

USING MACHINE LEARNING FOR DRUG DISCOVERY

By Ashray Shetty



Problem with drug discovery?

- Drugs can bind to more than one target
- Sometimes the connection can be weak
- Only around **10%** of candidate drugs are approved
- Half of drugs rejected due to lack of **effectiveness**



How to identify drug's target?

Based on structure

- very complex
- very expensive
- cannot be automated



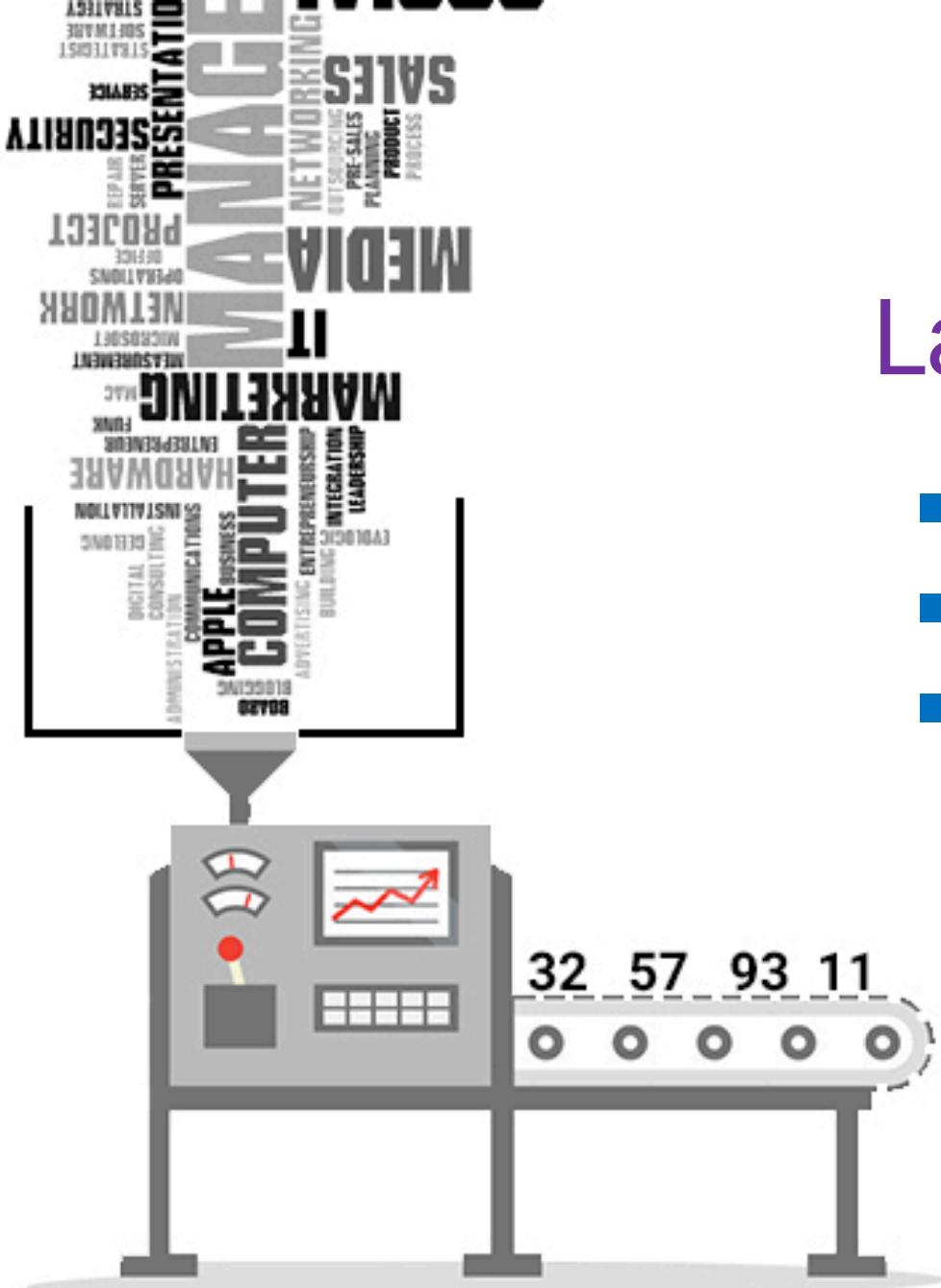
Using chemistry

- Automatized
- but expensive
- and time-consuming



Can we use computational approach???



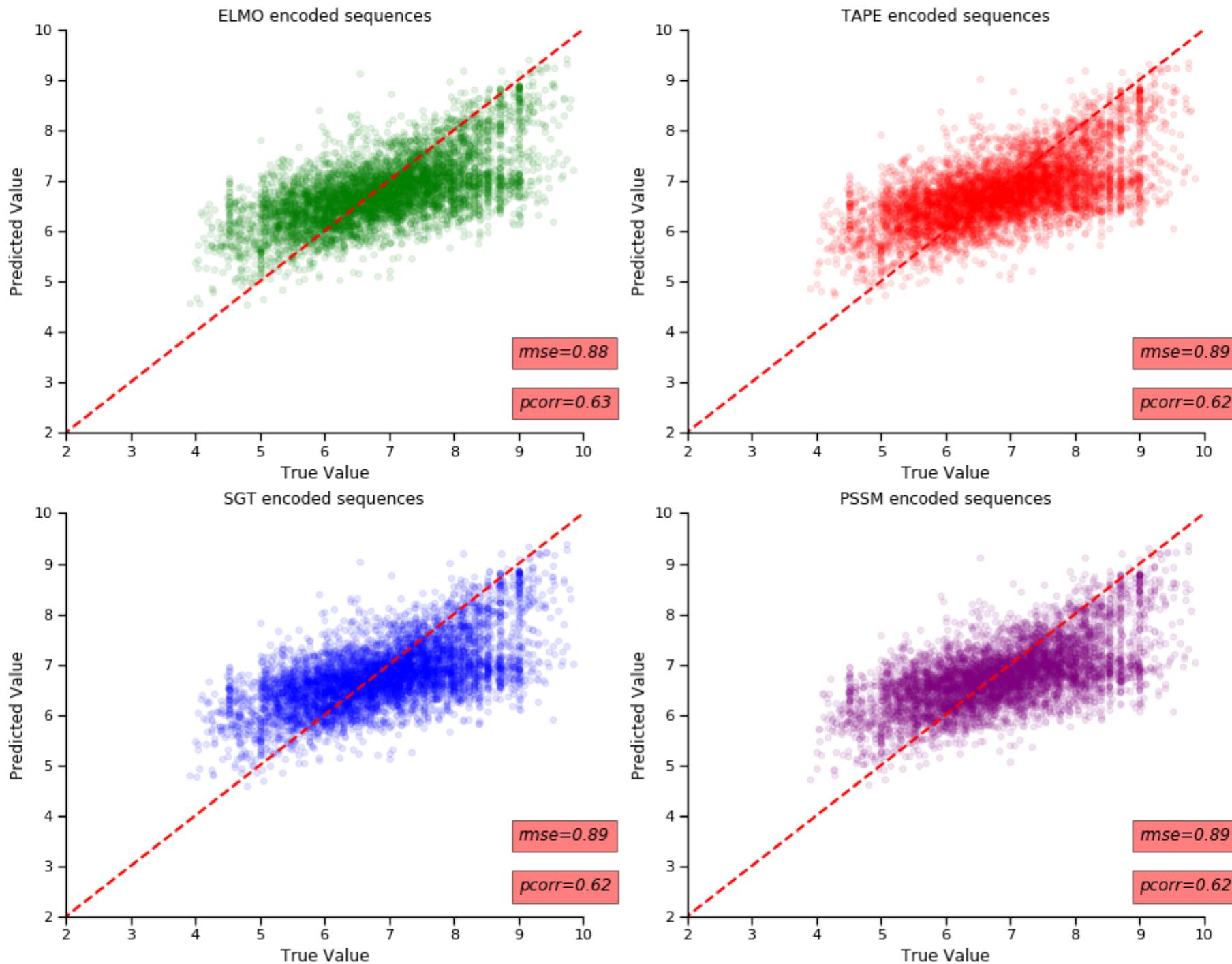


Language processing

- Turn letters to numbers
- Extract hidden meaning
- Useful for machine learning

Project overview:

- Dataset containing 89000 drug-target binding value
- Drugs have potential use as anticancer medication
- Used 4 different language processing methods for representing proteins



Drug Screening

Drugs

CHEMBL1085626	CHEMBL3670408
CHEMBL1761502	CHEMBL1099296
CHEMBL3690390	CHEMBL3421636
CHEMBL106992	CHEMBL471520
CHEMBL2386817	CHEMBL1835747
CHEMBL214329	CHEMBL3286814
CHEMBL115769	CHEMBL3329399
CHEMBL272938	CHEMBL3642385
CHEMBL3948663	CHEMBL2349008
CHEMBL1078721	CHEMBL1762534

Targets

P35968	P49841
Q16539	P52333
Q05397	P12931
P43405	Q96GD4
060674	P36888
P10721	P06239
014965	P11802

Targets

	P35968	Q16539	Q05397	P43405	O60674	P10721	O14965	P49841	P52333	P12931	Q96GD4	P36888	P06239	P11802	
Drugs	CHEMBL1085626	6.4	6.6	5.9	7	7.1	6.8	7.3	6.5	6.7	6.7	7.4	7.2	6.7	6.7
CHEMBL1761502	6.2	6.8	5.7	6.6	6.7	6.4	6.2	6.2	6.6	5.5	6.4	6.3	5.5	6.2	
CHEMBL3690390	6.6	7.1	5.6	7	6.8	6.9	6.6	6.7	6.7	6	6.6	7	6.3	6.9	
CHEMBL106992	6.5	7.1	6.4	7	7	6.9	6.5	6.7	6.5	6.9	6.3	6.8	7	6.5	
CHEMBL2386817	6.9	6.8	5.8	7.1	6.9	7	6.6	6.8	6.4	6.3	6.8	7.2	6.6	6.6	
CHEMBL214329	6.3	6.9	6.6	7.2	7.4	6.8	7.3	6.4	6.8	6.8	6.7	7.5	6.7	7.6	
CHEMBL115769	6.2	6.9	5.3	6.8	6.5	6.6	6	6.3	6.4	5.6	6.4	6.6	6.2	6	
CHEMBL272938	6.8	6.8	5.7	6.8	6.4	7.1	6.6	6.5	6.7	6.1	7	7.1	6.2	6.5	
CHEMBL3948663	6.7	6.8	5.9	7.3	7.2	6.7	6.9	6.6	6.6	6.6	6.8	7.1	6.6	6.8	
CHEMBL1078721	6.4	6.3	6	7.3	7.3	6.9	6.9	6.6	6.5	6.6	6.7	7.3	6.6	6.8	
CHEMBL3670408	6.7	6.8	5.7	7	6.8	6.5	5.6	6.4	6.9	5.6	7.3	6.6	6	6.6	
CHEMBL1099296	6.5	6.8	5.6	6.7	6.7	6.5	6.2	6.2	6.3	5.6	6.7	6.8	6.2	6.4	
CHEMBL3421636	6.3	6.8	5.7	6.8	6.6	6.5	6.4	6.6	6.3	6.5	6.4	6.8	6.5	5.9	
CHEMBL471520	6.6	6.5	6.8	6.6	6.8	6.8	6.7	6.6	6.7	6.5	6.6	6.7	6.7	6.5	
CHEMBL1835747	6.5	6.5	5.7	6.8	6.8	7.2	6.4	6.3	6.4	6.5	6.5	7.9	6.7	6.4	
CHEMBL3286814	6.4	6.8	6.3	6.8	7.1	6.6	6.4	6.6	7	6.3	6.5	6.5	6.4	6.7	
CHEMBL3329399	6.8	7	5.7	6.7	7	6.7	6.3	6.2	6.3	6	6.7	6.9	6.5	6.1	
CHEMBL3642385	6.6	6.4	5.6	6.7	7.2	6.6	6.6	6.5	6.1	6.8	6.7	7	6.9	6.6	
CHEMBL2349008	7.2	6.9	5.7	6.8	6.8	7	6.4	6.7	6.3	6	6.8	7	6.7	6.3	
CHEMBL1762534	6.7	6.7	5.9	6.7	7.1	6.7	6.5	6.6	6.4	6.7	6.5	6.9	6.6	6.4	

Benefits of using ML based screening

- Organizing workflow during drug screening
- Allows us to fill the blank in our understanding of drug target activity space
- Leading to saving time and cost associated with testing
- Better able to detect and avoid side effects associated with drug

Thank you