CASE STUDY

Cyclica's MatchMaker™ complements DNA Encoded Library (DEL) experimental techniques for predicting drug discovery targets

Opportunity
The study aims to understand the strengths and synergies between DNA encoded libraries (DEL) technology and Cyclica's MatchMaker™ for improving the drug discovery process.

Technology
DNA encoded libraries (DEL) is an affinity-based experimental screening method to identify hit compounds. Cyclica's MatchMaker™ is a deep-learning framework for protein-ligand interactions.

Solution
The conclusions highlight that combining experimental DEL data with computational predictions from MatchMaker™ can offer enhanced predictivity. This synergy between the methods can lead to more accurate predictions than using each method individually, providing an opportunity to improve the drug discovery process.

Summary

DNA encoded libraries (DEL) are an in vitro method for identifying hit compounds for a given protein target (review). As such, DEL competes with computational methods such as virtual screening and Machine Learning (ML) driven drug/target interaction (DTI) prediction. Cyclica has built MatchMaker™, a single deep-learning framework for protein-ligand interactions across the entire proteome. Here, in collaboration with WuXi AppTec, a global company that provides a broad portfolio of R&D and manufacturing services that enable companies in the pharmaceutical, biotech and medical device industries to advance discoveries and deliver groundbreaking treatments to patients, we perform a comparative analysis of DEL vs. MatchMaker™ in finding molecules that bind a well-studied protein kinase, Aurora kinase A. We also investigate the possibility of combining both methods to obtain an improved hit rate over both. We find that Cyclica's MatchMaker™ has as much power to predict binding as the experimental DEL data, amongst a set of 41 molecules for which we obtained experimental on-DNA validation results. Furthermore, we find that combining both sets of data can substantially increase predictivity.

Methods

WuXi AppTec performed a DEL screen on Aurora kinase A and identified 28,173 target-enriched molecules in the binding fraction of the pull-down experiment. They resynthesized (on-DNA) a selection of these molecules and obtained ASMS binding data on 41 of them and their byproducts (25 binders, 16 non-binders). Cyclica compared predictive MatchMaker™ scores between the binders and non-binders to assess whether MatchMaker™ scores predict validation results. We also used a pareto optimal combination of multiple factors to obtain compound rankings by predictive, experimental and combined criteria for a more general assessment of relative predictive power and synergy between the experimental and computational methods.

Results

Figure 1 shows the distribution of MatchMaker™ scores between the binders and non-binders. It is clear that there is considerable predictivity in the scores. Notably, 100% of the molecules with scores greater than -2 are binders. The ROC-AUC value for predicting binding by MatchMaker™ score alone is 0.7. The difference between scores of binders and non-binders is significant, using the Mann–Whitney U test (p-value < 0.02).

To exploit this observed predictivity better, we used a Pareto optimal combination of factors to compute a "fitness" for each molecule, using our POEM technology [1]. Three types of fitness were computed: 1) "fitness_del", a purely experimental combination of counts from the DEL data, 2) "fitness_predict", a purely predictive combination of

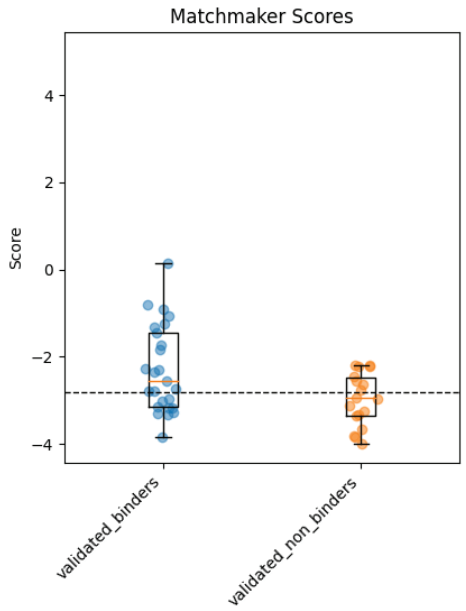

Figure 1. Distribution of MatchMaker™ scores between binders and non-binders for the target protein kinase.

factors derived from the molecular structure including MatchMaker™ score, and 3) "fitness", a combination of both predictive and experimental factors. We ranked all molecules by each of these values and plotted the inverse log ranks defined as *− log(rank/N)* of the tested molecules (**Figure 2**).

It appears that both predictive and experimental factors are useful in predicting validation, and that the combination does even better. To understand how independent from each other the predictive and experimental fitness ranks are, we plotted them against each other (**Figure 3**). It is clear that there is substantial complementarity between the two fitness ranks. For example, there are 8 validated molecules identified by predictive log rank >3 (horizontal dashed line), and 9 validated molecules by experimental log rank >2 (vertical dashed line), but 13 molecules can be identified if both criteria are used together.

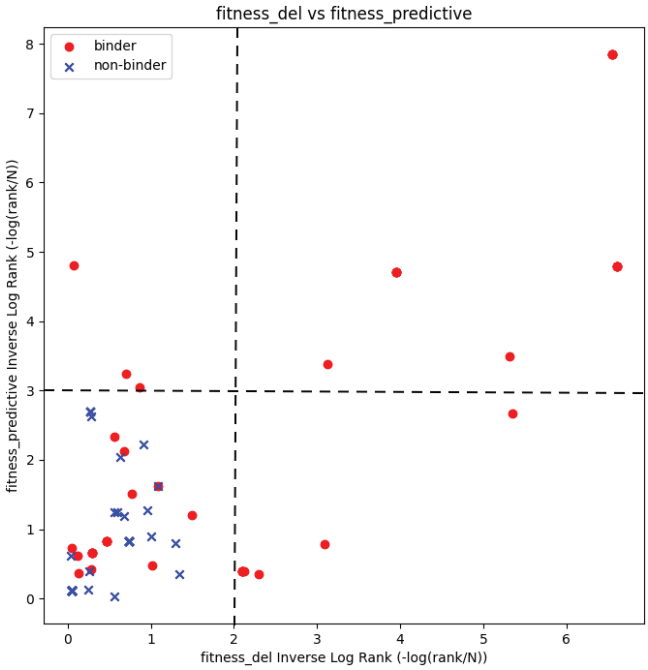

Figure 3. Comparison of fitness ranks between experimental evidence provided by DEL technology and computational predictions of Cyclica's MatchMaker™ technology.

Conclusions

This study aims to provide a better understanding of the strengths and potential synergies between the well-established DEL technology and emerging

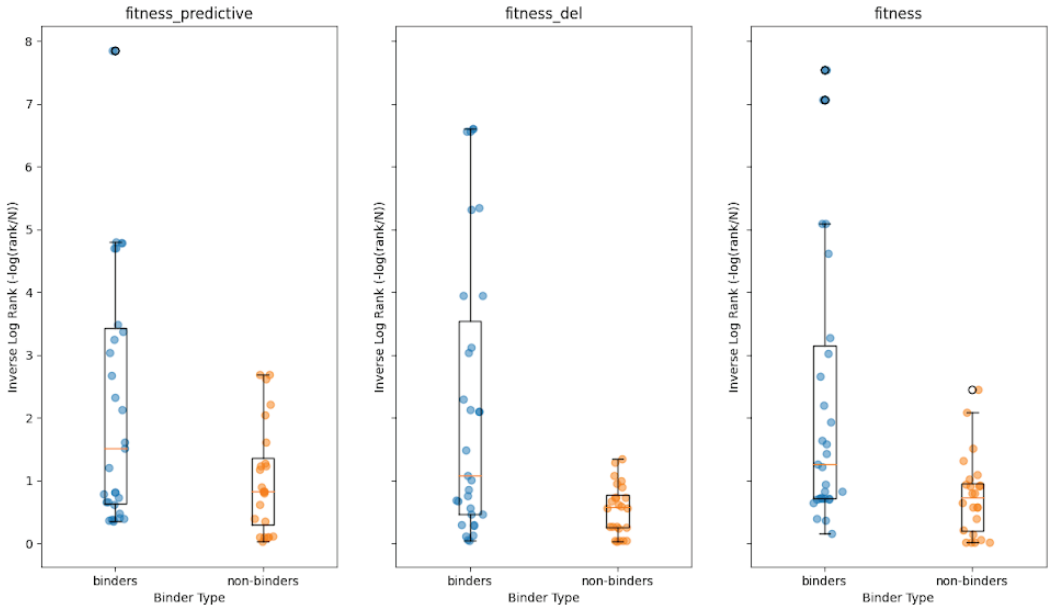

Figure 2. Distribution of fitness scores for molecules, including both binders and non-binders, using "fitness_del", "fitness_predict" and "fitness" as three different pareto optimal based fitness scoring calculation methods. ROC-AUC values for predictivity are 0.67, 0.63, and 0.68, respectively.

computational methods like MatchMaker™. Our findings indicate that combining the experimental DEL data with the computational predictions from Cyclica's MatchMaker™ technology can offer enhanced predictivity. The synergy between these methods could lead to more accurate predictions than relying on each method individually, providing an opportunity to improve the drug discovery process. One limitation of the current study is the retrospective nature of our analysis, which does not allow us to estimate realistic hit rates. In addition to the on-DNA validation data, we also plan to obtain results from off-DNA resynthesis for a different target in a future analysis. To achieve that, we plan to select molecules using the predictive methods and submit them for off-DNA resynthesis and assay.

References

1. Andrew E Brereton et al. (2020) “Predicting drug properties with parameter-free machine learning: pareto-optimal embedded modeling (POEM)” Mach. Learn.: Sci. Technol. 1 025008