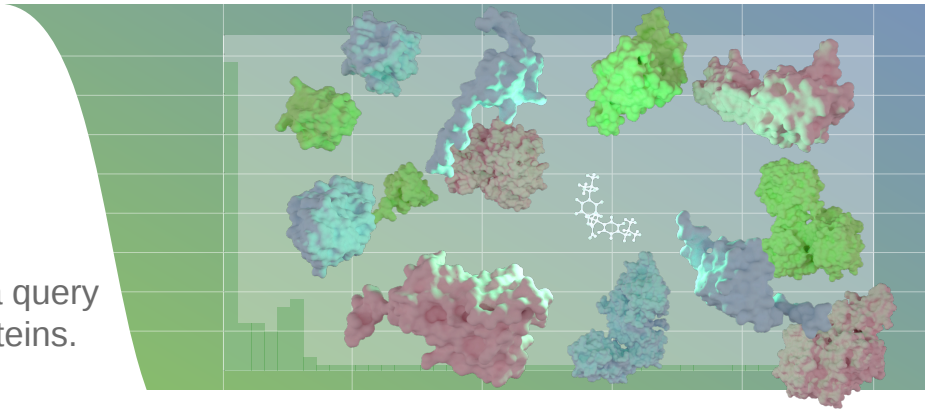
CYCLICA

Assessment of Drug Target Identification and Ranking using PROBE$_X$

Cyclica's PROBE$_X$ proteome docking tool accurately predicts interactions between a query drug and all structurally-characterized proteins.



PROBLEM

Mapping a ligand's interactions with diverse proteins experimentally can be expensive, while existing knowledge-based *in silico* solutions are dependent on prior biochemical knowledge.

TECHNOLOGY

Ligand Express™ tools: PROBE$_X$ Proteome Docking

SOLUTION

PROBE$_X$ proteome docking predicts protein interactions for small molecules. The software's success in ranking previously known drug-target interactions for 53 recently-analyzed compounds is reported herein.

INTRODUCTION

Mapping protein interactions to a drug of interest can elucidate the compound's mechanism of action for beneficial or adverse pharmacologies and/or discover its other novel uses. A recent resurgence of phenotypic assays in pharmaceutical discovery fuels a growing need for computational mechanism of action predictions[1]. Experimental approaches to explore a drug's possible target(s) are not universally applicable. For example, drug target exploration via resistance selection is limited to toxic compounds like antibiotics or oncology therapeutics. Once plausible targets are identified through these experimental screens, biological targets are further validated by low-throughput assays, such as gene-specific knockdown experiments. Computational approaches can reduce the cost of experimental screening step(s) by providing immediate hypotheses for knockdown validation. Existing knowledge-based computational methods depend heavily on the availability of existing drug-target interactions, since their inferences are based on drug-drug and/or protein-protein similarity.

PROBE$_X$ is a hybrid computational procedure that integrates molecular docking and protein surface matching to intelligently sample the landscape of known proteins for potential interaction sites. In this approach, surface features of favorable docking sites guide the search to new sites on different proteins. The surface matching procedure is computationally fast, affording a complete scan of all structurally-characterized proteins. Protein surfaces sharing physico-chemical properties similar to docking hits are subsequently evaluated with more computationally expensive molecular docking simulations.

METHODOLOGY

We assessed the drug-target identification potential of PROBE$_X$ by reporting the observed ranks of known targets for 53 recently-analyzed compounds. For this study, we limited known targets to associations curated by DrugBank (March 2016) whereby the target protein has a publicly-available macromolecular structure representing at least 30% of the total protein sequence. Proteins subjected to molecular docking simulations by PROBE$_X$ were ranked according to their docking scores. Ranks of proteins that represent negative surface matching prediction were evenly distributed after those of docked hits.

We also compared ranks of known targets to the ranks of a randomized negative control. For each known target of the 53 analyzed compounds, fifty protein decoys were randomly sampled from the PROBE$_X$ structural library of 342,723 protein structure models representing 30,584 different UniProt entries. Randomized decoys are used to provide a baseline for the rank of known target (ROKT) plot in **Figure 1**, negative data to construct a Receiver Operating Characteristic (ROC) curve in **Figure 2**, and to compile enrichment factors for **Table 1**. Enrichment factors are given by the following equation:

$$EF_x = \frac{Known\ targets\ in\ the\ top\ x\%\ of\ ranks}{Decoys\ in\ the\ top\ x\%\ of\ ranks}$$

RESULTS

The distribution of known target ranks is provided in **Figure 1.** This figure illustrates that most known targets (77%) rank in the top 15%. The remaining 23%, shown as a flat bar, represent targets that were not selected by the surface matching algorithm for subsequent docking, and therefore represent false negatives. The figure also shows a prominent

Table 1. 26 Enrichment scores of PROBE$_X$-predicted drug-targets interactions relative to ranks of randomly selected targets from the PROBE$_X$ structure library.

Top X% of Predictions (%)	Enrichment Factor (EF$_X$)
0.1	64.3
0.5	45.0
1.0	31.4
5.0	9.3
10.0	6.3

peak in the first bar, representing a strong enrichment of targets among top-ranked protein structures. Detailed enrichment scores are also provided in **Table 1**, whereby we observe a 64-fold enrichment of protein targets in the top 0.1% (~30 proteins) of ranked proteins. The Receiver Operating Characteristic (ROC) curve in **Figure 2**, further demonstrates the predictive power of PROBE$_X$ relative to randomized decoys (negative data). In this figure, the steep rise in predictive power (left side of the curve) is evidence of strong enrichment of genuine targets among top ranked compounds. Proteins that are not selected for docking do not have a score, leading to a truncated ROC curve. In order to estimate the AUC, we extrapolated the curve with a straight line segment as shown in **Figure 2**.

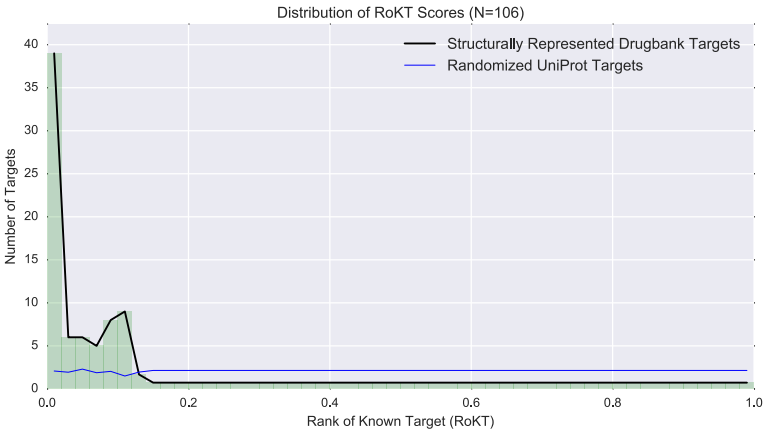


Figure 1. Histogram distribution of known target ranks for 53 compounds analyzed by PROBE$_X$, representing 106 drug-protein interactions (black line and green bars). A second line (blue) showcases the distribution of decoy proteins randomly-selected from the PROBE$_X$ structure library, normalized to match the area of known targets ranks.

SUMMARY

Cyclica's PROBE$_X$ technology offers a computational alternative to costly experimental techniques for screening investigational compounds in search of interacting proteins and polypharmacologies. Described thus far is the computational enrichment of putative targets through proteome docking alone. Further target enrichment can be achieved with Cyclica's DIVE$_X$ technology, by coupling PROBE$_X$ docking hits with additional biological information, such as phenotype, tissue type, etc. For an example, see Cyclica case study: "DIVE$_X$ Uncovers Potentially Fatal Drug-Protein Interactions", which demonstrates how systems biology information can supplement proteome docking predictions to shortlist potential molecular interactions responsible for the fatal, neurotoxic side effects of experimental drug BIA 10-2474.

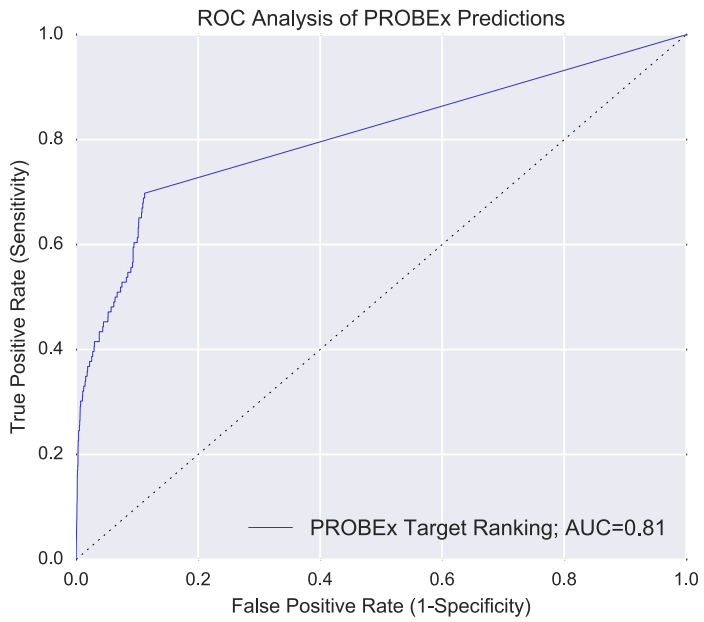


Figure 2. Receiver Operating Characteristic (ROC) curve demonstrating the sensitivity and specificity of PROBE$_X$ as an alternate representation of the data presented in figure 1. The ROC analysis compares the ranks of known drug-target interactions (positive data) relative to randomly selected decoy interactions (negative data). The dotted diagonal edge represents the line of no-discrimination.

In contrast to knowledge-based computational tools, the use of protein structure information and molecular docking allows PROBE$_X$ to identify new connections between drugs and protein without the need for prior drug-protein interaction data as a basis for inference. Typically, drug-protein interaction information is limited to well-characterized proteins, such as previously-studied disease targets. Eliminating this dependency allows PROBE$_X$ to screen a broader range of proteins, requiring only a 3D model structure. Recent efforts aimed at predicting drug-target interactions without biochemical data for similar molecules, report 63% sensitivity and 81% specificity[2]. In contrast, PROBE$_X$ reaches predictive rates of 70% sensitivity and 89% specificity (**Figure 2**).

RESOURCES

1. Wagner, B. K. & Schreiber, S. L. The Power of Sophisticated Phenotypic Screening and Modern Mechanism-of-Action Methods. *Cell Chem Biol.* **23**(1), 3-9 (2016).

2. Naveed H. et al. An integrated structure- and system-based framework to identify new targets of metabolites and known drugs. *Bioinformatics*. **31**(24), 3922-9, (2015).

CYCLICA INC.

18 King St East, Suite 801
Toronto, Ontario, M5H 1A1, Canada
1-416-304-9201

101 Main Street, Floor 14
Cambridge, Massachusetts,
02014-1519, United States