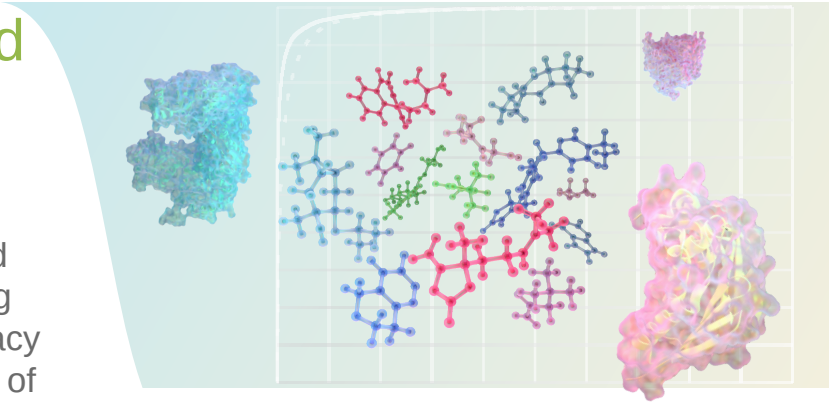
CYCLICA

MatchMaker: A Leap Forward in Proteome Screening Beyond Molecular Docking



MatchMaker combines molecular biophysics and deep learning (DL) to predict binding of new drug molecules to all proteins with high speed, accuracy and generalizability, moving beyond the reliance of molecular docking.

OPPORTUNITY

Drug/target interaction databases and 3D structure databases offer complementary advantages in machine learning (ML)-based drug-target prediction algorithms

TECHNOLOGY

MatchMaker

Ligand Express Proteome Screening

SOLUTION

Cyclica has pioneered a novel DL strategy for DTI predictions, that generalizes to new molecules with exceptionally high accuracy, distinguishing binders from non-binders with 98.3% accuracy.

INTRODUCTION

Fast, accurate, and generalizable predictions of drug-target interactions (DTI) have the potential to transform pre-clinical stages of drug discovery. In living systems, a drug's efficacy, polypharmacology, toxicity, and side effects are mediated through interactions with tens to hundreds of proteins found in the human proteome. The drive to profile the holistic impact of a drug on cellular systems has inspired numerous in silico and experimental approaches, each with their own strengths and weakness. While computational approaches offer undeniable speed and cost advantage over experimental techniques, accuracy and generalizability have historically challenged in silico methodologies.

Advanced ML approaches such as deep learning are increasingly popular in drug discovery, as they can generate accurate models that leverage large quantities of highly dimensional data. However, not all models are truly generalizable and fail to live up to their reported accuracies in real world settings. In particular, DTI predictions are particularly prone to overestimation as a consequence of compound series bias, ie. the presence of multiple similar ligands in a DTI database interacting with the same protein[1].

To address the need for accurate and generalizable DTI prediction model, we have developed MatchMaker, a deep learning algorithm that synthetically augments the millions of known DTIs found in public databases with biophysical information from 3D structure databases. Through stringent testing, we demonstrate accuracy and generalizability that outperform leading edge computational technologies, even outperforming experimental methods such as thermal proteome profiling (TPP).

METHODOLOGY

Ranking test set: We randomly selected 100 molecules (the "ranking test set") from the DTI database STITCH 5.0 in which each molecule has at least one very high confidence target, i.e. with a confidence score of one. Most molecules had only one high confidence target, but there were seven with two, one with four and one with five targets, for a total of 114 known targets.

Model training: We trained a MatchMaker model with our currently best known DTI data set and hyperparameters. The DTI data was expanded with 19 negative drug/target pairs for each positive, generated by pairing the target with randomly chosen molecules with no evidence of interaction. To avoid compound series bias[1], we excluded all molecules from the DTI data that were similar to any of the ones selected for the ranking test set. Molecules were considered similar when their Tanimoto Similarity (TS) based on Morgan 3 fingerprints exceeded 0.5. The DTI set was then further divided randomly into a training and validation set, using a modified approach to cluster-cross-validation[2]. Briefly, the partition was done along the lines of Tanimoto clusters of 0.75 similarity, again to avoid series bias. The MatchMaker model was trained on the training set, and training was monitored using the validation set.

Individual binding vs. non-binding discrimination: The MatchMaker model is trained with positive and negative cases to discriminate between drug/target combinations that bind in reality (as represented in DTI databases) and those that don't. We measured the performance of the model by predicting binding for all pairs in the validation set and comparing those predictions with the known labels. Since those labels were excluded from model training, the ability of the model to predict them represents its ability to generalize beyond its training data.

Rank of known targets: Since the purpose of MatchMaker is primarily in proteome screening, we also evaluated the ability of the model to rank all considered proteins such that true targets are found near the top of the list. For this purpose, we computed MatchMaker predictions for all 100 molecules of the ranking test set against all 8717 considered human proteins, noting the ranks of all 114 high confidence targets in their respective lists.

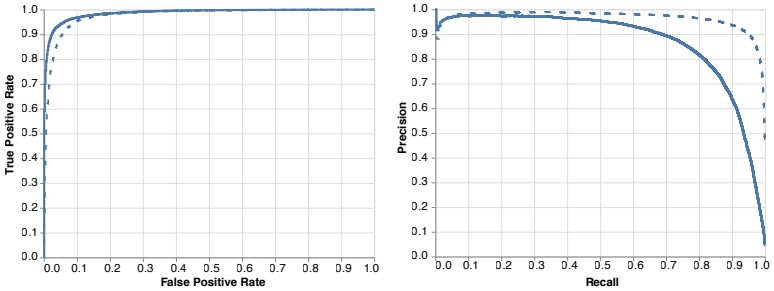


Figure 1. ROC (A) and precision recall (B) curves for discrimination of binding vs. non-binding pairs in the validation set of 378,158 pairs. Approximately 5% of the pairs were labeled positive. The area under curve (AUC) is 0.986 for ROC and 0.87 for PR. The dashed lines represent the balanced model, with 1:1 positives to negatives in training and validation data.

RESULTS

Individual Binding vs. non-binding discrimination: MatchMaker achieves an accuracy of 98.3% in discriminating interacting from non-interacting pairs in the validation set. **Figure 1** shows the Receiver Operating Characteristic (ROC) and the Precision/Recall (PR) curves for the discriminator, the areas under the curve (AUC) are 0.986 and 0.87, respectively. The validation set consisted of 378,158 pairs, of which approximately 5% were labeled positive. We also computed a balanced model, which achieves an accuracy of 92.6% on validation data with 1:1 positives to negatives. We focus on the 1:19 model here because it is superior in ranking known targets.

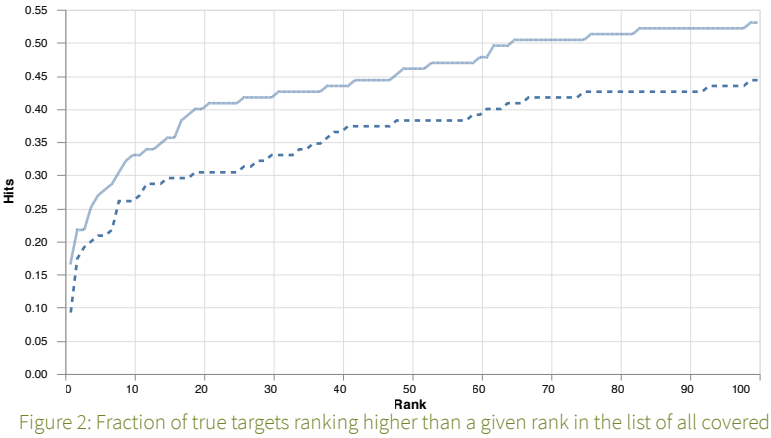


Figure 2: Fraction of true targets ranking higher than a given rank in the list of all covered proteins (8717 for the present model). The area under the full curve (area under accumulation curve (AUAC)), is a measure sometimes used for ranking accuracy and computes to 0.911. The dashed line represents the balanced model.

Ranking of known targets: Figure 2 shows the cumulative fraction of true positives amongst all 114 actual positives in the ranking test set. One third of all positives are detected within the top 10 (0.12%), and one half within the top 64 (0.73%). **Table 1** quantifies predictive power in proteome screening relative to Cyclica's first generation Ligand Express®, Proteome Screening technology, which used a unique combination surface matching technology to identify potential binding sites and molecular docking to rank proteins. This first-generation technology was designed with a focus on generalizability and has historically performed well on blind tests, but partial reliance on molecular docking led to high computational demand and limited prediction accuracy. Matchmaker-powered Proteome Screening significantly outperforms its predecessor in ranking known targets, particularly in the top 0.1% (top 9 proteins).

Table 1: Comparison of ranking performance between traditional Ligand Express based on molecular docking and the MatchMaker deep learning method. Shown is the enrichment factor EFx for five different x, which is defined as the number of true positives in the top x ranks (as plotted in Figure 2) divided by the number expected for a random ranking (114·x, here).

Top %	LE with Docking	MatchMaker
0.1	64.3	321.7
0.5	45.0	88.7
1.0	31.4	52.2
5.0	9.3	12.9
10.0	6.3	7.4

Comparative Benchmarking: There are no standardized benchmarks or community challenges available yet to evaluate accuracy and generalizability of proteome-scale DTI predictions. We instead compared MatchMaker directly to Secure DTI, a leading peer-reviewed technology for DTI predictions published by Hie et al. Oct 2018 in Science[3]. The analogous technology trains deep learning models on STITCH DB and uses global protein features derived from domain composition. In evaluating their models however, Secure DTI only excludes training DTIs representing exact ligand matches to testing DTIs (TS=1.0). In other words, isomers, and structural analogs of training molecules may be present in their testing set.

Table 2: Comparison of individual binding prediction performance between Secure DTI[3] and MatchMaker. Shown for both methods are areas under ROC and PR curves, and in addition for MatchMaker, target ranking AUAC values

METHOD	ROC-AUC	PR-AUC	RANKING AUAC
Secure DTI (1:1 neg)	0.95	0.95	
MatchMaker (1:1 neg)	0.975	0.966	0.877
MatchMaker (19:1 neg)	0.986	0.87	0.911

In contrast, we have applied very stringent filters to exclude all training DTIs with ligands remotely similar to those found in testing DTIs (TS>=0.5). This ensures that reported accuracies are representative of real-world applications, where new lead scaffolds are characterized in preclinical stages of pharmaceutical R&D. Despite applying a more realistic and challenging testing criteria, MatchMaker outperforms Secure DTI when retrained with an analogous 1:1 positive-to-negative training data ratio (**Table 2**).

SUMMARY

MatchMaker exhibits unprecedented accuracy in predicting ligand binding for individual drug/target pairs (98.3%), and substantially improves proteome screening accuracy relative to Cyclica's first-generation technology. MatchMaker can screen millions of molecules against the entire human proteome and has the capacity to further augment accuracy by merging public and private sources of DTI data and 3D structures of protein-ligand complexes. By synthetically augmenting DTI data with biophysical information, MatchMaker provides fast, accurate, and generalizable DTI predictions for proteome-scale applications.

RESOURCES

1. Mayr A., *et al.* Large-scale comparison of machine learning methods for drug target prediction on ChEMBL. *Chemical Science.* (2018)

2. Mayr A., *et al.* DeepTox: Toxicity Prediction using Deep Learning. Front. *Environ. Sci.* (2016)

3. Hie, B., *et al.* Realizing private and practical pharmacological collaboration. *Science.* (2018)

CYCLICA INC.

18 King St East, Suite 801
Toronto, Ontario, M5C 1C4, Canada
1-416-304-9201