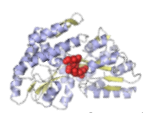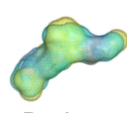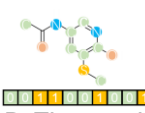# *In Silico* Drug Discovery Platform

In Silico Tools for Drug Discovery

Hit Finding	Hit to Lead	Lead Optimization

Virtual Screening (VS)
- Docking
- 3D Shape VS
- Pharmacophore VS
- 2D Fingerprint VS

Experimental + VS Combo
- DEL+ Machine Learning (ML)
- ASMS + ML
- DEL+ ASMS + ML

Molecule Design & Evaluation
- Molecule Generation
- Scaffold Based Molecule Generation
- Fragment Based Molecule Generation
- Docking
- Molecular Dynamics
- DEL_SAR

ADMET Optimization
- ADMET Prediction

Binding Estimation
- MM / GB (PB) SA

In Silico Services for Hit Finding

	Structure Based	Ligand Based		
VS	Co-crystal or AF2 Predicted Structure	3D shape based	Pharmacophore based VS	2D Fingerprint based
Diversity	5 Million	20 Million	5 Million	1 Billion
Libraries Source for VS		WuXi AppTec GalaXi (20 Million) Commercial Vendors (10 Billion)		
Service Model		FFS/Customization Services tailored to specific needs		
Timeline	4 to 5 weeks	3 to 4 weeks		
Note	Co-crystal / predicted structure needed	Query molecule needed		

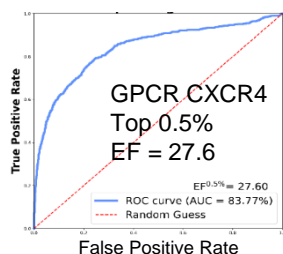|	DEL / ASMS Enhanced VS		
Experimental + VS Combo	DEL + ML	ASMS + ML	DEL + ASMS + ML
Diversity	10 Billion	270K	
Libraries Source for VS	GalaXi, Enamine, Mcule in-stock and virtual lib.	WuXi AppTec ASMS lib.	
Timeline (Wet + dry)	4 + 3 weeks	5 + 1 weeks	8 + 1 weeks
Note	DEL screening needed	30K ASMS screening needed	DEL + 30K ASMS screening needed

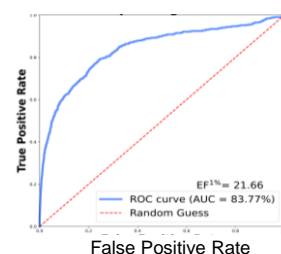# Representative Workflow & Case Studies for Hit Finding

SBVS

Cocrystal CXCR4
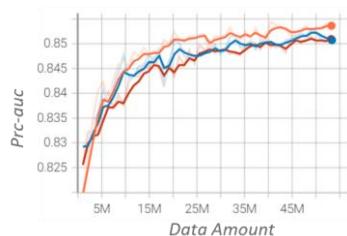
Receiver Operating Characteristic

GPCR CXCR4
Top 0.5%
EF = 27.6

$EF^{0.5\%} = 27.60$
ROC curve (AUC = 83.77%)
Random Guess

False Positive Rate

Receiver Operating Characteristic

$EF^{1\%} = 21.66$
ROC curve (AUC = 83.77%)
Random Guess

True Positive Rate

False Positive Rate

DEL + ML

DEL_ML enhanced VS

Model performance

Prc-auc

Data Amount

Docking evaluation

ML06: from this workflow;
CDD-1713: from patent; DEL01: from DEL

| DEL01 | 0.17 |
| CDD-1713 | 0.16 |

CDD-1713   DEL01

0   0.2   0.4   0.6   0.8   1
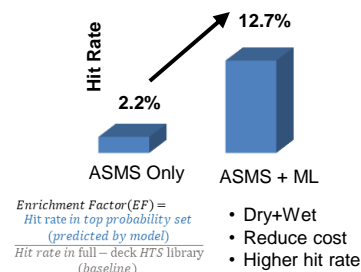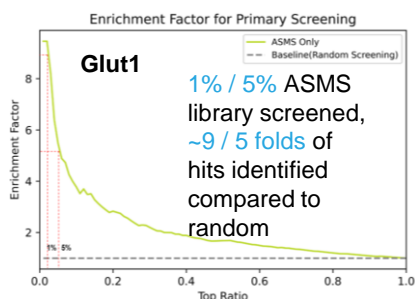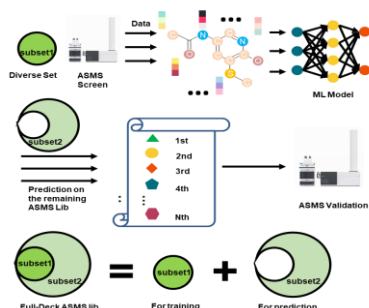Tanimoto Similarity of ML06 with 2 hits
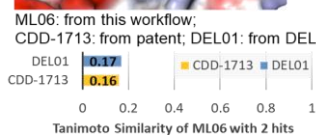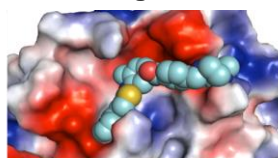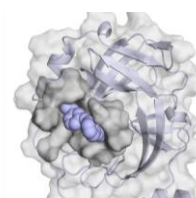
IC50 & Co-crystal
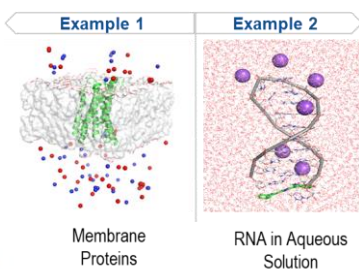
Co-crystal obtained (unpublished data)
Novel scaffold with patent space
LE > 0.28, suitable for further LO

ASMS+ML

Diverse Set — ASMS Screen — Data — ML Model

1st 2nd 3rd 4th Nth

Prediction on the remaining ASMS Lib — ASMS Validation

Full-Deck ASMS lib = subset1 + subset2
For training   For prediction

Enrichment Factor for Primary Screening

ASMS Only
Baseline(Random Screening)

Glut1

Enrichment Factor

1% 5%

Top Ratio

1% / 5% ASMS library screened, ~9 / 5 folds of hits identified compared to random

Hit Rate

12.7%

2.2%

ASMS Only   ASMS + ML

$$Enrichment\ Factor(EF) = \frac{Hit\ rate\ in\ top\ probability\ set\ (predicted\ by\ model)}{Hit\ rate\ in\ full-deck\ HTS\ library\ (baseline)}$$

• Dry+Wet
• Reduce cost
• Higher hit rate

Hit to Lead & Lead Optimization

Molecular Dynamics Platform

Example 1   Example 2

Membrane Proteins   RNA in Aqueous Solution

ADMET Prediction Model

Category	Prediction Metrics	Model	Category	Prediction Metrics	Model
Absorption	• Caco-2 cell permeability • MDCK cell permeability • PAMPA permeability • Kinetic solubility • Thermodynamic solubility	• XGBoost • Random Forest	Metabolism	• CYP450	• Deep Learning
			Excretion	• Clearance_Hepatocyte • Clearance _Microsome • T1/2	• Deep Learning
Distribution	• BBB Score	• Polynomial Regression	Toxicity	• hERG Blocker • Tox21 (ER, AR pathway)	• Deep Learning