Combinatorial Optimization of Antibody Libraries via Constrained Integer Programming

Conor F. Hayes[1,*], Andre R. Goncalves[2], Steven Magana-Zook[2], Jacob Pettit[2], Ahmet Can Solak[2], Daniel Faissol[2],
Mikel Landajuela[2,*] (landajuelala1@llnl.gov)

[2]Lawrence Livermore National Laboratory, Livermore, CA, USA [1]Cognizant AI Lab, San Francisco, CA, USA
*Equal contribution. Work performed while C.F.H. was at LLNL.

Abstract

Cold-start antibody campaigns need starting libraries that are both diverse and high quality, but existing workflows either require expensive screening or give little direct control over diversity. ProtLib-Designer (PLD) formulates library design as a multi-objective ILP with explicit constraints on library size, positional reuse, and repeated mutations, using AI-guided scores from protein language models and inverse folding. The optimizer directly targets user-specified library budgets, making it practical when downstream screening capacity is fixed and every sequence slot matters. Across Trastuzumab, D44.1, and Spesolimab, PLD outperforms LMG, MODIFY, and SPEA2 on BEU and hypervolume while retaining practical residue entropy and competitive humanness. The result is a training-free library design workflow with interpretable controls instead of a learned diversity heuristic. **Code:** `https://github.com/LLNL/protlib-designer`
Keywords: Combinatorial Optimization, Integer Linear Programming, Antibody Library Design, Diversity Constraints, Protein Language Models, Deep Mutational Scanning.

Motivation

- **The problem.** Directed-evolution antibody campaigns need diverse, high-affinity starting libraries even when no assay data exist yet. In this cold-start setting, a good library must cover sequence space while still placing many candidates near the best predicted trade-offs. The design problem is therefore combinatorial, budget-constrained, and inherently multi-objective.
- **Why existing methods fall short.** Experimental DMS and simulation-heavy workflows are expensive, while ML generators such as LMG and MODIFY do not expose first-class diversity control [1, 2]. Pareto-only search such as SPEA2 can find strong points, but it does not reliably deliver a user-requested library size [8]. In practice, users need a fixed-size library with both objective quality and broad mutation coverage.
- **Our contribution.** PLD plugs protein language model and inverse-folding scores into a diversity-constrained ILP. The result is a library optimizer with a small set of interpretable knobs instead of a learned diversity surrogate. That makes it easy to adapt the same framework to different antibody systems and screening budgets.

Methods

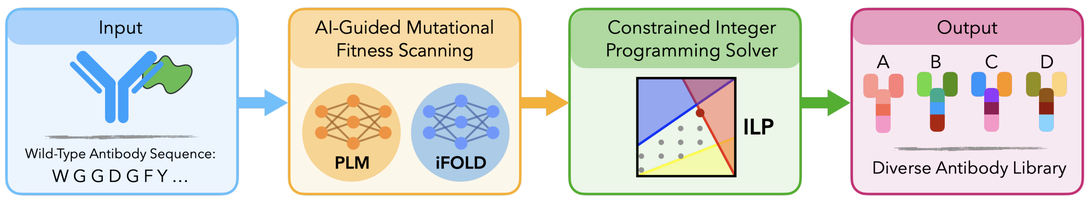

Figure 1: ProtLib-Designer overview. Starting from an antibody-antigen complex, PLD scores candidate mutations with protein language model and inverse-folding objectives, then solves a constrained ILP to return a fixed-size, diverse antibody library.

In-silico deep mutational scanning. Each candidate mutation receives two additive scores: an intrinsic PLM term from wild-type marginal likelihood and an extrinsic inverse-folding term conditioned on the antibody-antigen complex [6, 3, 4]. Lower scores are better. This gives PLD a paired view of sequence plausibility and structure-aware compatibility without requiring task-specific retraining.

$$
\begin{aligned}
s_{ij}^{\mathrm{PLM}} &= -\log p(x_i = a_j \mid \mathbf{w}) + \log p(x_i = w_i \mid \mathbf{w}) \\
s_{ij}^{\mathrm{IFOLD}} &= -\log p(x_i = a_j \mid \mathbf{w}_{<i}, \mathrm{struct}(\mathbf{w}, \mathbf{g})) \\
&\quad + \log p(x_i = w_i \mid \mathbf{w}_{<i}, \mathrm{struct}(\mathbf{w}, \mathbf{g})).
\end{aligned}
$$

Multi-objective ILP per mutant. Binary variables z_{ij} choose amino acid j at mutable position i. For objectives $Q = \{\mathrm{PLM}, \mathrm{IFOLD}\}$, PLD solves

$$
\begin{aligned}
\min_{\mathbf{z}} \quad & \left\{ \sum_{i=1}^{N} \sum_{j=1}^{M} s_{ij}^{q} z_{ij} \right\}_{q \in Q} \\
\text{s.t.} \quad & \sum_{j=1}^{M} z_{ij} \le 1 \ \forall i, \qquad n_{\min} \le \sum_{i=1}^{N} \sum_{j=1}^{M} z_{ij} \le n_{\max}, \\
& z_{ij} = 0 \text{ when } a_j = w_i, \qquad z_{ij} \in \{0, 1\}.
\end{aligned}
$$

Sampling $\boldsymbol{\lambda} \sim \mathrm{Dirichlet}(1, 1)$ sweeps different Pareto trade-offs without retraining any model. In the reported experiments, each system uses a fixed 1,000-sequence budget and sequence-level mutation counts constrained between $n_{\min} = 5$ and $n_{\max} = 8$.

ProtLib-Designer: solve-and-remove. For $k = 1, \ldots, K$, PLD solves the ILP, stores the best mutant, then removes a neighborhood around every prior solution. The ball radius ϵ suppresses near-duplicates, δ_1 limits overused positions, δ_2 limits repeated mutations, and optional ξ cycles the mutation count across the allowed range. This solve-and-remove loop is what converts strong single-mutant optimization into controllable library-level diversity.

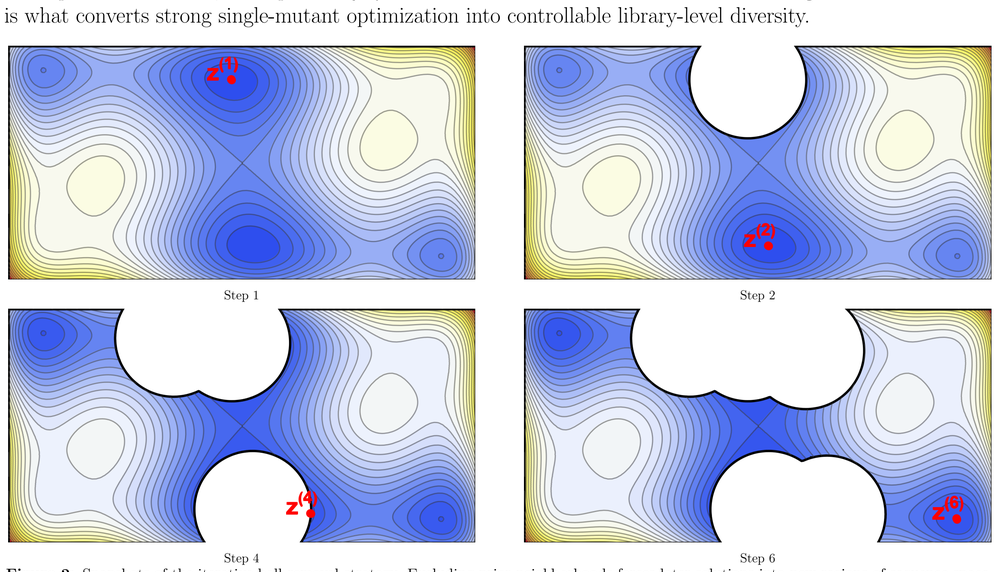

Figure 2: Snapshots of the iterative ball-removal strategy. Excluding prior neighborhoods forces later solutions into new regions of sequence space.

How the user-facing knobs behave. The diversity controls act at library level rather than at the score level. That makes each knob easy to interpret before any screening data exist.

What you tune. Only four knobs matter in practice: ϵ, δ_1, δ_2, and $(n_{\min}, n_{\max})$. ϵ controls local redundancy, δ_1 limits position reuse, and δ_2 limits mutation reuse across the whole library. There is no training loop and no learned diversity proxy, so users can directly decide how broad or conservative the final library should be.

Results

PLD is evaluated on Trastuzumab and D44.1 [5] and on Spesolimab [7], where it is the only method that consistently combines batch-size control with strong multi-objective coverage. These three systems span increasing search-space difficulty, providing a useful stress test for whether diversity constraints remain effective outside a single benchmark.

Objective-space behavior (Trastuzumab). PLD libraries cluster near the Pareto front in clear stratified layers, while LMG and MODIFY spread broadly across the objective space [1, 2]. SPEA2 reaches strong points, but it cannot reliably populate the requested 1,000-mutant library [8]. The visible layering is expected: each solve-and-remove step peels away a local neighborhood, forcing subsequent ILP solves to occupy a new region of sequence space.

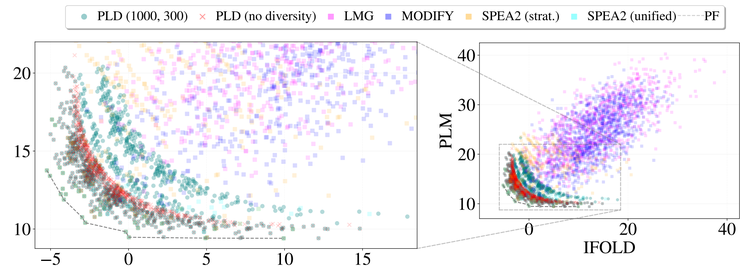

Figure 3: PLD stays near the Pareto front in layered bands, while LMG and MODIFY scatter broadly.

Headline metrics.

System	Method	Ent. ↑	BEU ↓	HV ↑	Human. ↑	Oracle (%) ↑
Trastuzumab	**PLD**	**3.22**	**4.62**	**2231.9**	**-1296.7**	**58.2**
	LMG	4.75	10.70	1937.7	-1309.8	17.1
	MODIFY	3.97	10.14	2012.4	-1308.3	18.6
D44.1	**PLD**	**2.93**	**1.84**	**1262.6**	**-1815.9**	—
	LMG	5.42	10.81	705.8	-1830.8	—
	MODIFY	4.75	12.60	711.9	-1838.8	—
Spesolimab	**PLD**	**3.03**	**-7.28**	**552.3**	**-940.3**	—
	LMG	6.01	4.50	118.7	-949.4	—
	MODIFY	5.68	2.24	89.8	-948.6	—
	SPEA2 strat.	2.99	-6.51	552.3	-942.6	—
	SPEA2 unif.	3.10	-6.19	528.0	-941.6	—

Table 1: Expanded baselines. PLD keeps the best BEU and matches the best HV while maintaining entropy above 2.9.

Reading the benchmark table. High raw entropy by itself is not enough: LMG and MODIFY often achieve broader residue usage by drifting farther from the Pareto front, whereas PLD concentrates diversity in stronger-scoring regions. The expanded table preserves the same practical pattern: PLD keeps near-front coverage while returning a fixed-size, diverse library.

Pareto-front recovery. PLD recovers the full Pareto front on all three systems; on Spesolimab, only PLD and stratified SPEA2 match the front. This matters because Spesolimab exposes a 47-position design space, where naive diversification becomes especially hard.

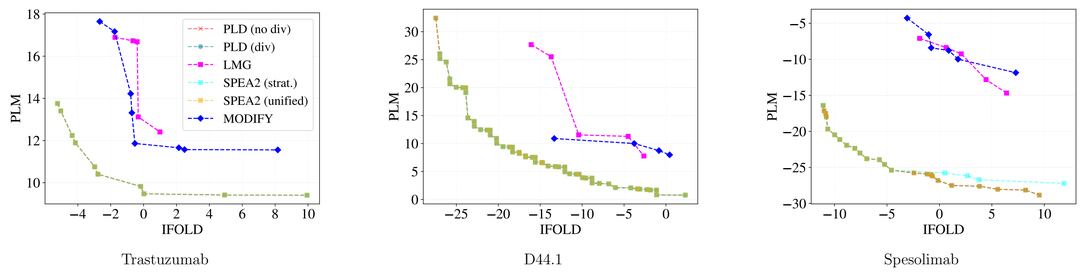

Figure 4: PLD recovers the full Pareto front on all three systems; only PLD and SPEA2 (stratified) match it on Spesolimab despite its 47-position search space.

Why the constraints matter most on hard systems. The Spesolimab case is the most revealing because its 47 mutable positions create a much larger search space than Trastuzumab or D44.1. In that regime, unrestricted diversification tends to waste budget on distant but low-quality mutants, while purely objective-driven search reuses the same local motifs. PLD keeps both pressures in play at once by solving for quality and diversity in the same optimization step.

Operational takeaway. Across all three systems, PLD is the only method that jointly gives strong objective values, explicit diversity control, and deterministic library size. That combination is what makes it practical for real screening campaigns rather than just Pareto-front analysis.

Takeaways

- PLD is a multi-objective ILP with first-class diversity constraints for cold-start antibody library design.
- Only four knobs matter in practice: ϵ, δ_1, δ_2, and $(n_{\min}, n_{\max})$, with no training or learned diversity proxy.
- PLD achieves the best average-rank behavior across Trastuzumab, D44.1, and Spesolimab against LMG, MODIFY, and SPEA2 while preserving fixed library size.
- Diversity is enforced during optimization rather than added later as a filtering heuristic, and explicit library size maps naturally onto fixed assay budgets such as 96-, 384-, or 1,000-member campaigns.
- Open source: `github.com/LLNL/protlib-designer`

Limitations/Future work. PLD currently assumes an antibody-antigen complex structure, and total ILP cost grows linearly with library size. Natural next steps are breadth optimization, richer QAP-style formulations, and extending the same optimization logic to broader protein library design settings.

References

[1] Thomas A Desautels, Kathryn T Arrildt, Adam T Zemla, Edmond Y Lau, Fangqiang Zhu, Dante Ricci, Stephanie Cronin, Seth J Zost, Elad Binshtein, Suzanne M Scheaffer, et al. Computationally restoring the potency of a clinical antibody against omicron. *Nature*, pages 1–8, 2024.
[2] Kerr Ding, Michael Chin, Yunlong Zhao, et al. Machine learning-guided co-optimization of fitness and diversity facilitates combinatorial library design in enzyme engineering. *Nature Communications*, 15, 2024.
[3] Brian L Hie, Varun R Shanker, Duo Xu, et al. Efficient evolution of human antibodies from general protein language models. *Nature Biotechnology*, 2023.
[4] M. H. Hoie, Alissa M. Hummer, T. H. Olsen, et al. Antifold: Improved antibody structure-based design using inverse folding, 2024. Preprint.
[5] Derek M. Mason, Simon Friedensohn, Cédric R. Weber, et al. Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning. *Nature Biomedical Engineering*, 5:600 – 612, 2021.
[6] Joshua Meier, Roshan Rao, Robert Verkuil, Jason Liu, Tom Sercu, and Alexander Rives. Language models enable zero-shot prediction of the effects of mutations on protein function. *bioRxiv*, 2021.
[7] Amir Shanehsazzadeh, Julian Alverio, George Kasun, et al. Igdesign: In vitro validated antibody design against multiple therapeutic antigens using inverse folding. *bioRxiv*, 2024.
[8] Eckart Zitzler, Marco Laumanns, and Lothar Thiele. SPEA2: Improving the strength pareto evolutionary algorithm. Report, ETH Zurich, Computer Engineering and Networks Laboratory, Zurich, 2001-05.
Acknowledgements. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC. LLNL-CONF-868784.