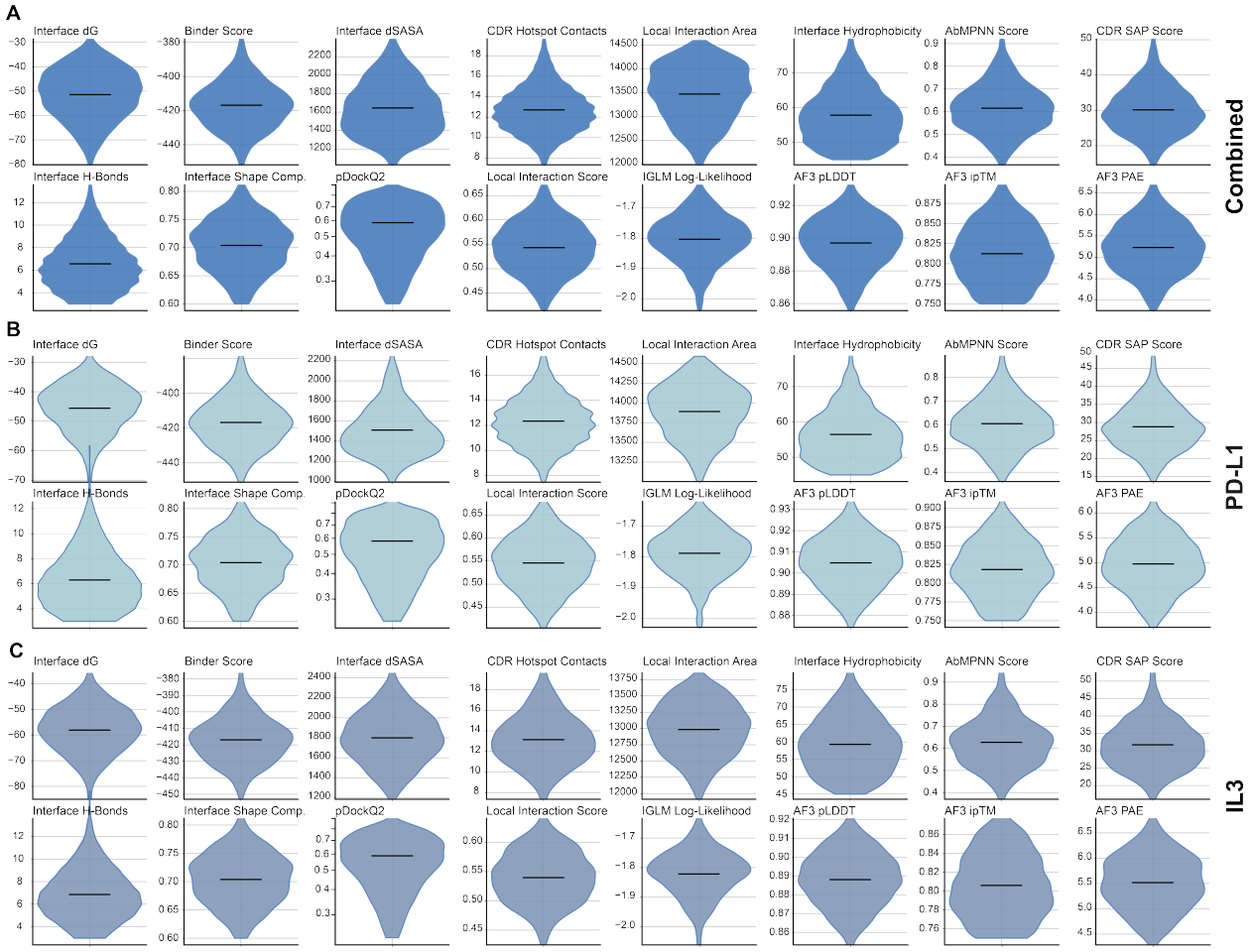

Figure S2 | **Scores distribution of key computational metrics for all passing nanobody designs.** (**A**) Distributions across all successful designs pooled from different targets. (**B**, **C**) Specific distributions for PD-L1 (**B**) and IL3 designs (**C**).

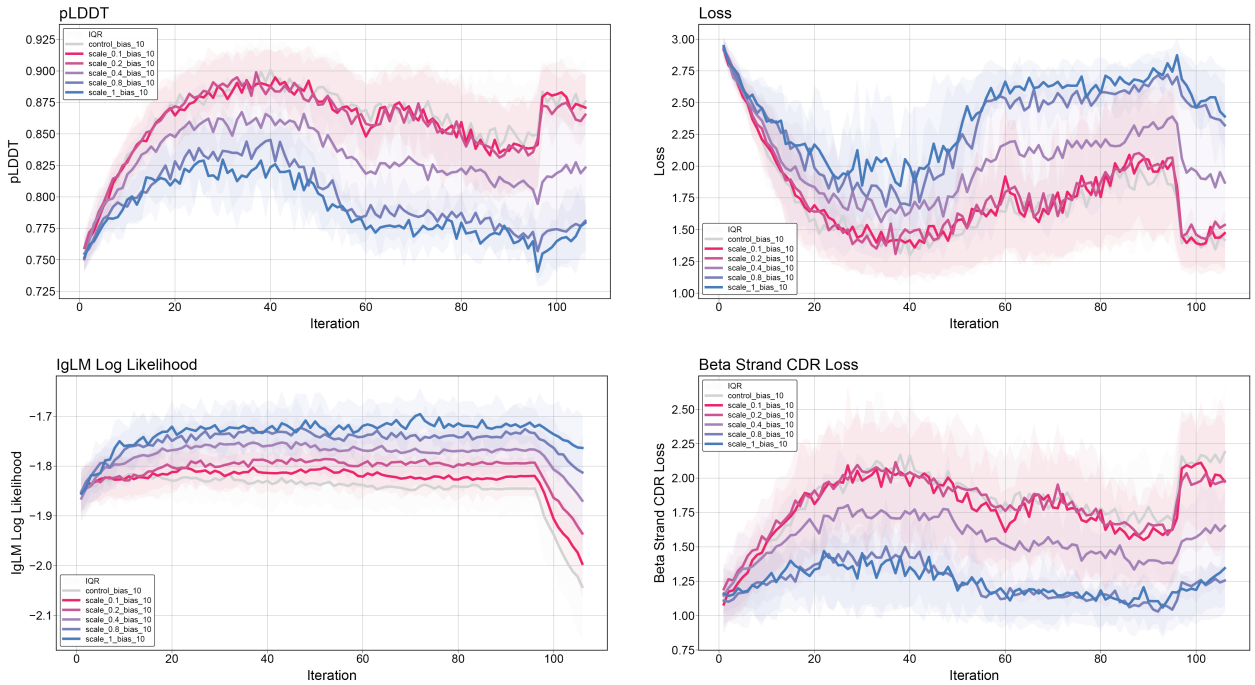

Figure S3 | **Effect of IgLM guidance over entire design trajectories.** The effect of incorporating IgLM-based guidance during the design trajectories across 110 iterations is shown, evaluated under different bias scaling factors (0.1, 0.2, 0.4, 0.8, 1.0) compared to a control without IgLM guidance.

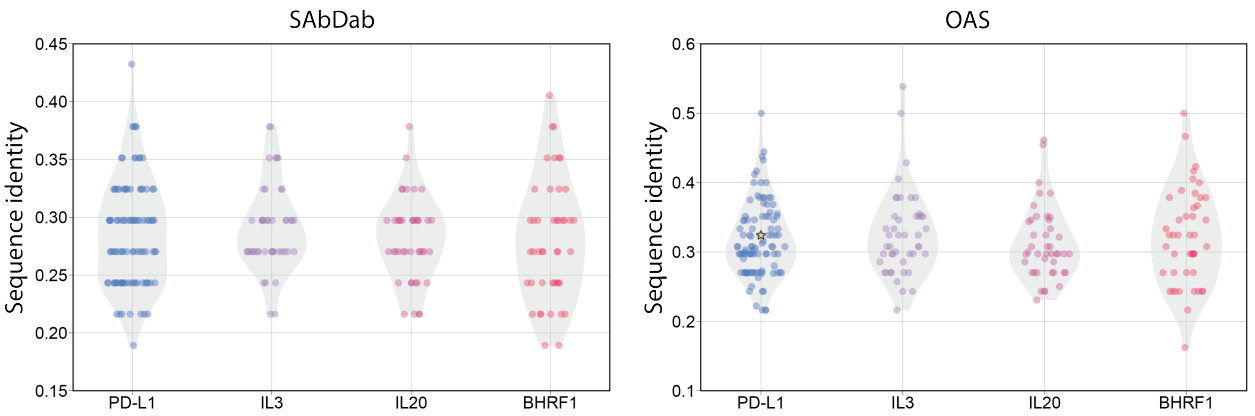

Figure S4 | **CDR sequence similarities to antibodies in SAbDab and OAS.** Alignments were created against the entire query sequence using `mmseqs` with a sensitivity of 3.0. The top hit was extracted and CDR similarity was calculated as the sequence identity of the top hit in the CDR positions.

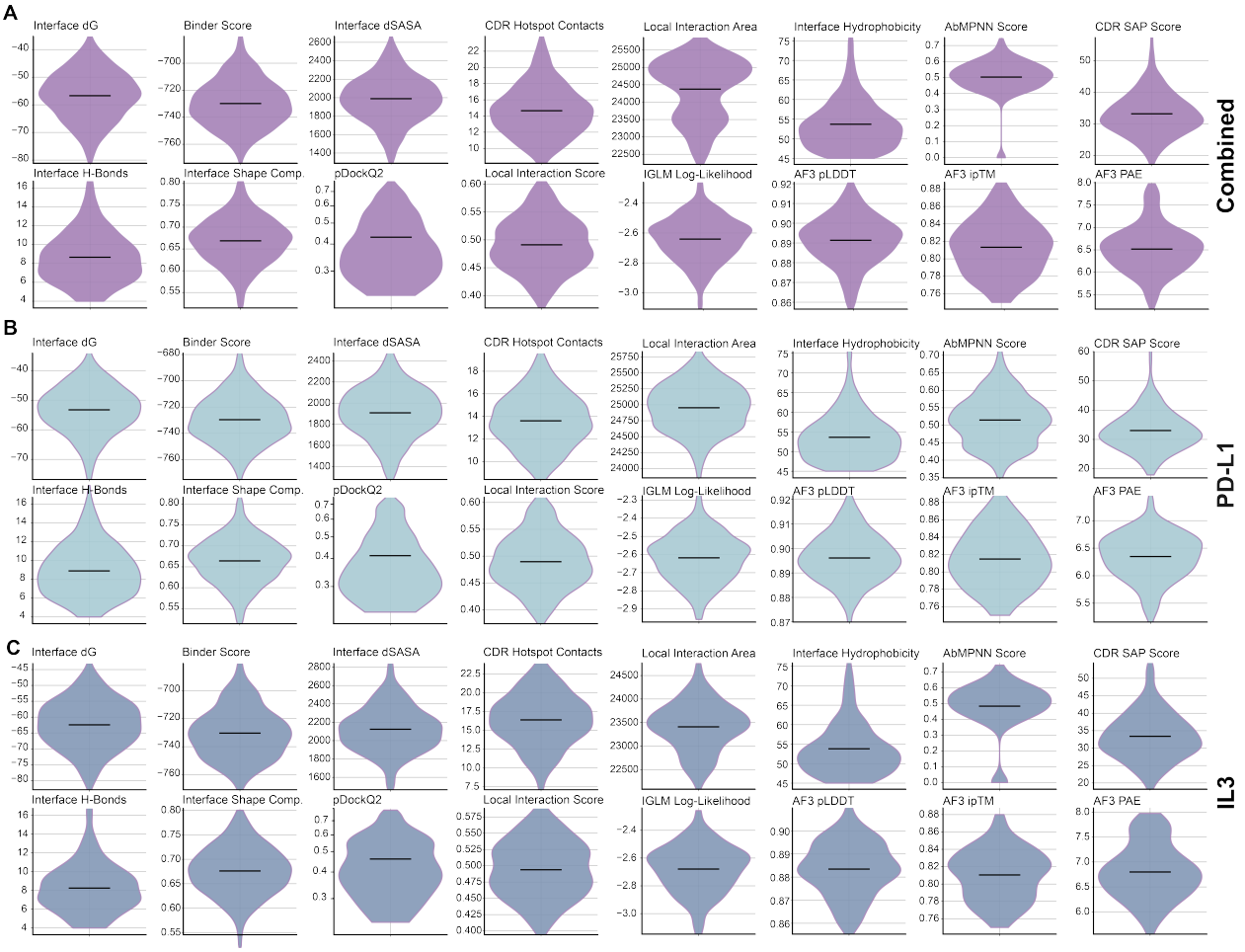

Figure S5 | **Scores distributions of key computational metrics for passing scFv designs.** (**A**) Distributions across all successful designs pooled from different targets. (**B,C**) Specific distributions for PD-L1 (**B**) and IL3 (**C**) designs.

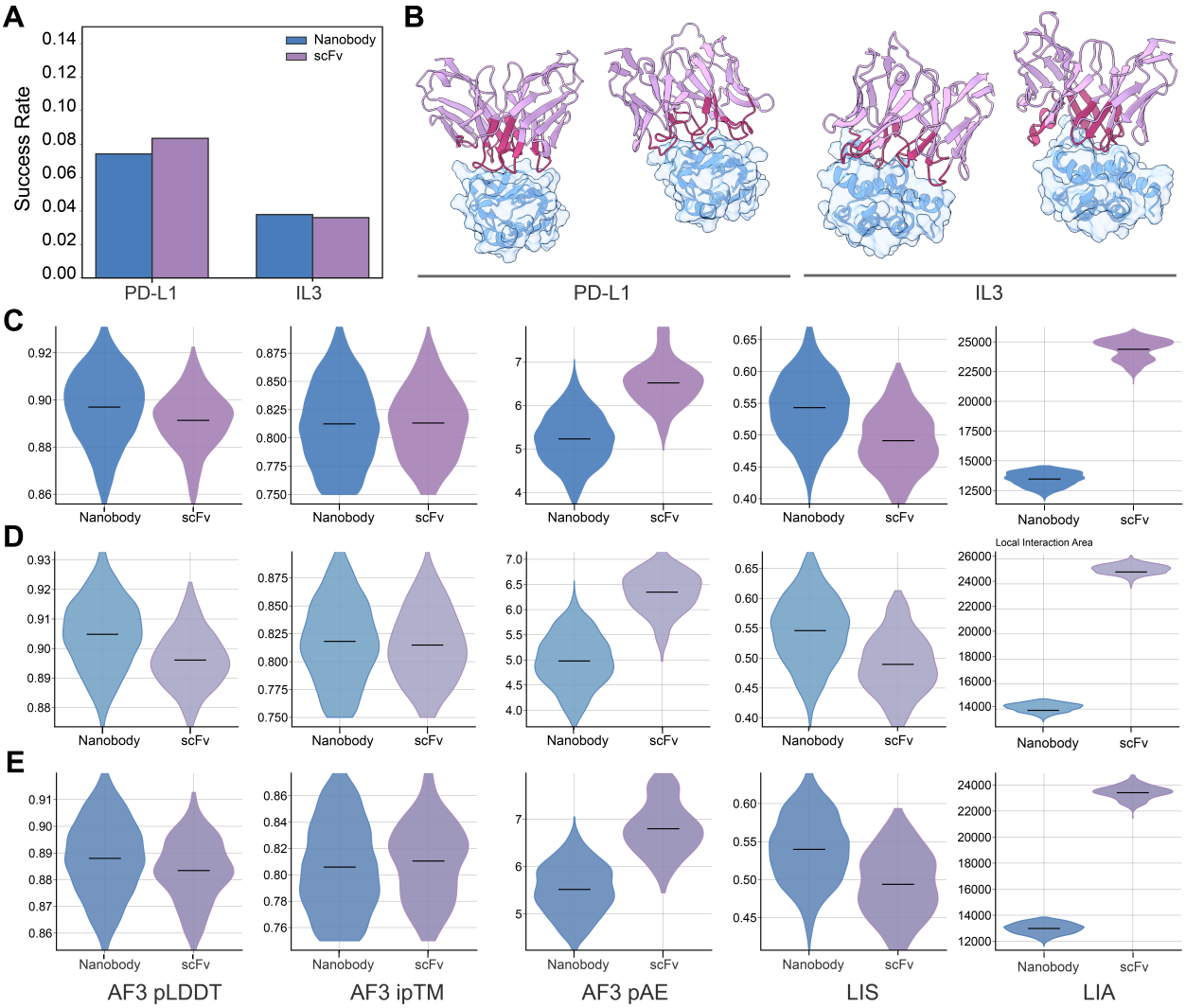

Figure S6 | **Design of scFv binders using Germinal.** (**A**) The success rates (defined as the number of passing designs divided by all designs) are shown for nanobody and scFv formats targeting PD-L1 and IL3. (**B**) Representative predicted binding structures of scFv binders targeting PD-L1 (left) and IL3 (right) are shown in complex with their respective antigens. (**C**) The distributions across all successful nanobody and scFv designs pooled from different targets. (**D,E**) Specific distributions for PD-L1 (**D**) and IL3 designs (**E**).

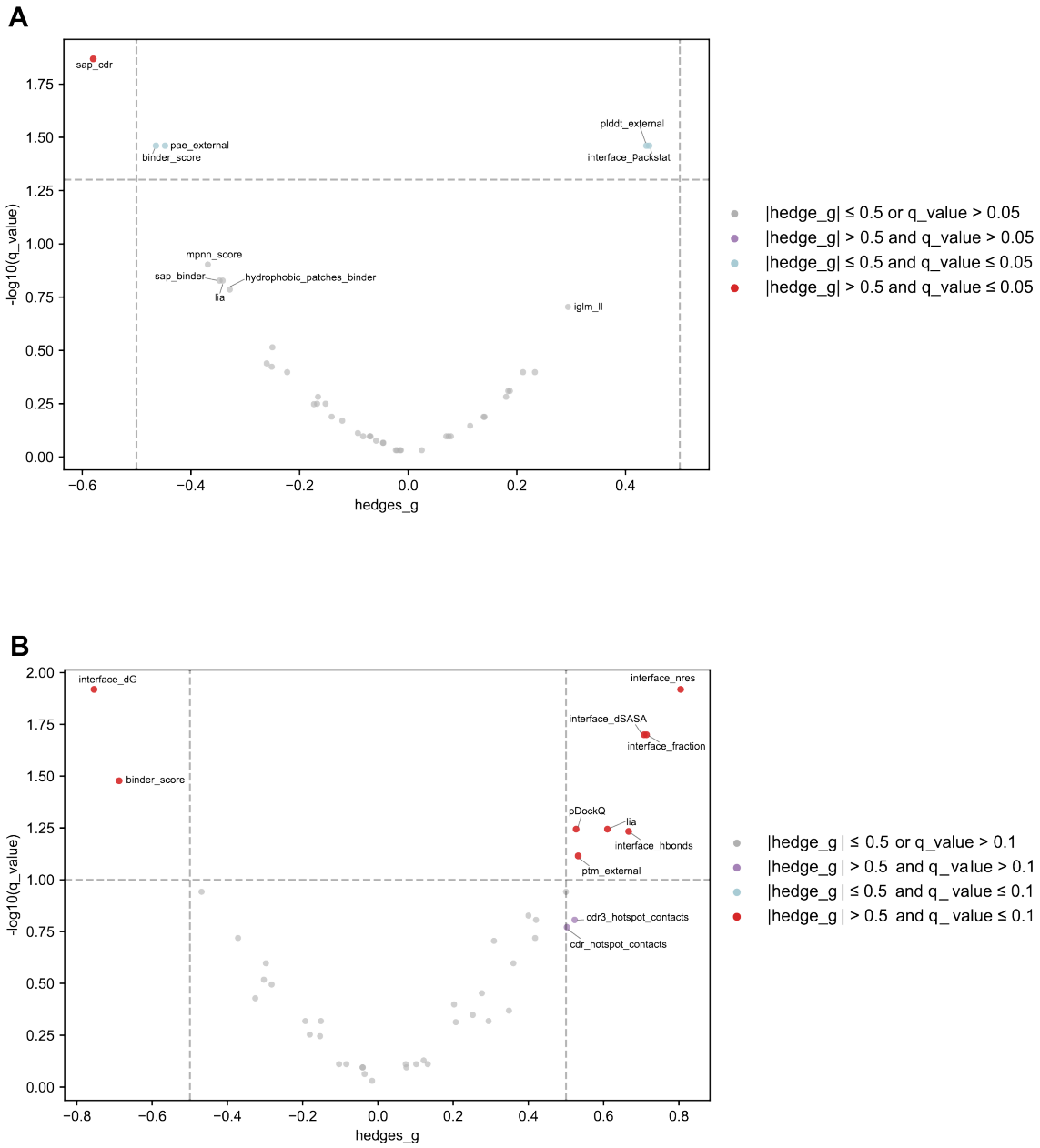

Figure S7 | **Identification of key computation metrics using differential analysis.** (**A**,**B**) Two groups of expressors were defined for expression analysis (**A**) based on the HiBiT read RLU values: top expressors (HiBiT read RLU > 1,000,000) and non-expressors (HiBiT read RLU < 150,000). Differential analysis was performed between these two groups, and metrics with Hedges' $g > 0.5$ and $q < 0.05$ were considered significant (red). Differential analysis was performed between BLI-verified binders and all other designs (**B**) to identify key metrics that affect binding. Metrics with Hedges' $g > 0.5$ and $q < 0.1$ were considered significant (red).

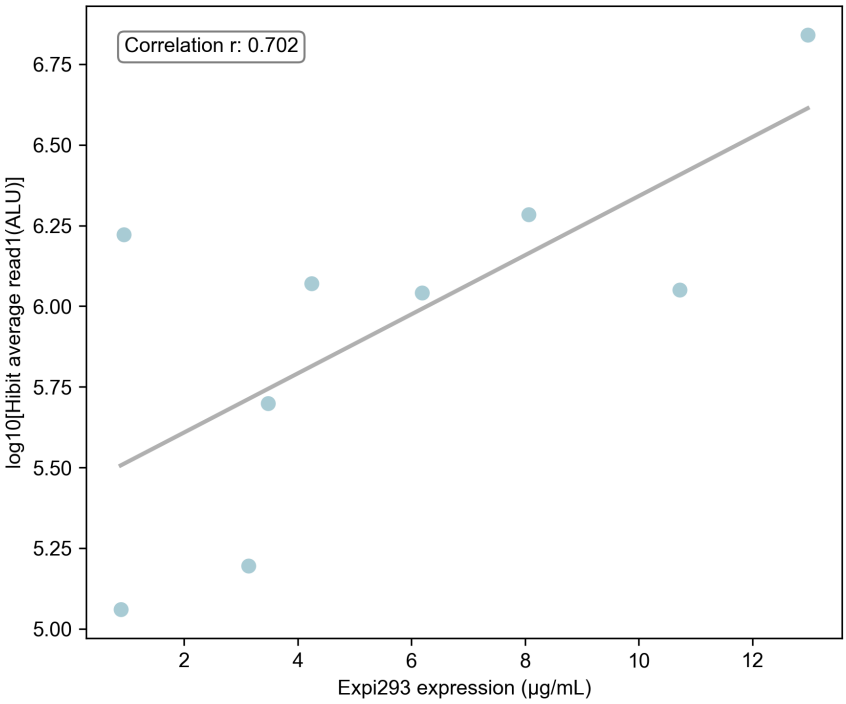

Figure S8 | **Correlation between luminescence signal (ALU) and Expi293 expression level.** The Expi293 expression data were obtained from the same batch of transfections used for the PD-L1 binder designs, with validated protein concentrations and recorded final volumes, to ensure consistent normalization across samples.

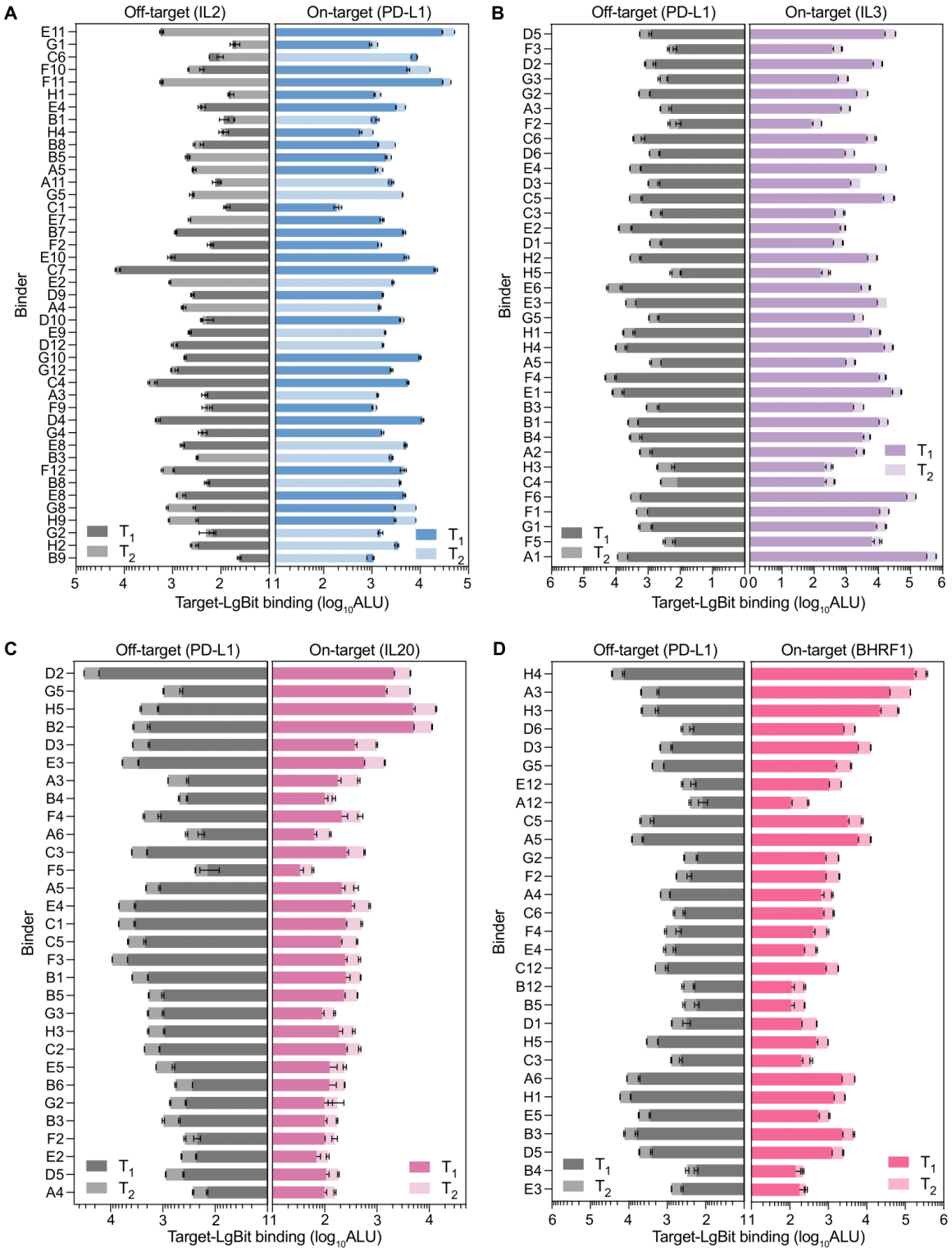

Figure S9 | **Raw on-target and off-target binding values at two time points.** Raw luminescence values from the split luciferase (SmBiT–LgBiT) binding assay for (**A**) PD-L1, (**B**) IL3, (**C**) IL20, (**D**) BHRF1 binders are shown. Each design was tested against its intended on-target antigen (right panels) and an off-target antigen (left panels) at two time points, T1 (dark shade) and T2 (light shade) (8 minutes after T1). Binders are plotted if they pass the expression threshold of >150,000 RLU as determined by the HiBiT assay for expression.

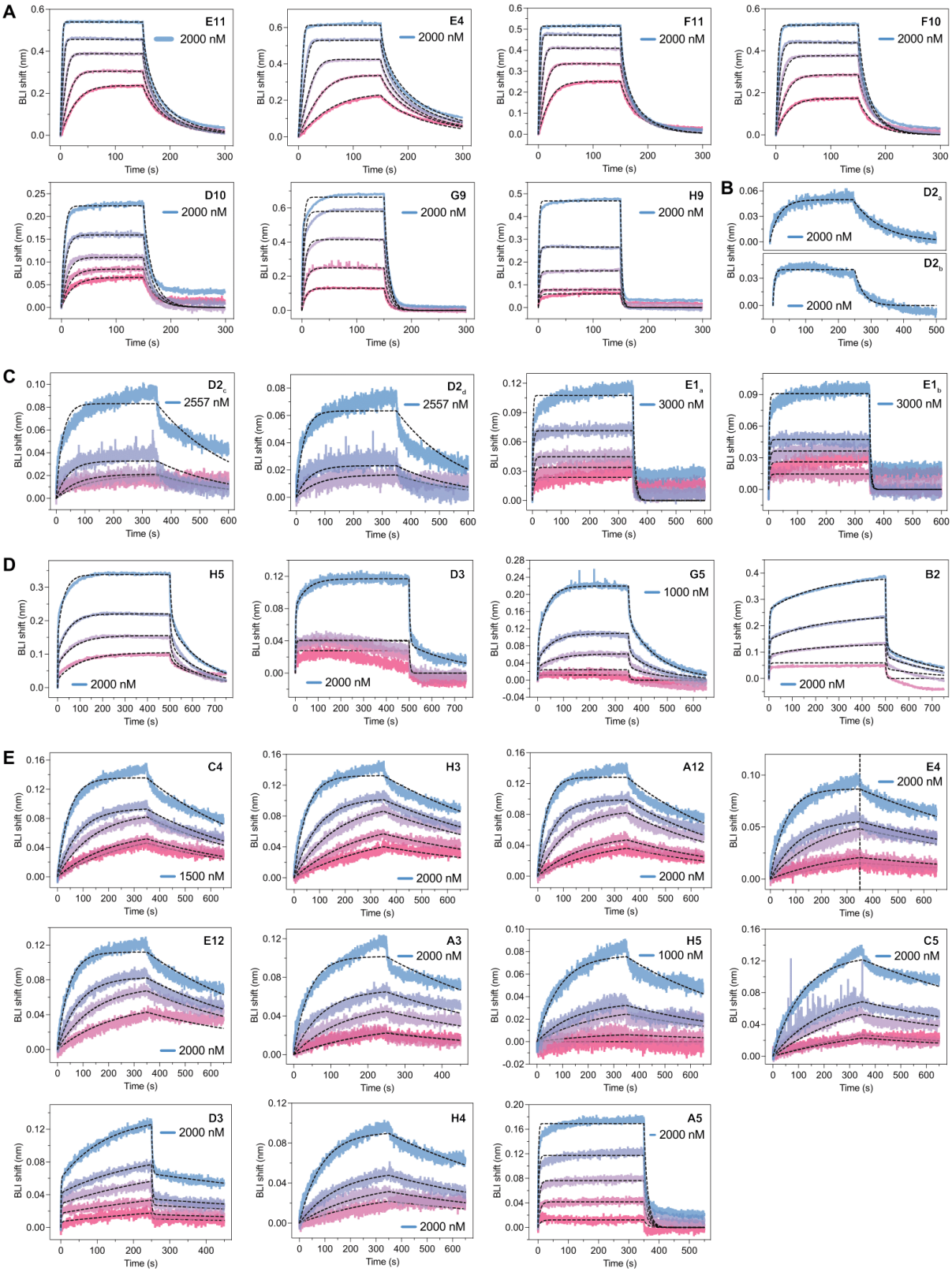

Figure S10 | **BLI sensograms for validated binders against all four target antigens.** (**A-E**) Sensograms for PD-L1 (**A**), IL3 (**B, C**), IL20 (**D**), and BHRF1 (**E**). For each BLI experiment, a titration of five concentrations was prepared by 2-fold serial dilution from the highest concentration tested, indicated on each plot.
For all BLI plots, binding curves are shown as colored lines, and kinetic fits are shown as black dashed lines.

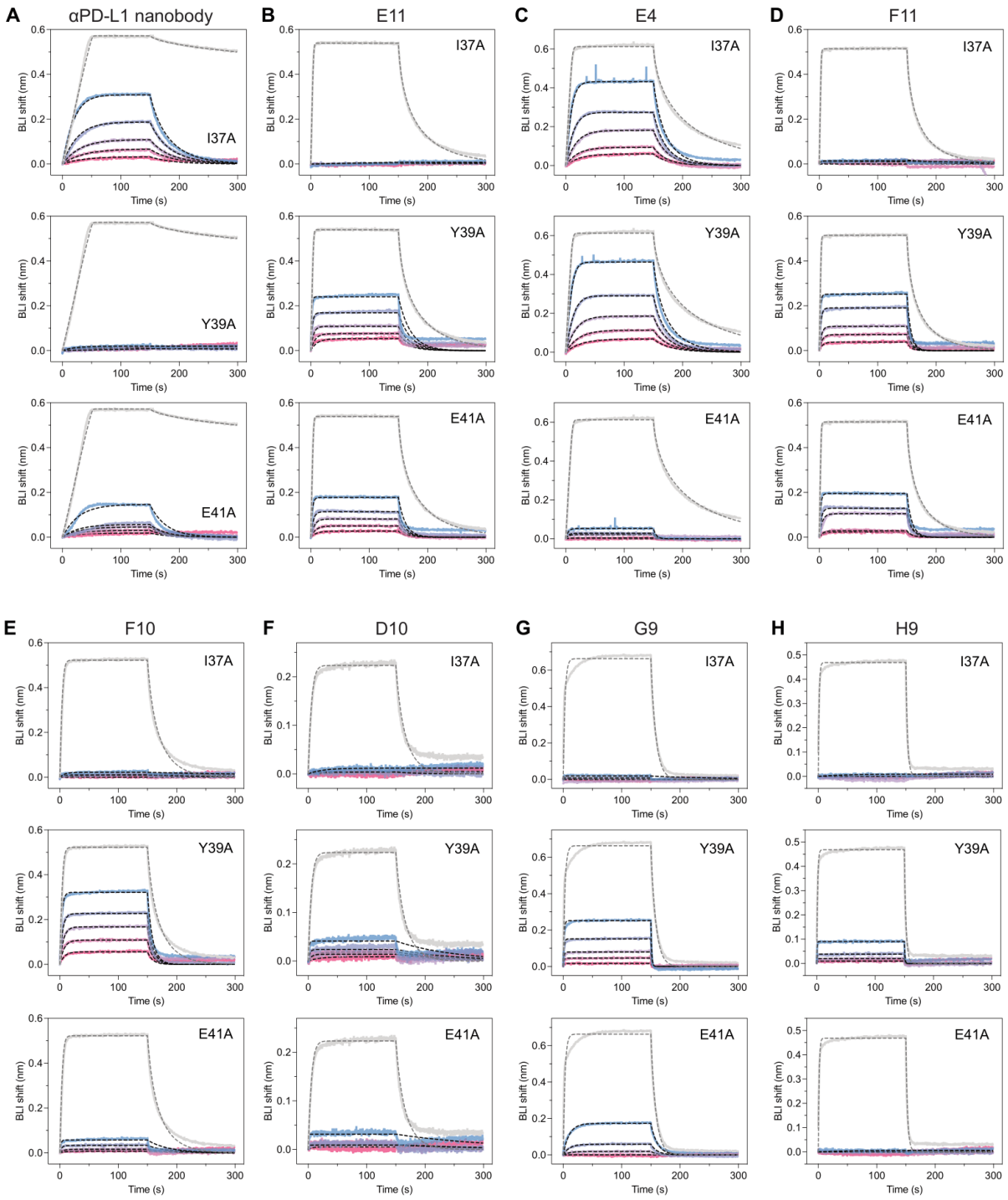

Figure S11 | **BLI sensograms showing the reduced binding of hit nanobody designs to PD-L1 variants harboring alanine substitutions in the Germinal-specified epitope.** Alanine substitution mutations were introduced in three distinct positions in the PD-L1 epitope to which binders were designed. A positive control nanobody against PD-L1 (αPD-L1 nanobody; **A**) and each of the BLI-validated binders (**B-H**) were assayed for binding against the three PD-L1 variants using BLI. K_D values reported in the main text were extracted from 1:1 binding models which were globally fit to each titration series (black dashed lines). In each binding experiment, a titration of five binder concentrations was prepared by 2-fold serial dilution from the highest concentration tested (2000 nM for all designed binders, 50 nM for αPD-L1 nanobody; blue BLI trace in each plot). The binding curve for each binder (at the highest binder concentration tested) to wild-type PD-L1 is shown on each plot as a gray trace, with a gray dashed line representing its kinetic fit to a 1:1 binding model.

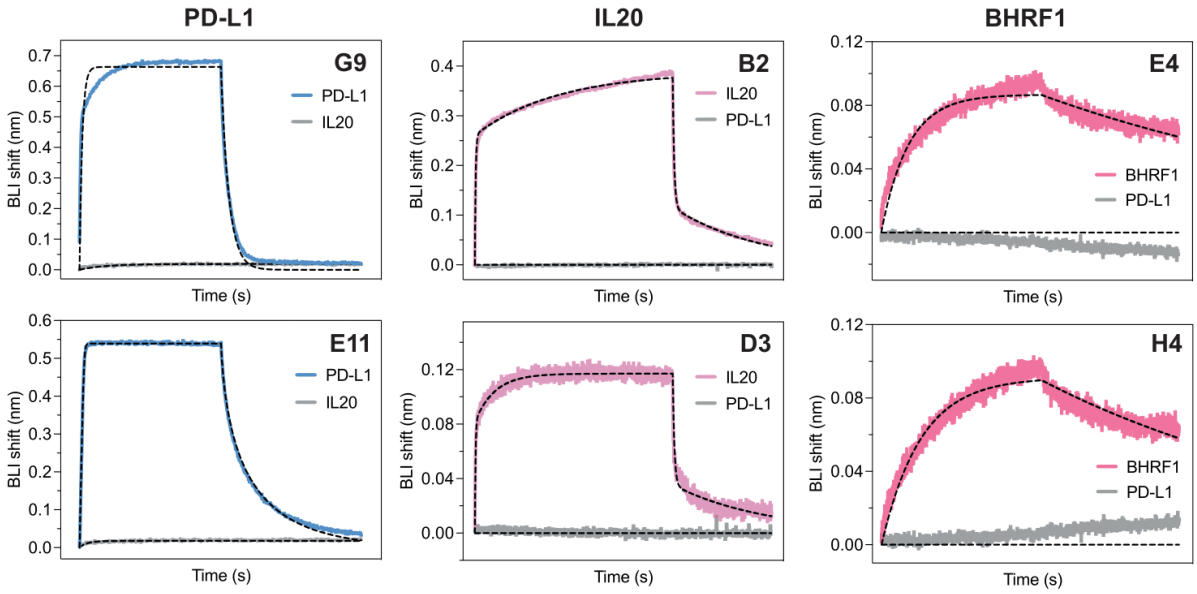

Figure S12 | **Testing target specificity of selected binders using BLI.** For each target antigen, two binders were selected based on their split-luciferase assay results: one with a low on-target/off-target binding ratio and one with a high ratio. These binders were tested on BLI against the corresponding off-target antigen at a concentration of 2000 nM to evaluate whether the split-luciferase–derived specificity metric correlates with BLI-based measurements. Binding responses to the off-target antigen are shown in gray, and the binding curves for the on-target antigen at 2000 nM are shown as colored traces, suggesting specific binding across scores of different range and that on-target/off-target ratios should not be used alone to define the specificity of designs. Binding curves are shown as colored lines, and kinetic fits are shown as black dashed lines.

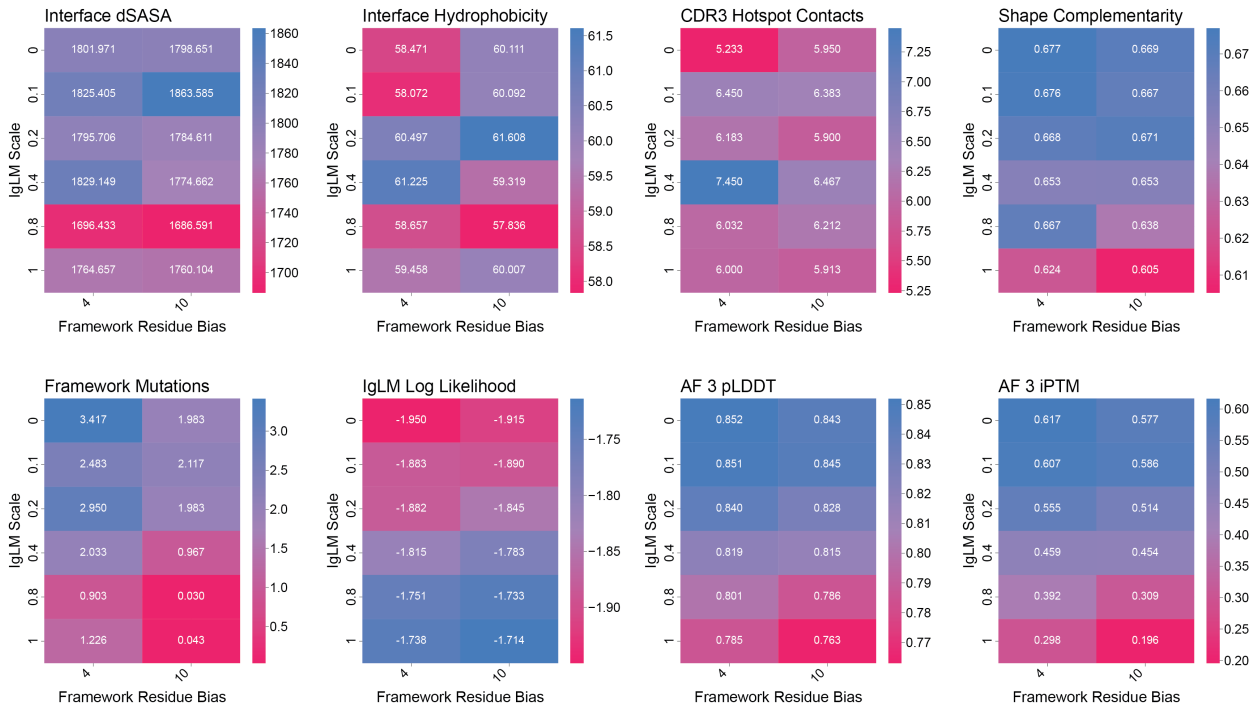

Figure S13 | **Comparison heatmap of selected metrics across a sweep of sequence biases and IgLM scale values.** Each cell reports the mean value of the corresponding metric over a small-scale run of 60 trajectories. Metrics in the top row show little dependence on either variable, whereas those in the bottom row exhibit clear trends as Sequence bias and IgLM scale vary.