The Screening Process

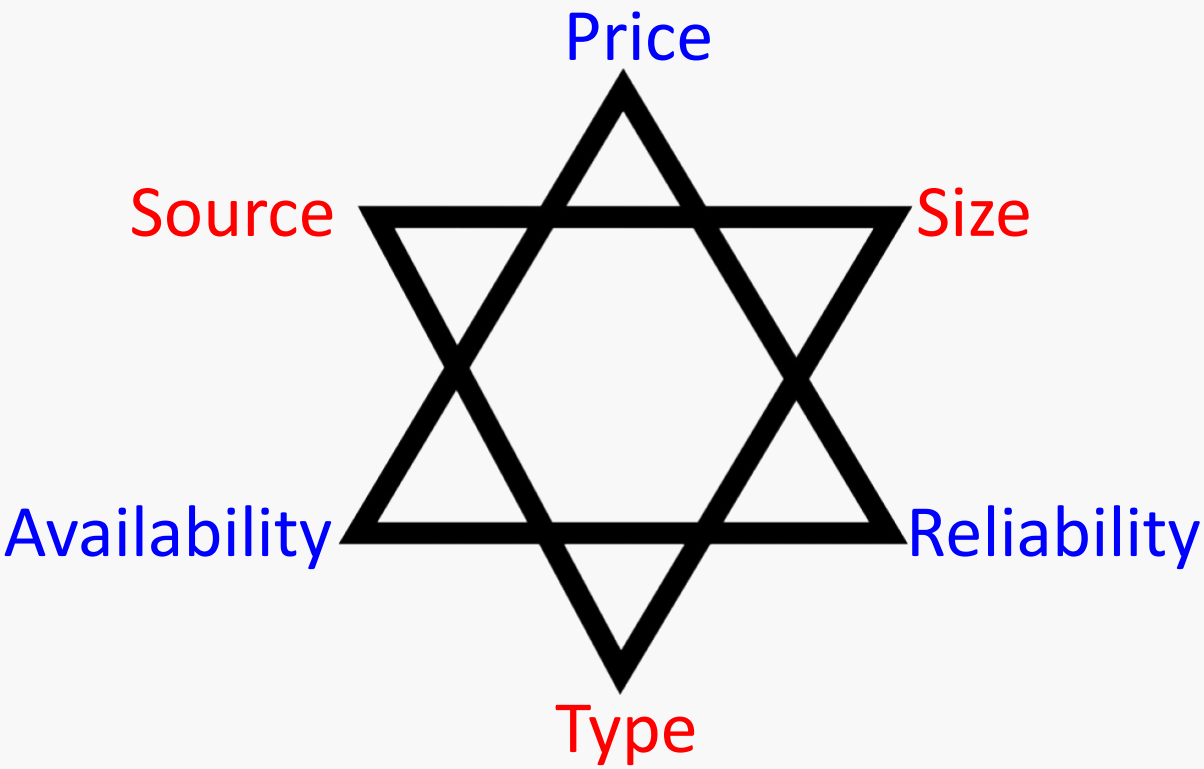
Hanoch Senderowitz Bar-Ilan University, Israel

1st MuTaLig COST Action Meeting, Lugano, July 2016

The screening library

A good screening library is critical to success

What makes for a good screening library (for a given project)?



Source

- In-house compounds
  - Available
  - Proprietary
  - Relevant (?)
- Commercially available compounds
  - Easily accessible
  - Screened by everyone
  - Not patentable
- Synthetically feasible compounds
  - Proprietary (?)
  - Focused
  - Novel
  - Require synthesis

- Cherry Picking
  - Biased screening
  - Specific compounds from different vendors
- Whole library
  - Unbiased screening
  - “Whole” library from a single vendor

Size

Is screening a game of numbers?

Table 1 Some popular estimations of the chemical space size

Size reduction mechanism is needed!!!

			stereoisomers		
1.3×10^{15}	≤38 atoms*	C, H	Acyclic stereoisomeric alkanes	Exhaustive enumeration	Blair and Henze [5]
10^{21}	<7 Å	40 functional	Neurological drugs	Combinatorial enumeration	Weaver and Weaver [8]

Mechanism Depends on knowledge

10^{33}	≤750 Da	C, N, O, F	Heptanes and hexanes including stereoisomers	Combinatorial enumeration	Weininger [23]
10^{33}	≤36 atoms, ≤500 Da	C, N, O, S, Hal	Stable compounds (stereoisomers are not taken into account)	Learning of exhaustively enumerated structures from	This work

The more we know, the better we can navigate through the chemical space

Gorse [27])

* The greatest number of compounds that is mentioned in the source

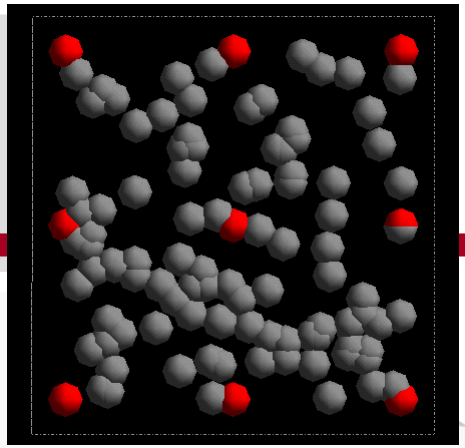
Polishchuk et al., J Comput Aided Mol Des (2013) 27:675–679

When we don't know anything

- Functional screening
- Phenotypic screening

Vendor	Library	# compounds
Multiple	ZINC	> 95,000,000
Multiple	emolecules	> 7,000,000
Enamine	HTS Collection	1,700,000

Diversity



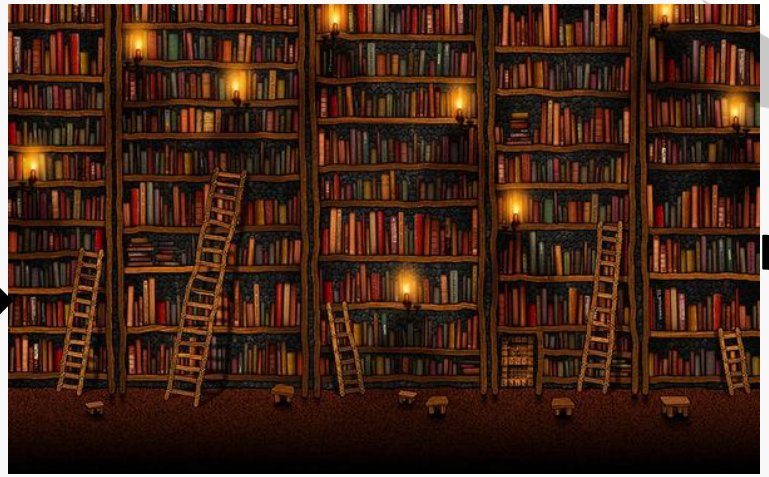
Vendor	Library	# compounds
Asinex	Gold & Platinum collection	292779
TimTec	ActiGlobe-50K	50000
ChemBridge	DIVERSet™-EXP DIVERSet™-CL	50000 50000
Maybridge	Screening Collection	55000
Enamine	Premium Collection	93 600
Sigma	MyriaScreen Diversity Collection	10000
ChemDiv	STOCK DIVERSITY COLLECTION	1500000

Could optimize-able hits be always obtained from a “master” library?

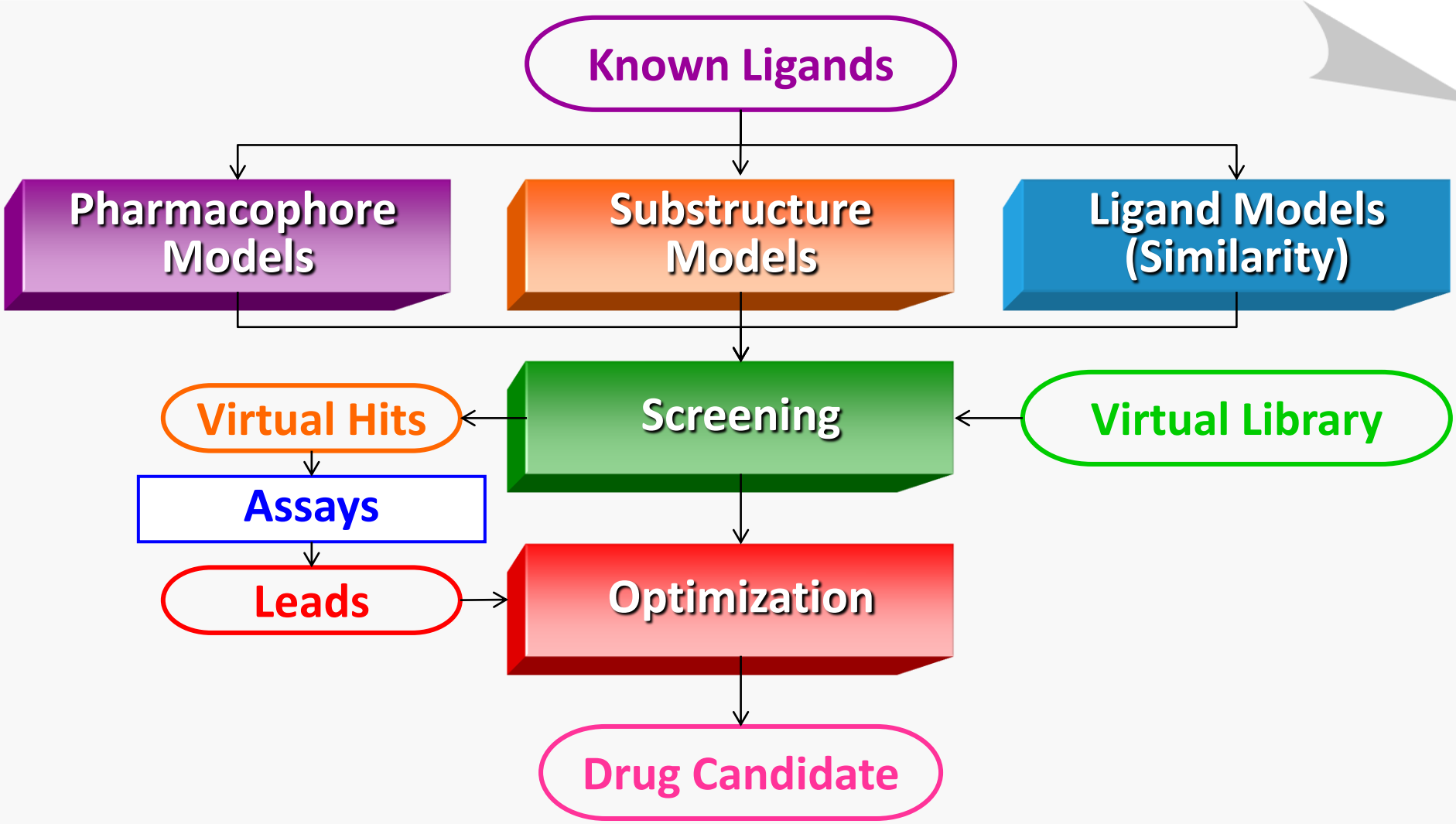
When we only know the target

Library	# compounds	Library	# compounds	Library	# compounds
Adenosine Receptors Targeted Library	21,957	Eccentric PPI Library	6,875	P2RX7 Antagonists Library	13,108
AgroChemical Library	55,436	Ephrin 4B library	7,906	PDZ PPI Library	4,586
Akt Targeted Library	12,328	Epigenetics Library	30,867	Peptidomimetic Library	13,973
Allosteric Kinases Inhibitors Library	26,615	Frequent Hitters Library	9,450	PI3K Targeted Library	19,898
Anti infective Library	8,675	G9a Inhibitors Library	13,132	Phosphatase Inhibitors	15,052
Anti bacterial/Anti viral Library	5,512	Glucocorticoid receptors Library	5,539	PKM2 Analogs	435
Anti fungal Library	6,278	GP 120 & GP 41 Libraries	19,974	PKM2 Modulators	8,403
Antimitotic Library	10,667	GSK3ß Targeted Library	4,896	Polymerase Library	5,771
Autophagy Targeted Library	17,687	HA2 Library	4,163	PPI CDI Library	142,000
Apoptotic Library	54,229	HDAC Library	20,413	Protein Kinases Target Platform Library	32,062
Aurora A B Kinases Library	10,360	Hedge Hog Pathway PPI Library	11,281	PPI Helix Turn Mimetics Library	21,558
Bcl2 Bax PPI inhibitors Library	26,279	Hsp90 Targeted Library	13,689	PRMT Library	32,049
Bcl2 PPI Inhibitors Library	11,188	h TERT Targeted Library	49,578	Proline Kinase Library	2,376
Beta 2 Adrenoligands Library	20,937	Indole Derivatives	11,948	Purinergic Library	3,732
Library of Small Molecule Inhibitors of beta Catenin Signaling	9,092	Ion Channels Target Platform Library	14,926	Quiescent Cancer Cell Pathways Set	25,874
Beyond the Flatland Library	58,698	TK Targeted Library	32,062	RAR (Nuclear receptors) Ligands Library	7,981
Bradykinin Library	18,574	KRAS Targeted Library	11,044	Recognition Elements PPI Library	27,152
Bromodomains Library	6,114	Ligand Gated Ion Channels Library	4,166	SH2 Library	14,111
Calcium Channels Library	10,638	Macl GPIb alpha Interaction Library	28,135	SH2 PTB Focused Library	7,333
Cancer Stem Cells Targeted Library 6	19,95	Matrix Metalloproteinases Targeted Library	9,017	Shape Helix Mimetics PPI Library	9,454
CB1 2 Library	17,185	MDM2 PPI Library	7,144	SmartTM Library	54,803
Chemokines Library2	20,84	MDM2 p53 interaction inhibitors Library	6,799	Soluble Diversity Library	9,624
CMet Library	16,421	MDM2 p53 PPI inhibitors targeted Library	18,274	Serine Proteases Inhibitors Library	38,233
CNS Targeted Library	32,313	MEF2 HDAC (class II) Modulators Library	6,058	Sulfotransferase Library	90,813
CXCR4 Targeted Library	11,248	Methyltrasferase Library	11,647	Targeted Diversity Library	46,817
CNS BBB Library	26,490	Modulators of Protein Protein Interactions (PPI) Library	127,936	TLR 8 ligands Library	844
Cysteine Proteases Inhibitors Library	8,602	Monoamine Transporters Library	7,990	Type II Kinase Inhibitors Library	8,302
Fragments Library	15,034	Na+ Channels Blockers/Antagonists Set	60,247	VEGFR Inhibitors Library	43,860
Developmental Pathway (Hh/Wnt) Set	2,413	NFkb Regulators Library	9,447	P24 Targeted Library	12,516
Cyclic Ugi PPI Library	10,582	Nonpeptide Peptidomimetics PPI library	22,380	Launched & Clinically Evaluated Drugs Library	266
DNMT Focused Library	38,769	NR Focused Library	1,760		

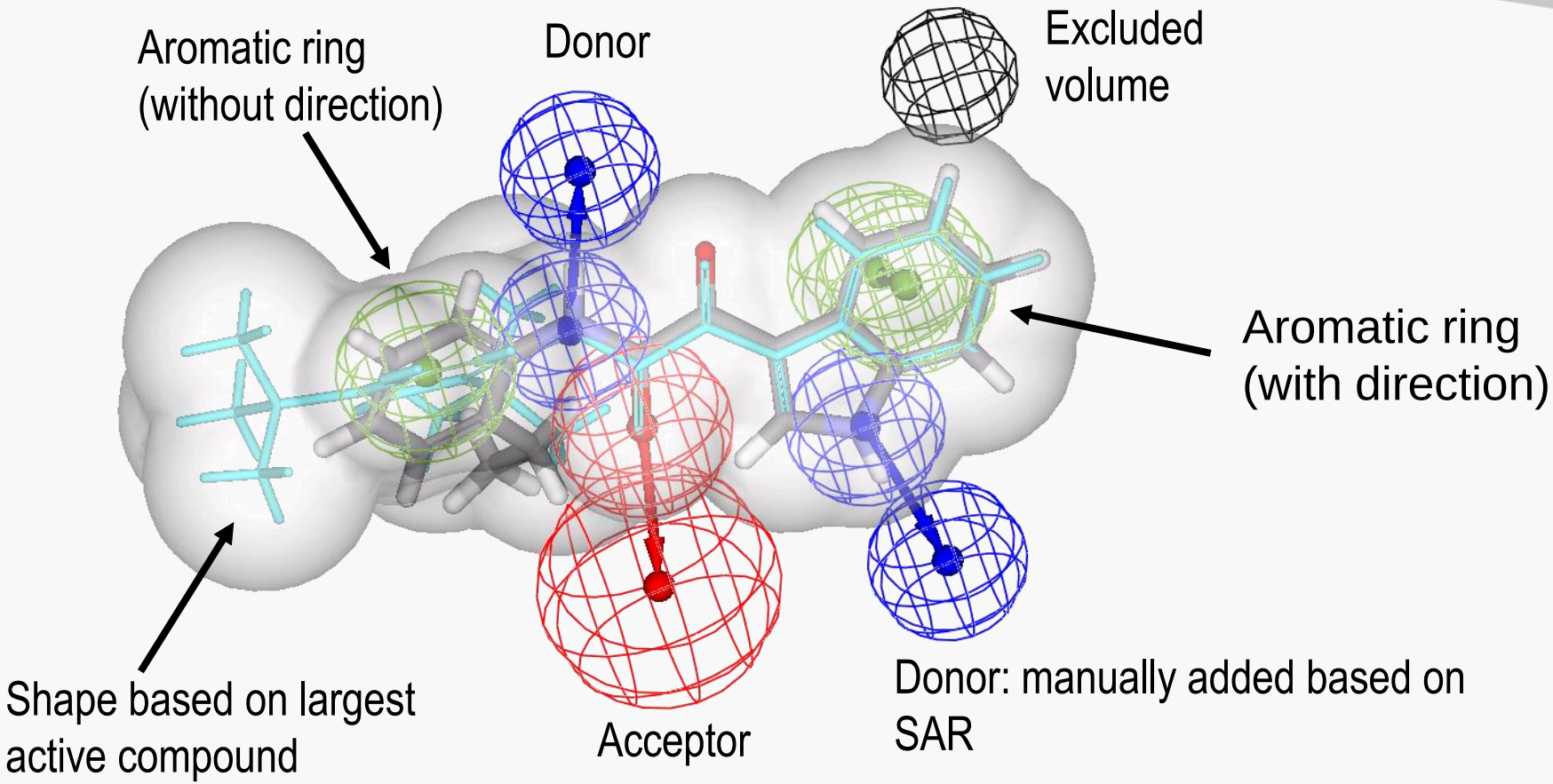
Virtual screening: Smart navigation through chemical space



When we know the ligands

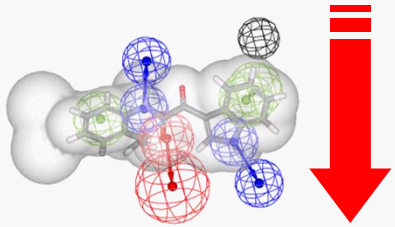


Pharmacophore: Example



Validate hypothesis

Sample library + reference compounds



Screen and score
library with hypothesis

Ranked list of compounds

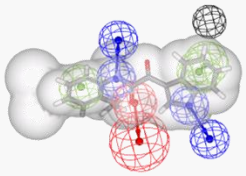
The hypothesis is considered valid if the known actives are highly ranked compared to the library compounds

A score cutoff was selected such that:

1. All weakly active compounds are below the cutoff
2. All Medium-highly active compounds are above the cutoff
3. Only 0.35% of the screening library are above the cutoff

Screen and rank library bases on hypothesis

738,500



Apply cutoff

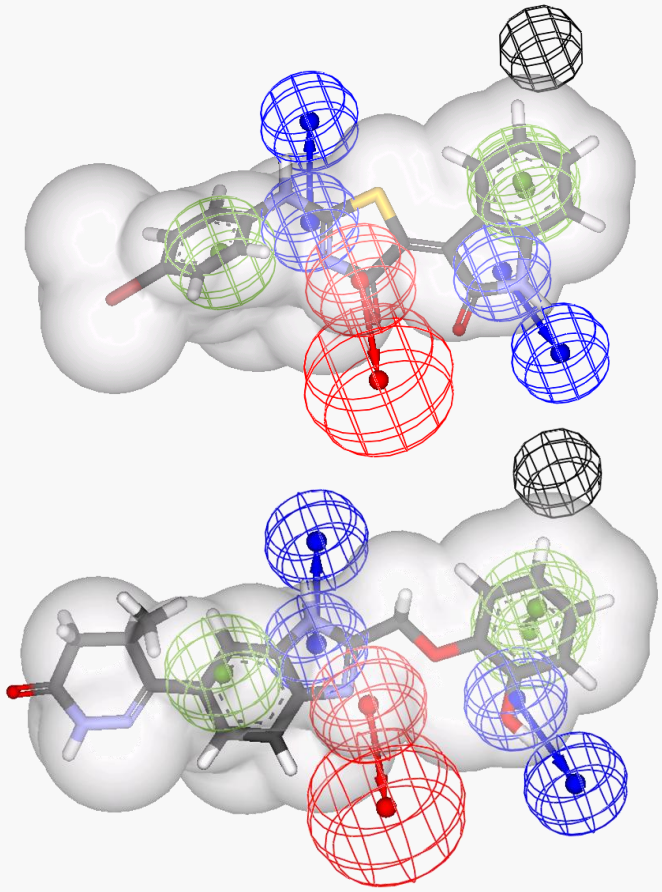
2,700

Similarity-Based Clustering

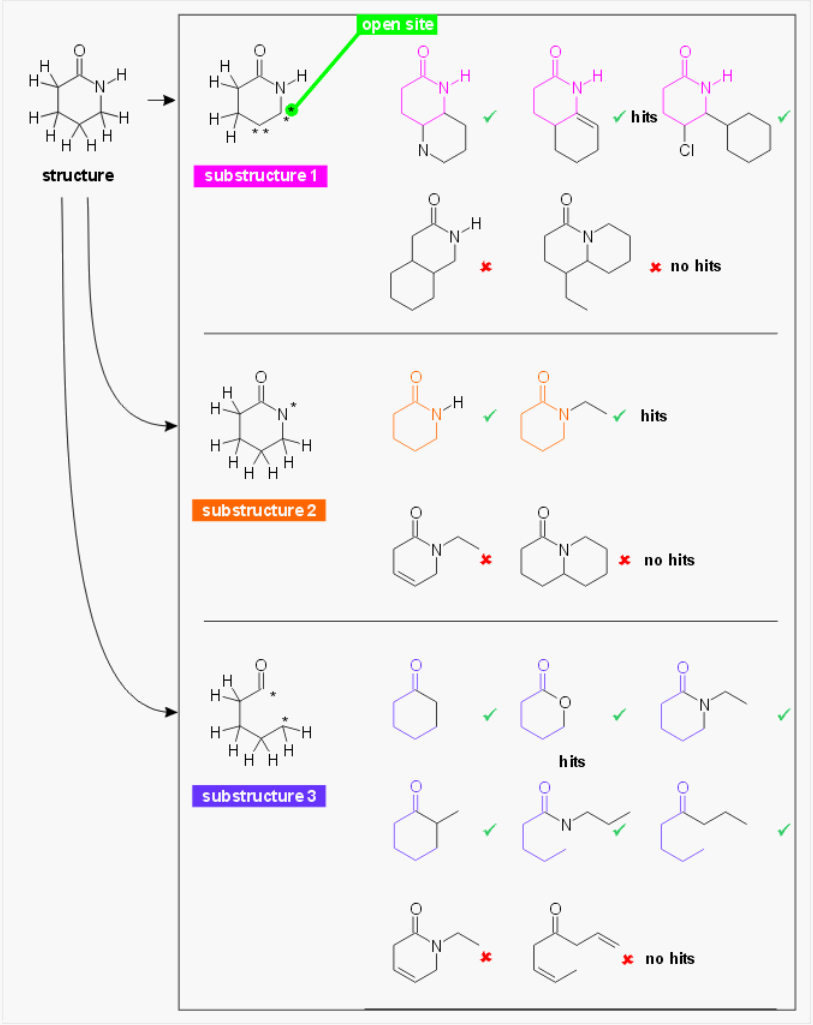
Visual inspection of cluster representatives and other members of “interesting” clusters

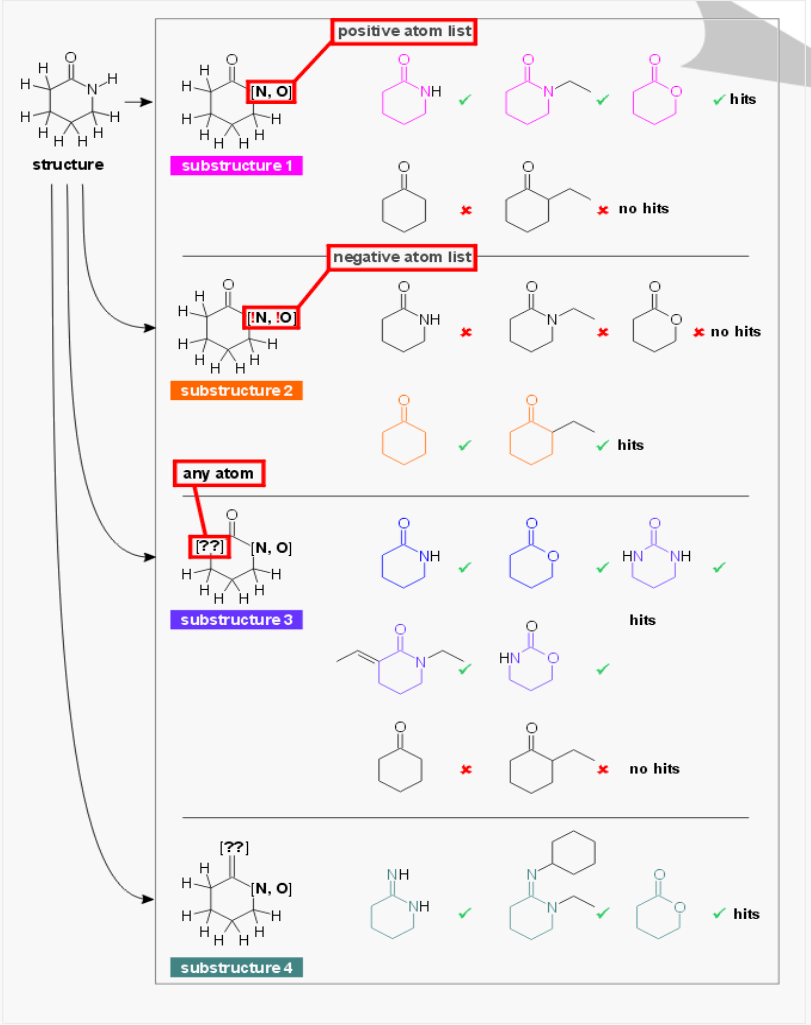
139

Examples



Substructure models





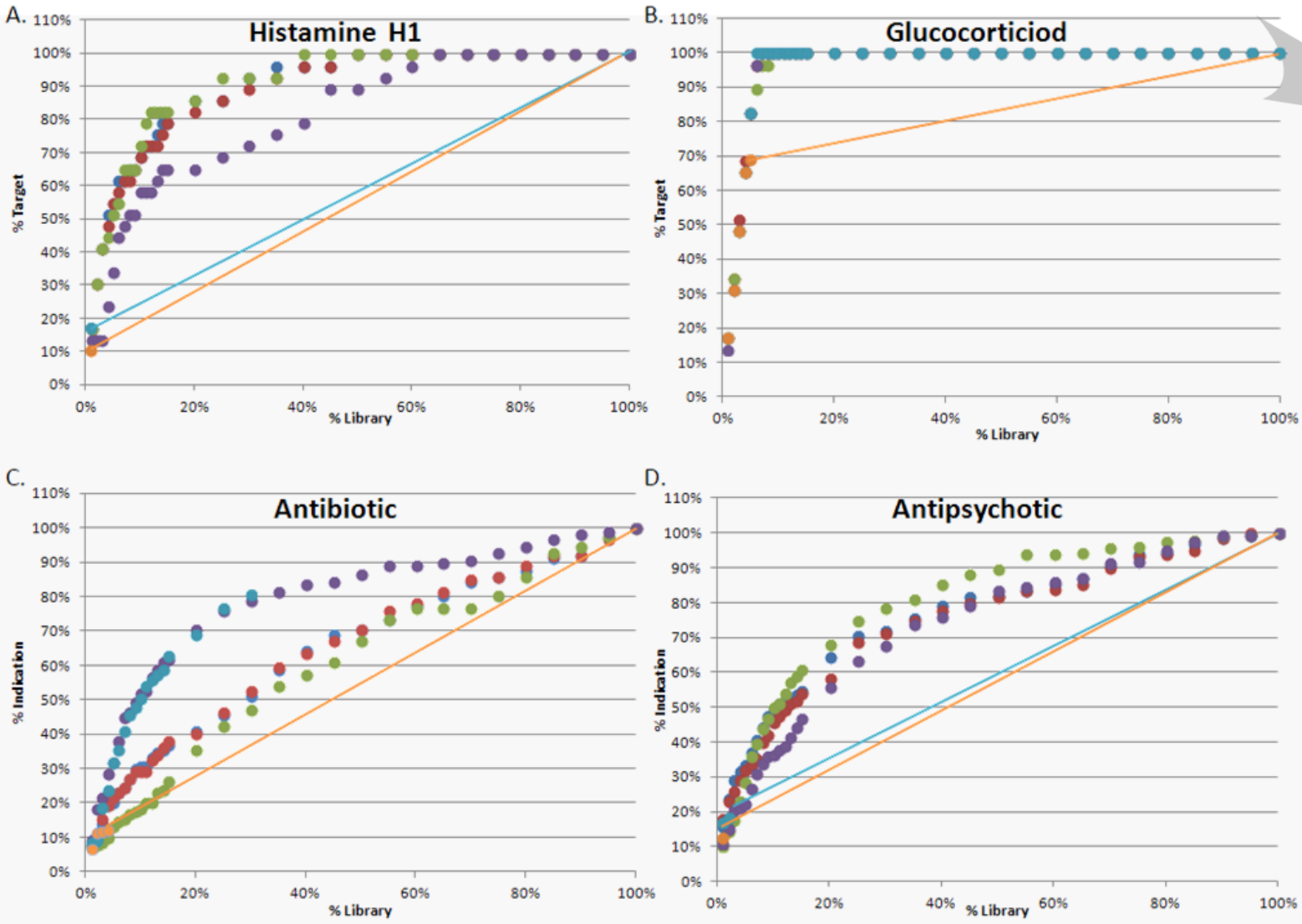
Ligand models: Similarity descriptors

- Evaluate descriptors based on their ability to select compounds belonging to the same target / indication as a reference active compounds.
- Indication particularly relevant to phenotypic screening
- Similarity evaluated by the Tanimoto coefficient

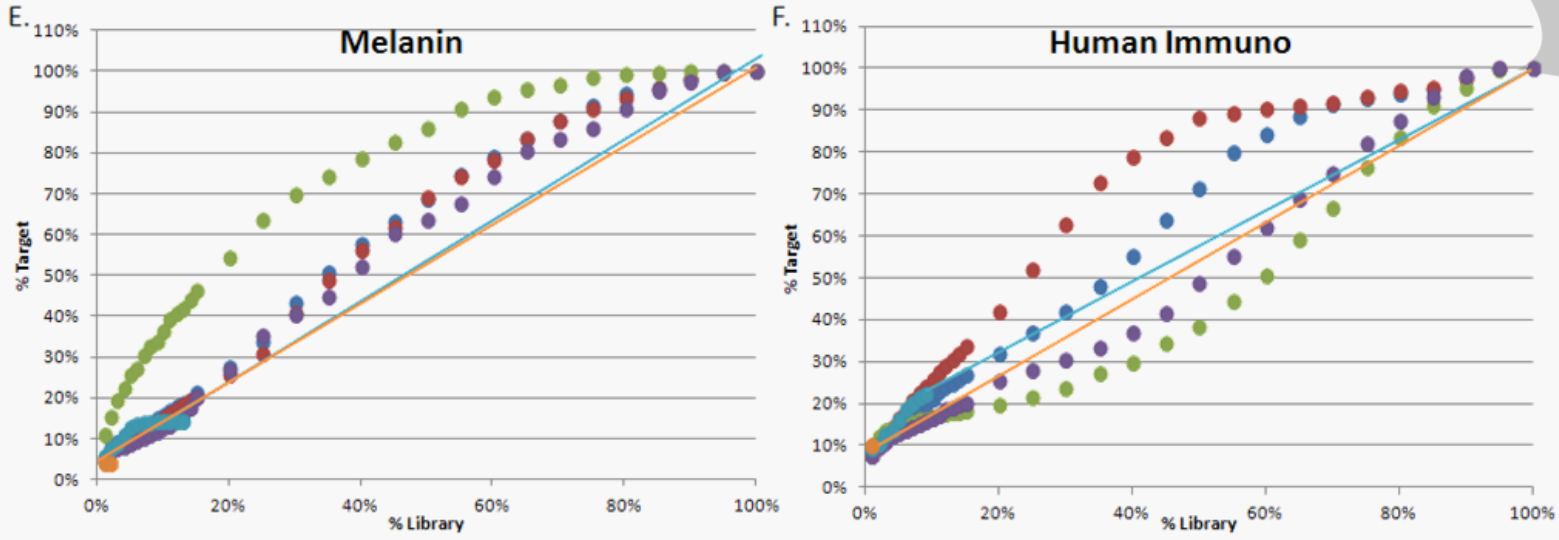
Database	Reference compounds	# Targets / Indication
DrugBank	Fluocinolone acetonide Carinoxamine	Glucocorticiod receptor Histamine H1 receptor
CMC	Haloperidol Lymecycline	Antipsychotic antibiotic
CHEMBL	CHEMBL488890 CHEMBL14759	Melanine concentrating hormone receptor 1 Human immunodeficiency virus type 1 protease

Gilad et al., J. Cheminformatics, 2015, 7:61

Similarity descriptors

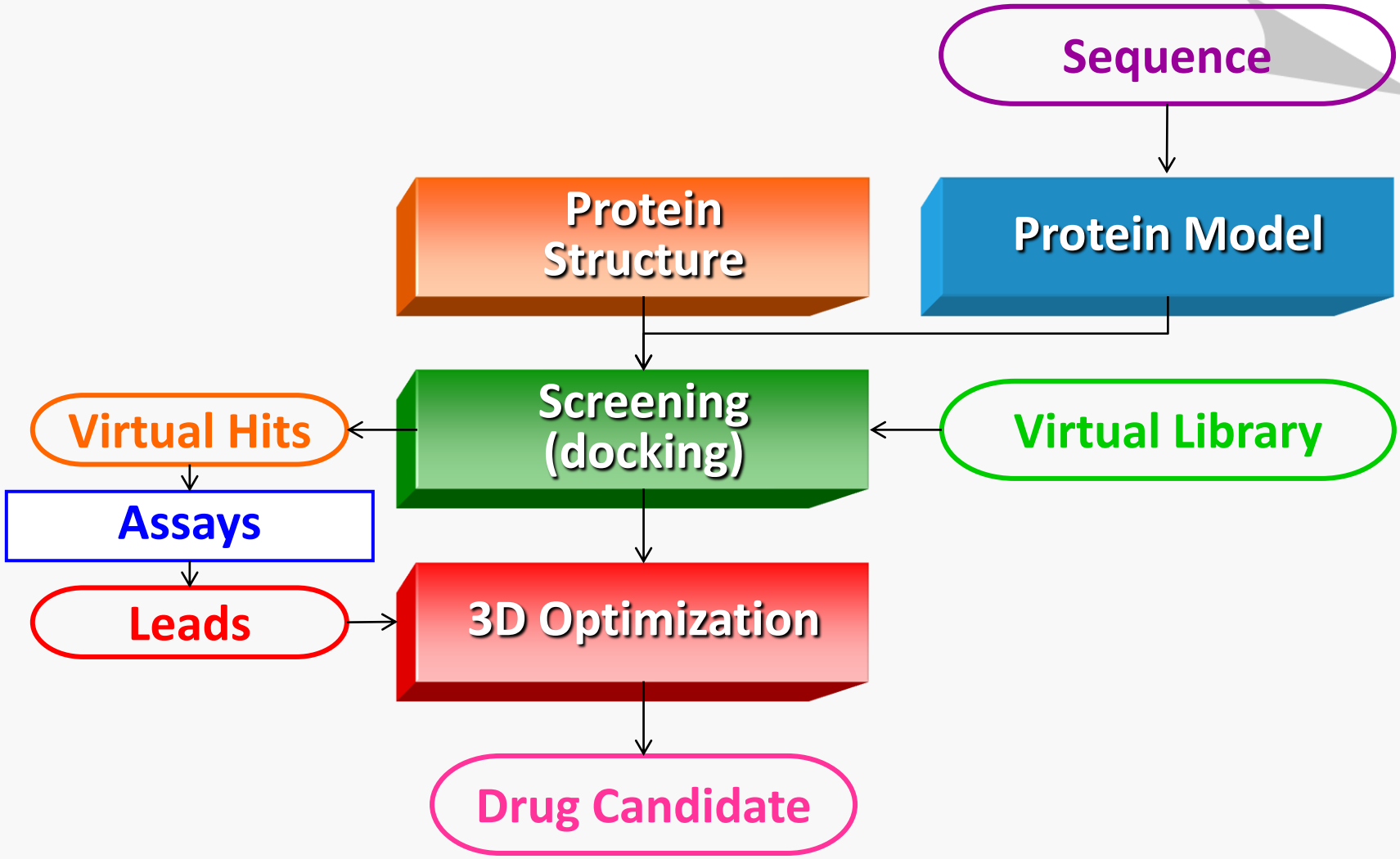


Similarity descriptors



- Enrichment averaged over entire curves and over 6 compounds
- ECFP_4, ECFP_6, MDL, PHFP_3 work well

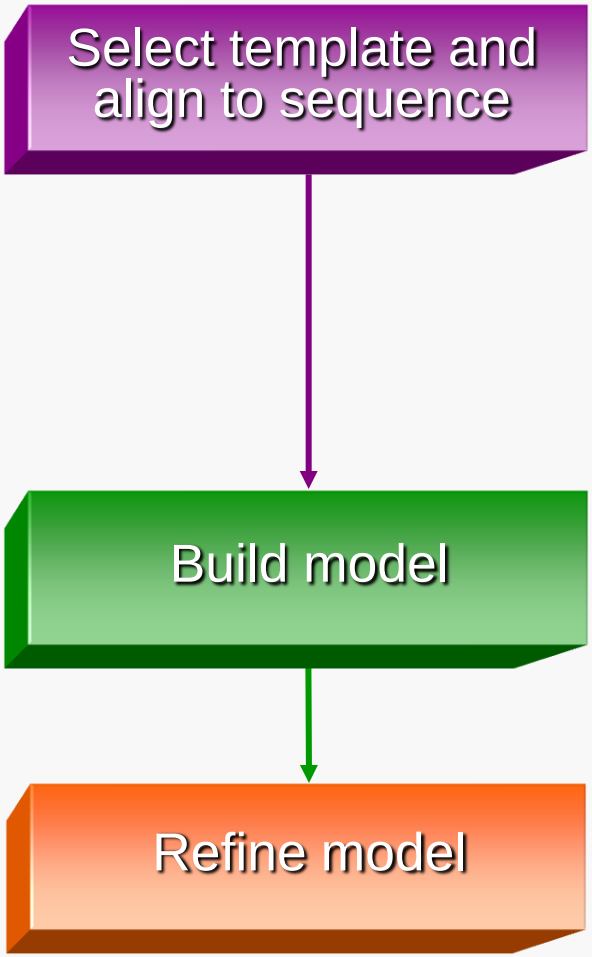
When we know the protein structure



Model development: Preparation of crystal structures

- Download structure from PDB
  - High resolution
  - Solved in the presence of a relevant ligand
- Prepare structure
  - Add hydrogens
  - Check structure for flipped Asn, Gln (look at H-bond pattern)
  - Assign protonation states (specific attention to His at binding sites)
  - Remodel loops
  - Look at conserved water molecules
  - Refine through MD?
- A crystal structure is a snapshot
- A crystal structure is the result of a highly biased selection procedure

Homology (comparative) modeling



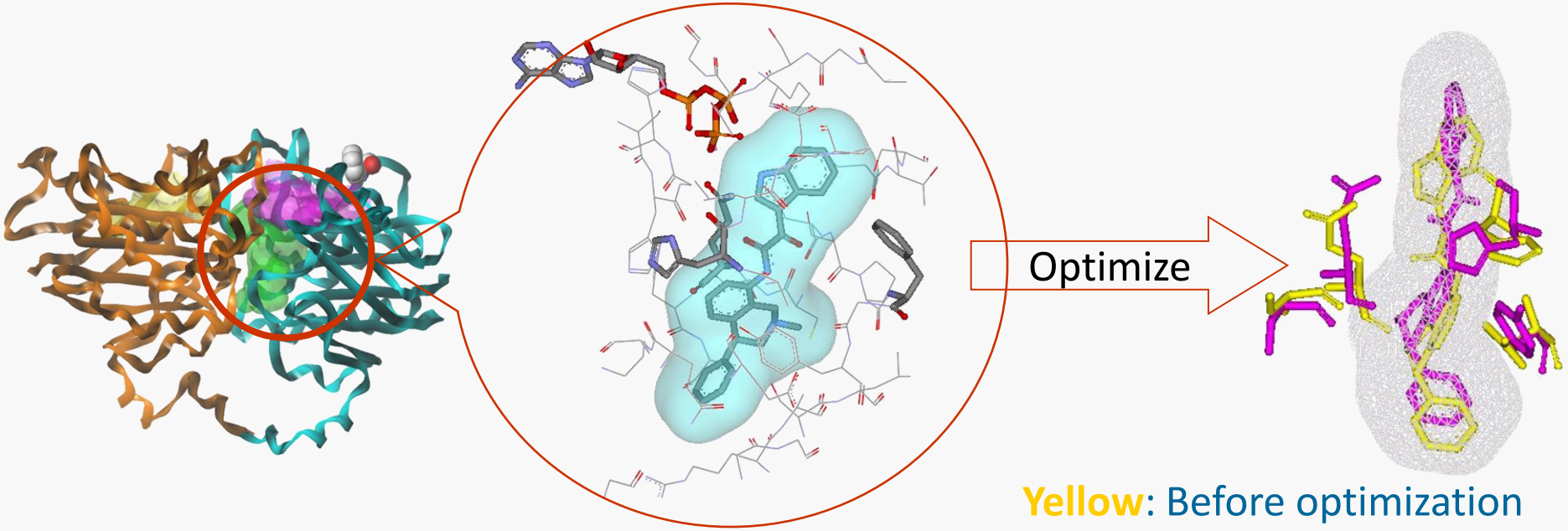
- 3D structure with high (>30%) sequence identity available
  - Pairwise alignment sufficient
- For templates with low sequence identity to target
  - Multiple sequence alignment
- **Always try to improve alignment manually**
- **Avoid gaps in secondary structural elements**

- Use template(s) to model core regions of target
  - Satisfy spatial constraints
  - Average Ca coordinates
  - Database searches of small fragments
- Loop modeling
- Side-chain modeling

- Energy minimization
- Molecular dynamics

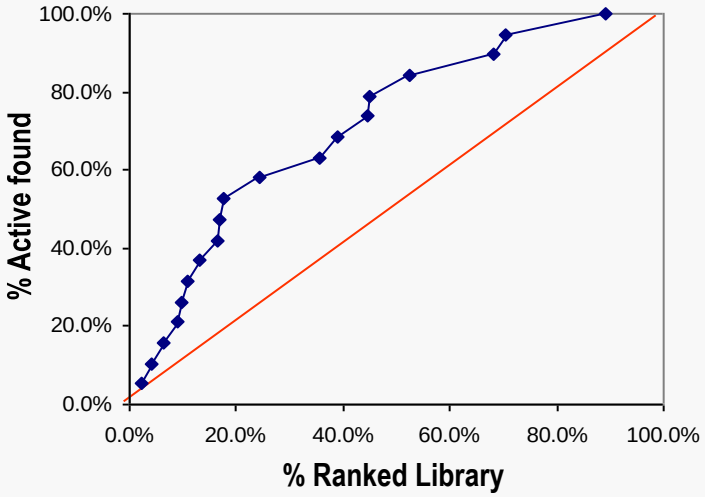
Virtual co-crystal

- Virtual co-crystal is a process by which the binding site is optimized in the presence of a potential ligand using MD simulations
- Past experience has taught us that docking studies perform better on a co-crystal structure rather than on an apo-protein structure

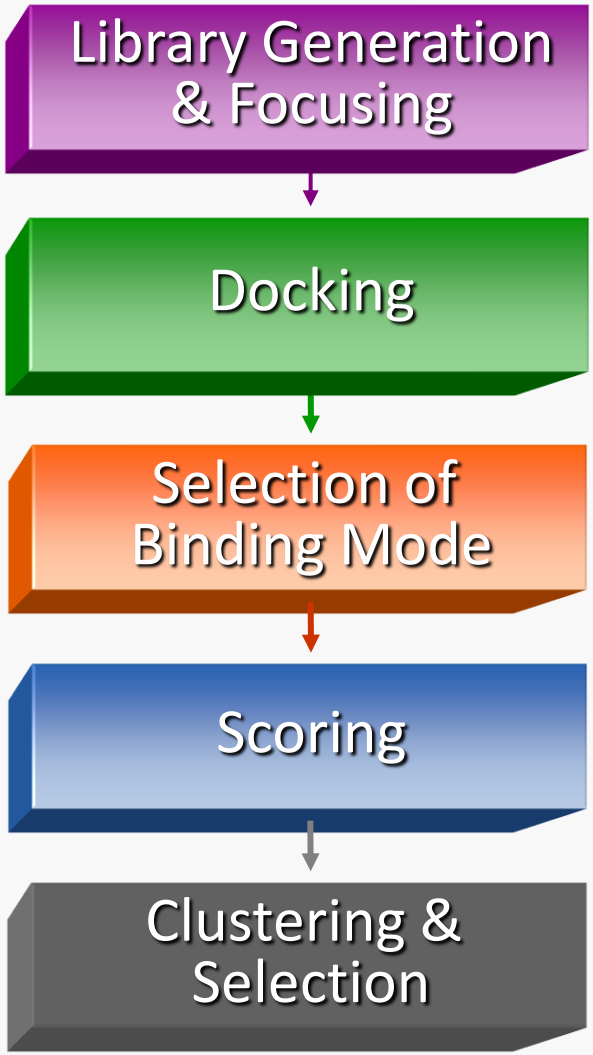


Model validation

- Is this a “healthy” protein?
  - Stereochemistry integrity of the model (use programs such as Procheck, Whatif, Prosa)
  - Stability during (long) MD simulation
- Is this your protein?
  - Target – template RMSD
  - Agreement with experimental data
  - Good enrichment



Virtual screening



- Starting point: 2D representation of compounds
- End point: Multiple 3D conformations of ~100K compounds
- Focusing based on known ligands and binding site characteristics

- Multiple docking tools (Glide, Autodock, Ligandfit, CDocker)

- Target driven (e.g., SiteMap)
- Ligand driven (e.g., pharmacophore)
- Scoring driven

- Multiple scoring functions
- Consensus scoring algorithm

- Clustering and selection of virtual hits (~100-300 per site)
- Visual inspection is critical

Example of library focusing

Chemical property profile based on known compounds

$3x10^6$

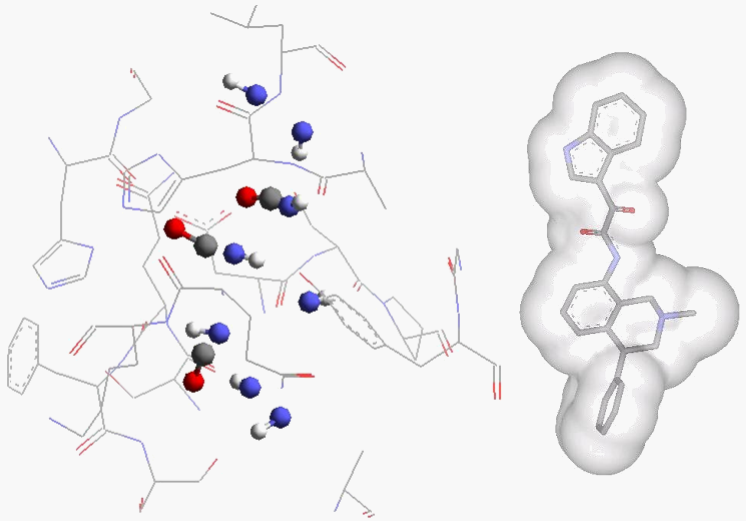
Information from binding-site analysis

Set upper bounds base on Lipinski rule-of-5

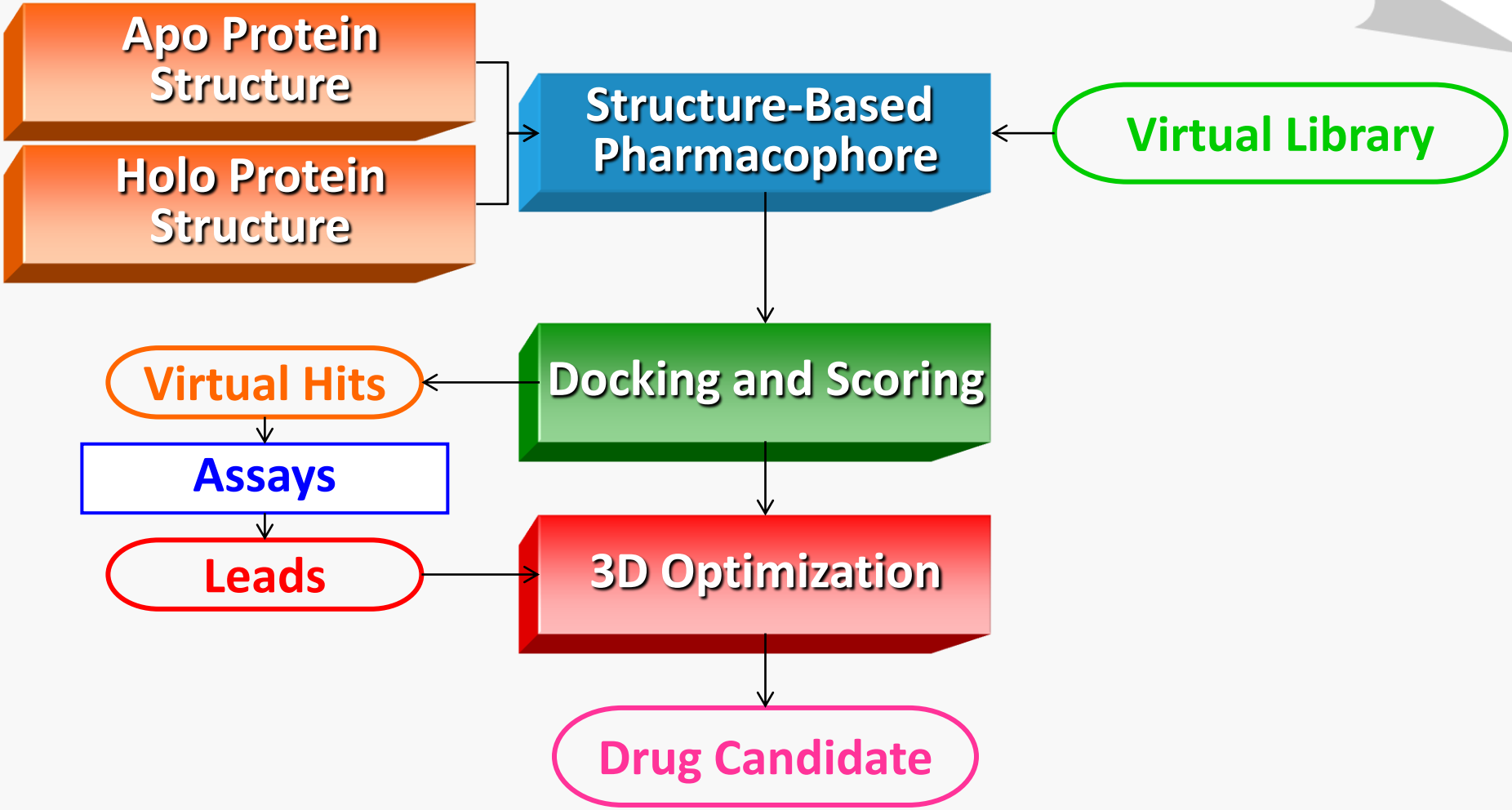
Structure-based focusing (interaction map & shape)

$1.5x10^5$

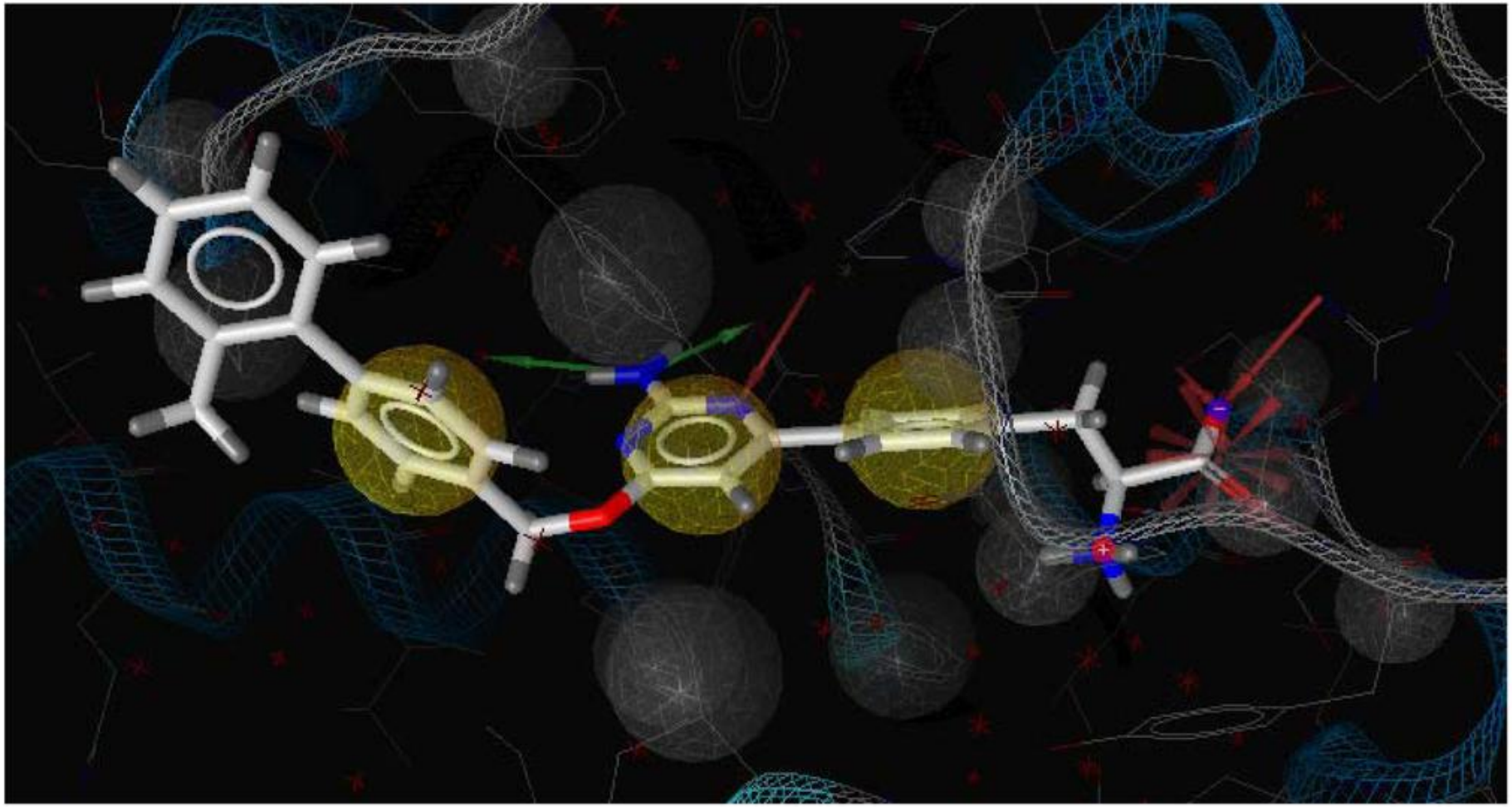
- MW: 250-600
- # H-bond acceptors: 1-4
- # H-bond donors: 1-5
- # rotatble bonds: 3-7
- # aromatic scaffolds: 1-3
- # N4, Carboxy, guanidine: 0



Combining ligand-based and target-based screening



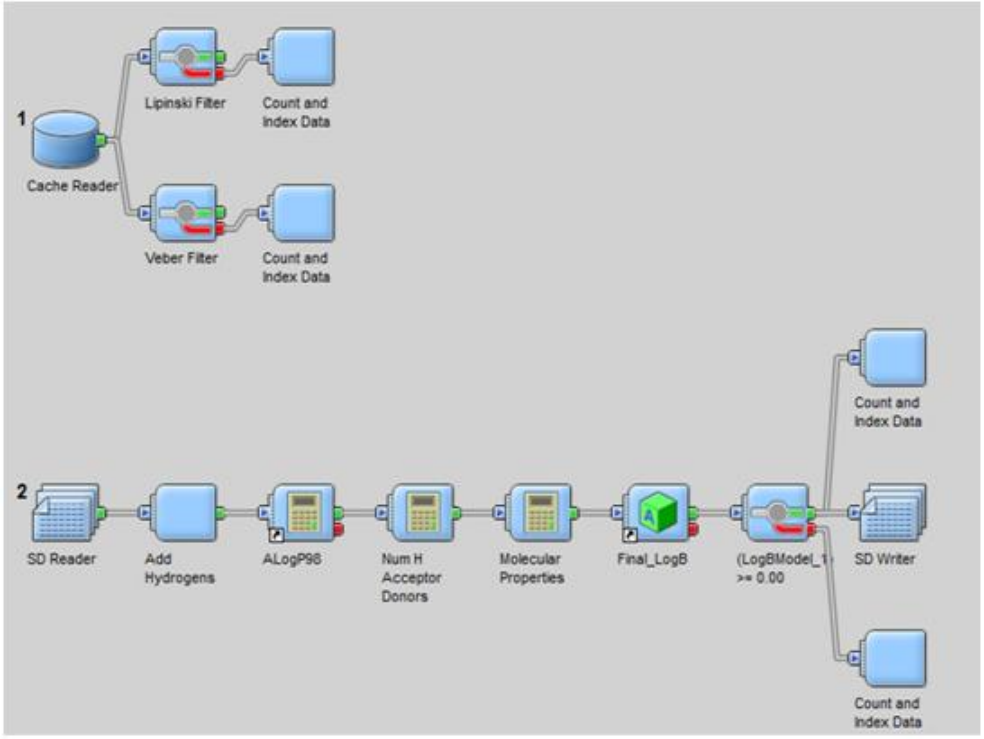
Combining ligand-based and target-based screening



Goal

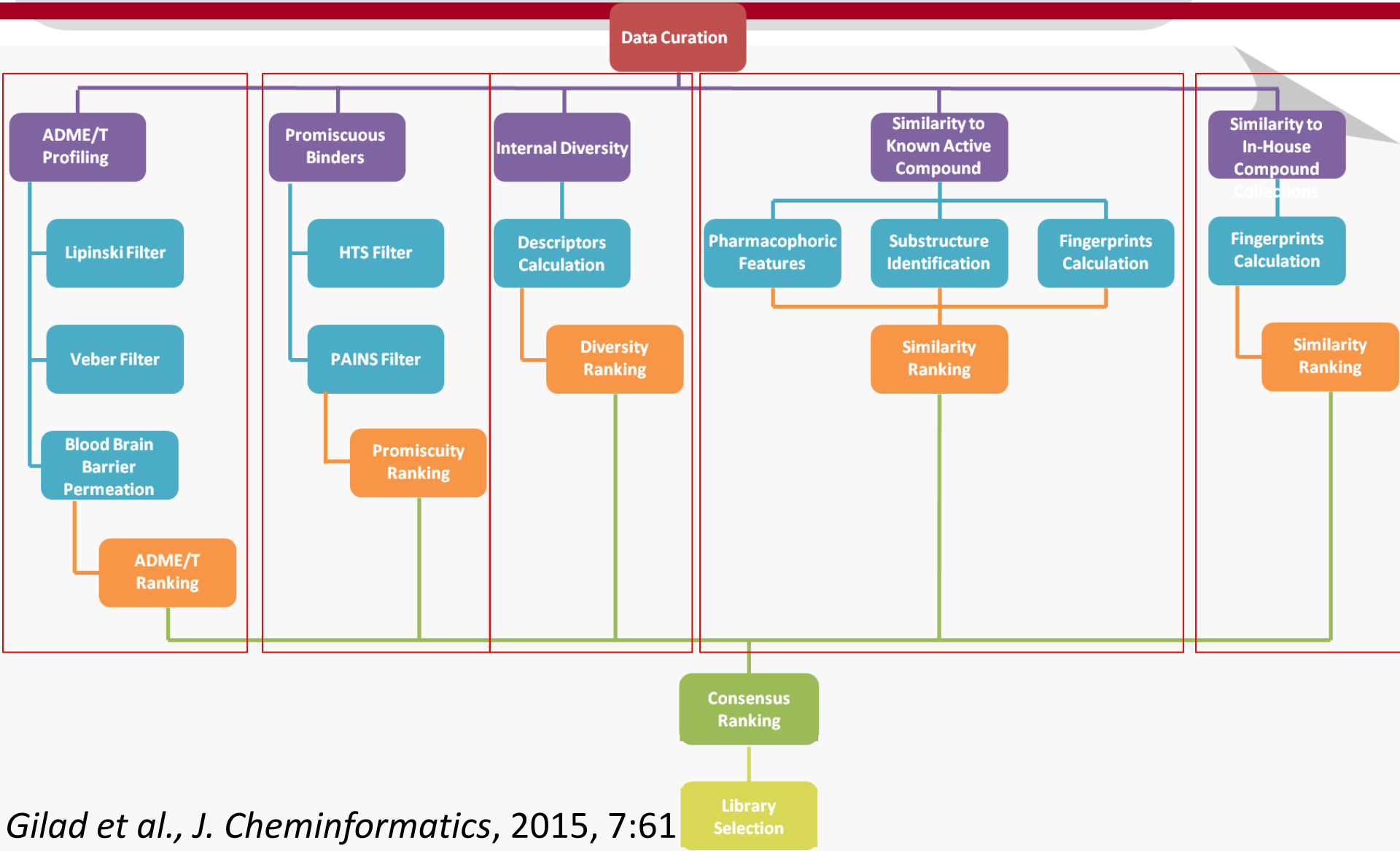
Development of a ***modular, customizable*** work flow for the evaluation and ranking of whole libraries for phenotypic screening

Libraries →



→ Ranks

A library selection workflow



Gilad et al., J. Cheminformatics, 2015, 7:61

Just an idea

Define biological targets

↓

Obtain experimental data (e.g., CHEMBL)

↓

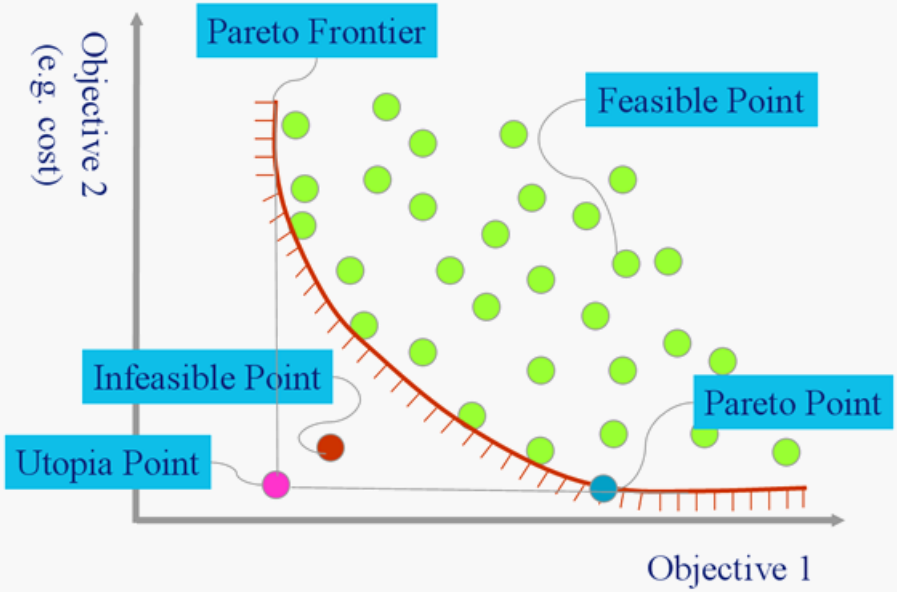
Develop predictive QSAR models

↓

Virtually screen databases using models

↓

Select compounds based on multi-objective optimization



Acknowledgments

Group members

- Avi Yosipof
- Yocheved Gilad
- Luba Simachev
- Netaly Khazanov
- Michael Zhenin
- Kobi Spiegel

Former Group members

- Oren Nahum
- Efrat Noy
- Hannah Avgy
- Gal Fradin
- Tamar Getter
- Shirin Kahremani
- Reut Gigi

ELA
ASSOCIATION EUROPÉENNE CONTRE LES LEUCODYSTROPHIES

B S F
United States – Israel
Binational Science Foundation

INNI
Israel National Nanotechnology Initiative

CYSTIC FIBROSIS FOUNDATION
THERAPEUTICS