From Incremental to Exponential: Integrating AI and DEL to Enable Discovery Across the Proteome

Heta A. Gandhi[1†], Elena Cáceres[1†], Cristiana Carpinteiro[2], Manuel Ravasqueira[2], Mridula Bontha[1], Brandon Bravo[1], Graham Carlson[1], Jim Davis[3], Ketki Dhamnaskar[1], John Eichenseer[1], Telmo Felgueira[2], Julian Fernandez[2], Emel Ficici[1], Jamie Furneisen[1], Jennifer Greufe[3], Chad Hewitt[1], Tim Kane[1], Tatiana Kennedy[1], Buckley Kolhauff[1], Emily Kruger[2], Ajay Kulkarni[1], Evan Landstrom[1], Calvin Larson[1], John Lee[1], Emily Low[1], Marie Malone[1], Ratul Mukerji[1], Ryan Pemberton[1], Chris B. Phelps[1], Prabhu Raman[1], Patricia Rocha[2], Sushmitha Regulapati[2], Jorge Sampaio[2], Nick Sanchez[1], Jose Santos[1], Julie Sheung[1], Vaishali Sridhar[1], Jeffrey Wu[1], Lik Hang Yuen[1], Stephanie Yung[1], Kiril Zelenkovski[2], Paul Novick[1]

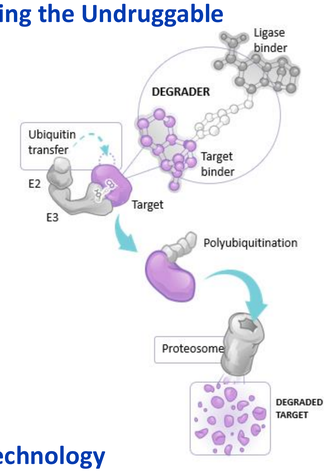
[1]Nurix Therapeutics, Inc., San Francisco, CA; [2]Loka Inc, San Francisco, CA; [3]Amazon Web Services, Seattle, WA

Targeted Protein Degradation for Drugging the Undruggable

Nurix Therapeutics combines DNA-Encoded Library (DEL) screening with machine learning to accelerate small molecule drug discovery.

This enables the design of bifunctional degraders that direct disease-causing proteins to the ubiquitin–proteasome system for targeted elimination.

Nurix aims to harness the full potential of DEL-guided AI to create an autonomous discovery engine—accelerating the design of next-generation degraders and expanding the reach of targeted protein degradation across previously inaccessible targets.



DNA Encoded Library (DEL) Screening Technology

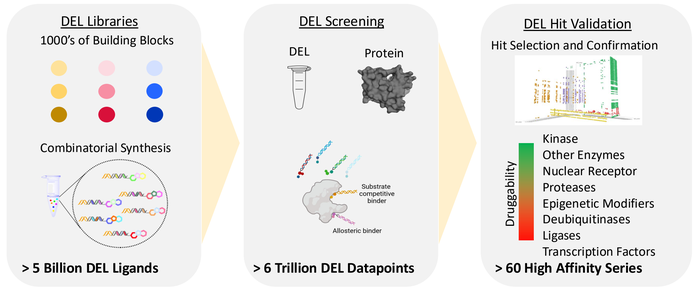
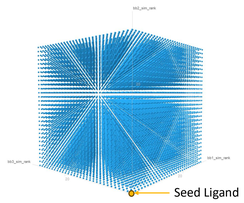


Figure 1: DEL Data Generation at Nurix DNA-Encoded Libraries[1] (DELs) are collections of small molecules encoded with unique DNA barcodes that record their synthetic history and building block identity. These libraries are pooled in a screening format that enables screening billions of compounds against a desired target. DELs enable rapid identification of binders by sequencing the enriched DNA tags after affinity selection of the desired target.

Leveraging DEL SAR for Hit Optimization

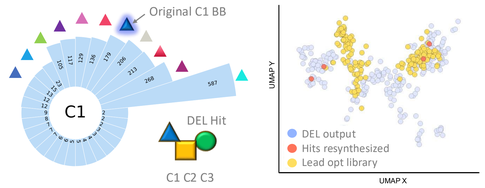


Zooming in to DEL Ligands constructed of 30 most similar BBs to those in a DEL ligand of interest (seed ligand).

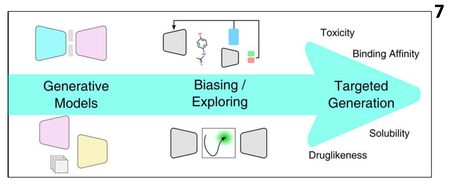
Typically, only a few compounds from the entire DEL output are resynthesized and validated – massively underutilizing the DEL platform's full potential.

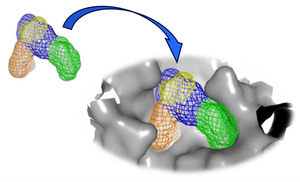
At Nurix, we take an expansive and data-driven approach to lead optimization that leverages broader insights from DEL outputs to improve downstream medicinal chemistry efforts. Multiple computational methods are used to enumerate parallel libraries and triage the designs.

Cheminformatics: Enables efficient SAR analysis, clustering, and visualization of DEL chemical space to prioritize analogs with optimal properties.



Generative ML: Designs novel compounds with better properties, that are also synthetically accessible. DEL ML and DEL Foundation models predict bioactivity while ADMET models predict molecular properties, which accelerates compound prioritization.





Computational Chemistry: Molecular docking, dynamics, and shape-based queries to model binding interactions and optimize ligand-target affinity at the atomic level.

Case Study: Target X

Chemists and computational scientists designed and triaged initial hit expansion libraries for two validated DEL hit series with the aid of DEL SAR insights

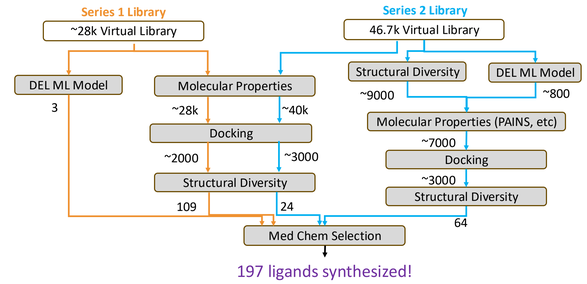


Figure 6: Initial hit expansion library triage workflow using computational methods, structural diversity and medicinal chemistry intuition.

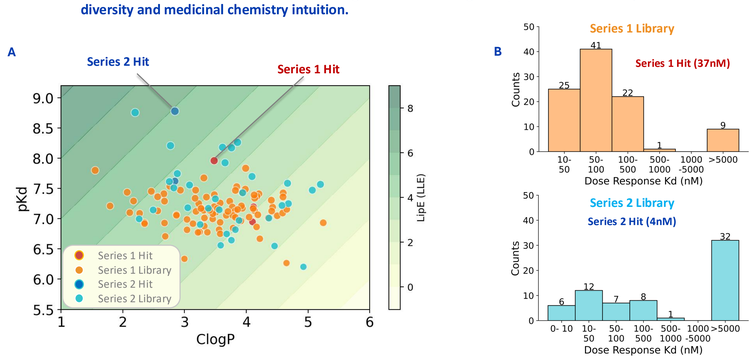


Figure 7: Initial hit expansion libraries improve molecular properties and lipophilic efficiency of the validated DEL hits and provide further information for medicinal chemistry optimization: A. Molecular lipophilicity plotted against SPR Kd (logscale) for two target X DEL expansion libraries; the color gradient shows the lipophilic ligand efficiency (LLE). Red and dark blue dots represent the reference Series 1 and Series 2 hits, respectively, and their enantiomers while orange and cyan colors correspond to the DEL expansion library molecules. B. Histograms of SPR Kds (nM) for the two libraries. These were not informed by co-crystal structures.

DEL Foundation Models (DEL FM) Enable *in Silico* DEL Experiments

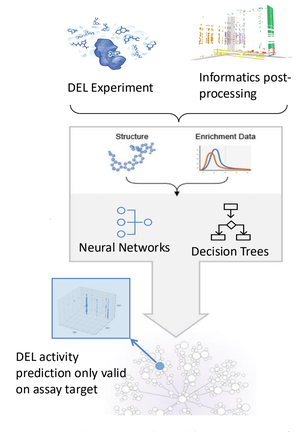


Figure 2: DEL ML Pipeline: This category of models takes enrichment data from a single DEL experiment and predicts activity based on molecular featurization. Models then are used for hit triaging, noise reduction, and informing synthesis follow-up[2-6]

These models predict binding to a single protein, are quick to train, and are effective on single target prediction tasks.

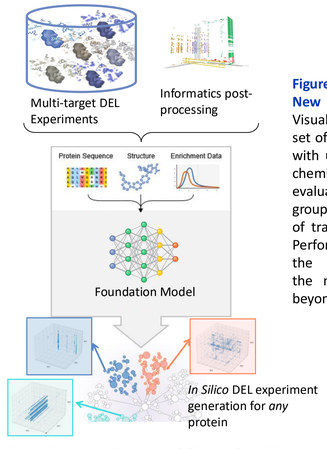


Figure 3: A Foundation Model Trained on DEL: DEL FM pipelines take enrichment data from multiple DEL experiments across different proteins and conditions. Models learn a generalizable relationship between protein sequence and molecular structure

DEL FM then predicts activity of each molecule, protein pair as if it were DEL data without experimental data prerequisites for either member

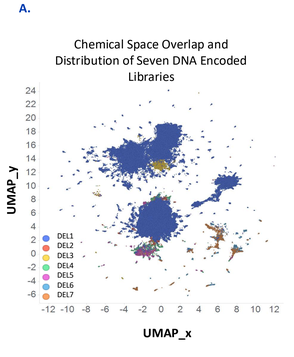
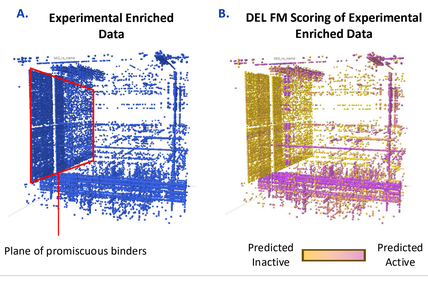


Figure 4: DEL FM Predicts New Chemical Matter: A. Visual depicting overlap of a set of DNA Encoded Libraries with unique sets of reaction chemistries. B. Procedure to evaluate performance. Each group was iteratively left out of training data for DEL-FM. Performance on each DEL in the holdout demonstrates the model's generalizability beyond baseline (dotted line)



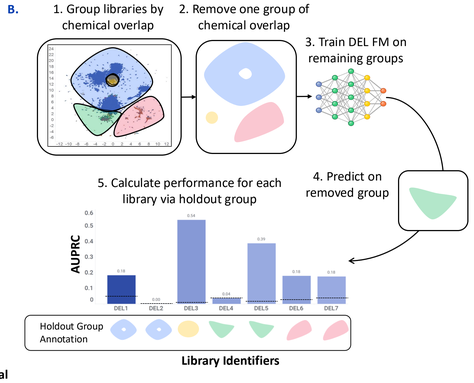


Figure 5: DEL FM Predicts and Denoises Enriched Chemical Matter: A. Enriched chemical matter of an unseen protein. B. Binders predicted by DEL FM largely recapitulate experimental data of a protein complex. Coloring of these molecules by DEL FM annotates important structures (purple) while ignoring promiscuous plane (yellow).

Conclusions

DEL datasets provide **large-scale, structured, and chemically diverse interaction data ideal** for training deep learning models. The scale of these models **enable robust pretraining and fine-tuning for downstream tasks** including lead discovery and optimization. This positions DEL-AI approaches to generalize beyond single screens, accelerating drug discovery even in low-data or novel target settings.

The **synergy between DEL and a multitude of complementary computational methods** drives a more efficient hit-to-lead process. Data-guided modeling and iterative experimentation accelerate and sharpen the expansion of promising hits into high-quality drug candidates.

By **integrating DEL-guided machine learning with computational chemistry and automated chemistry**, Nurix is building a discovery engine that moves degrader creation *from virtual to viable*—accelerating the path from computational prediction to clinically meaningful therapies.

References

1. Liu, J. et al. DNA-encoded Library Technology for Drug Discovery, Royal Society of Chemistry (2025).
2. Binder, P. et al. Partial Product Aware Machine Learning on DNA-Encoded Libraries. doi:10.48550/arxiv.2205.08020 (2022).
3. McCloskey, K. et al. Machine learning on DNA-encoded libraries: A new paradigm for hit finding. J. Med. Chem. 63, 8857–8866 (2020).
4. Ma, R. et al. Regression modeling on DNA encoded libraries.
5. Lim, K. S. et al. Machine learning on DNA-encoded library count data using an uncertainty-aware probabilistic loss function. J. Chem. Inf. Model. 19, 28–28 (2021).
6. Iqbal, S. et al. DEL+ML paradigm for actionable hit discovery – a cross DEL and cross ML model assessment. *ChemRxiv* doi:10.26434/chemrxiv-2024-2xrx4 (2024) .
7. Sousa, T. et al. Generative Deep Learning for Targeted Compound Design. J. Chem. Inf. Model. 61, 11, 5343-5361 (2021).

Presented at The Berkeley Biotech Research Day on July 22nd, 2025