



Centurion
UNIVERSITY
*Shaping Lives...
Empowering Communities...*

VIRTUAL SCREENING IN DRUG DISCOVERY



Centurion
UNIVERSITY

Chinmaya Chidananda Behera

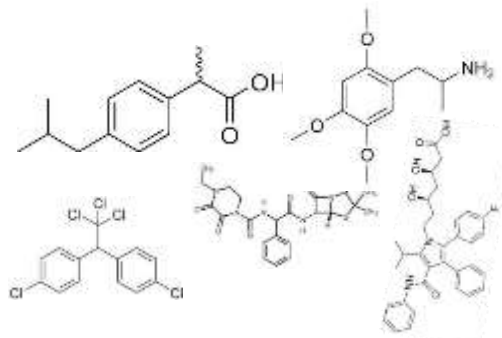
Asst. Professor (Pharmaceutical Chemistry)
School of Pharmacy and Life Sciences
Centurion University, Bhubaneswar



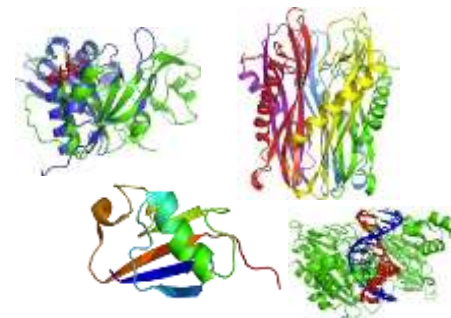
Centurion
UNIVERSITY

Shaping Lives...
Empowering Communities...

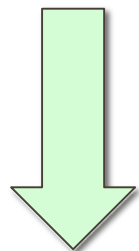
Challenge in Drug Discovery



Chemical Space
 $> 10^{20}$

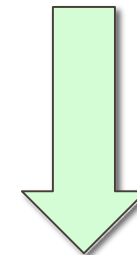


Biological Space
 10^4 to 10^5



“Available”
Compounds

Computational
Screening,
Experimental Screening



Validated
Targets



Centurion
UNIVERSITY

Shaping Lives...
Empowering Communities...

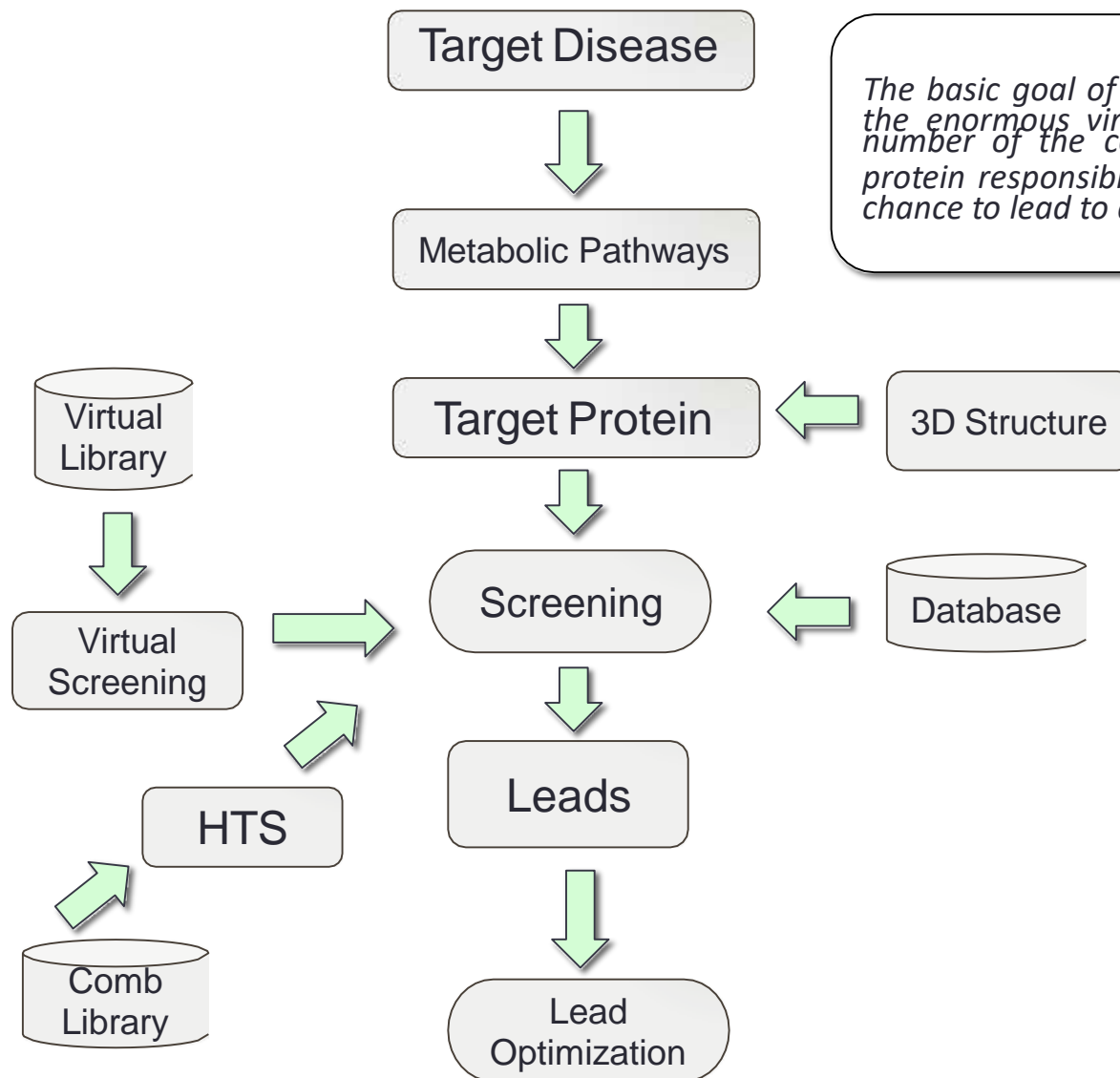
Choosing the right molecule

- Goal: to find a lead compound that can be optimized to give a drug candidate
 - Optimization: using chemical synthesis to modify the lead molecule in order to improve its chances of being a successful drug.
- The challenge: chemical space is vast
 - Estimates vary
 - Reymond et al. suggest there are ~1 billion compounds with up to 13 heavy atoms
 - There are ~65 million known compounds (example UniChem, PubChem)
 - A typical pharmaceutical compound collection contains ~1-5 million compounds
- High throughput screening allows large (up to 1 million) numbers of compounds to be tested
 - But very small proportion of “available” compounds
 - Large scale screening is expensive
 - Not all targets are suitable for HTS

Blum, L.C. & Reymond, J.-louis . J. Am. Chem. Soc. 131, 8732-8733(2009).



Screening Schema in Drug Discovery



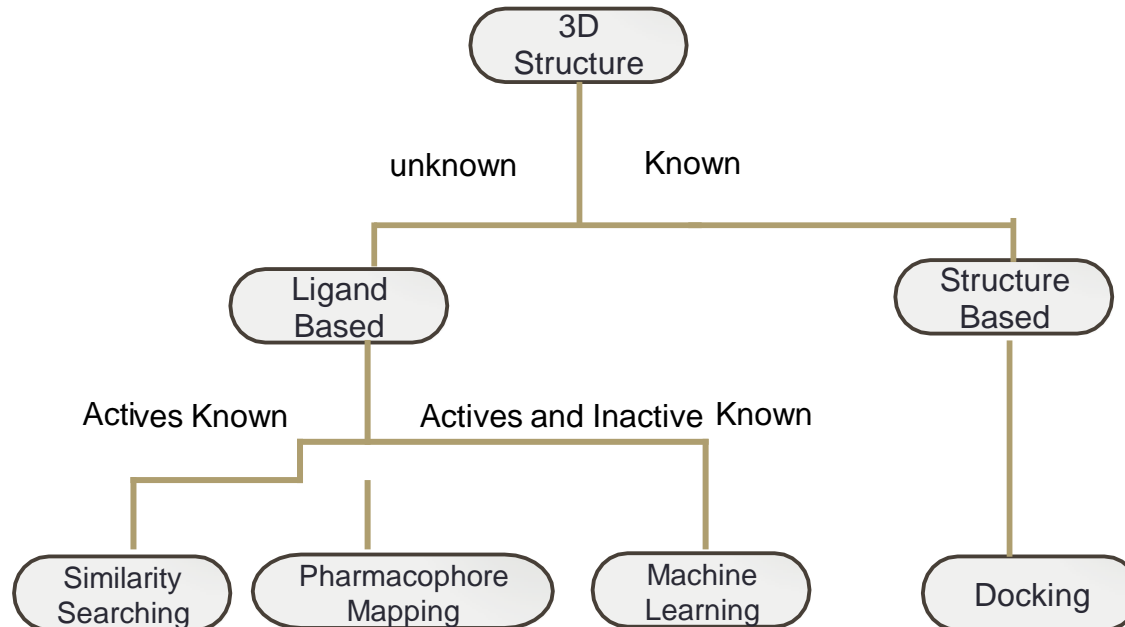
The basic goal of the virtual screening is the reduction of the enormous virtual chemical space, to a manageable number of the compounds that would inhibit a target protein responsible for disease and also have highest chance to lead to a drug candidate.



Centurion
UNIVERSITY

Shaping Lives...
Empowering Communities...

Virtual Screening



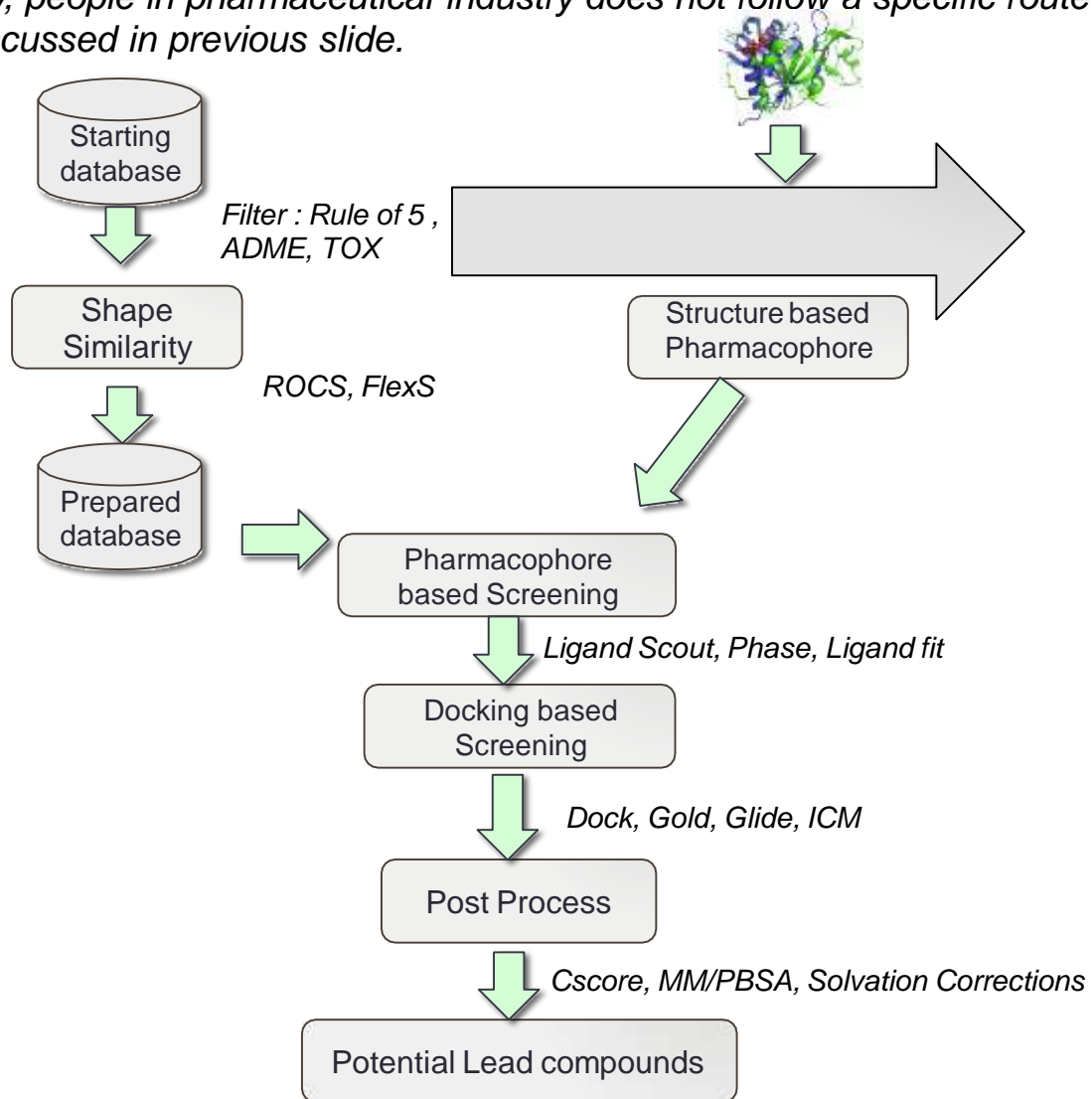
Depending upon structural and Bioactivity data available :

- One or more active molecule known perform similarity searching.
- Several active known try to identify a common 3D pharmacophore and then do 3D database search.
- Reasonable number of active and inactive known train a machine learning model.
- 3D structure of protein known use protein ligand docking.



Hybrid Virtual Screening

Mostly, people in pharmaceutical industry does not follow a specific route they follow a hybrid of methods as discussed in previous slide.



Cleaning Molecules

Remove isotopes, salts and mixtures

Protonation and normalization

Remove duplicates and invalid structures

Filtering Molecules

User defined or other filter

Remove problematic moieties using PAINS, Frequent Hitters etc.

PhyChem property descriptor calculation and filtration

Apply protonation at pH 7.4



Centurion
UNIVERSITY
*Shaping Lives...
Empowering Communities...*

Drug Like Properties

Drug-like properties are an integral element of drug discovery projects. Properties of interest to discovery scientists include the following:

- **Structural properties**
Hydrogen Bonding, Polar Surface area , Lipophilicity, Shape , Molecular Weight, Reactivity, pK_a
- **Physicochemical Properties**
Solubility, Permeability, Chemical Stability
- **Biochemical Properties**
Metabolism(Phase 1 and 2) , Protein and tissue binding, transport
- **Pharmacokinetics(PK) and toxicity**
Clearance, Half-life, Bioavailability, Drug-Drug Interaction, LD_{50}



Leadlike & Druglike

- **Leadlike**

- Molecular weight (MW) = 200–350 (optimization might add 100–200)
- clogP < 1.0–3.0 (optimization might increase by 1–2 log units)
- Single charge present (secondary or tertiary amine preferred)
- Importantly, exclude chemically reactive functional groups , ‘promiscuous inhibitors’, ‘frequent hitters’ and warheads
- Non-substrate peptides are suitable.

- **Druglike**

- Importantly, exclude chemically reactive functional groups , ‘promiscuous inhibitors’, ‘frequent hitters’ and warheads
- MW < 500
- cloP < 5
- H-bond donors < 5
- Sum of N and O (H-bond acceptors) < 10
- Polar surface area < 140 Å²
- Number of rotatable bonds ≤ 10



Centurion
UNIVERSITY

Shaping Lives...
Empowering Communities...

Filtering molecules using structural properties

Basic Washing –

- Removing Salts & Unwanted Elements

Filter out cationic atoms: Ca^{2+} , Na^{+} , etc.

Filter out metals:

Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, Y, Zr, Nb, Mo, Tc, Ru, Rh, Pd, Ag, Cd

Often the salt “filter” = keeping the largest molecule in the sdf entry.

- ALLOWED_ELEMENTS H, C, N, O, F, P, S, Cl, Br, I
- Check proper Atom Types by adding hydrogen and checks if O, N, C valences are correct.
- Check formal charge

Filter out Reactives (false positives for proteins)



Centurion
UNIVERSITY

Shaping Lives...
Empowering Communities...

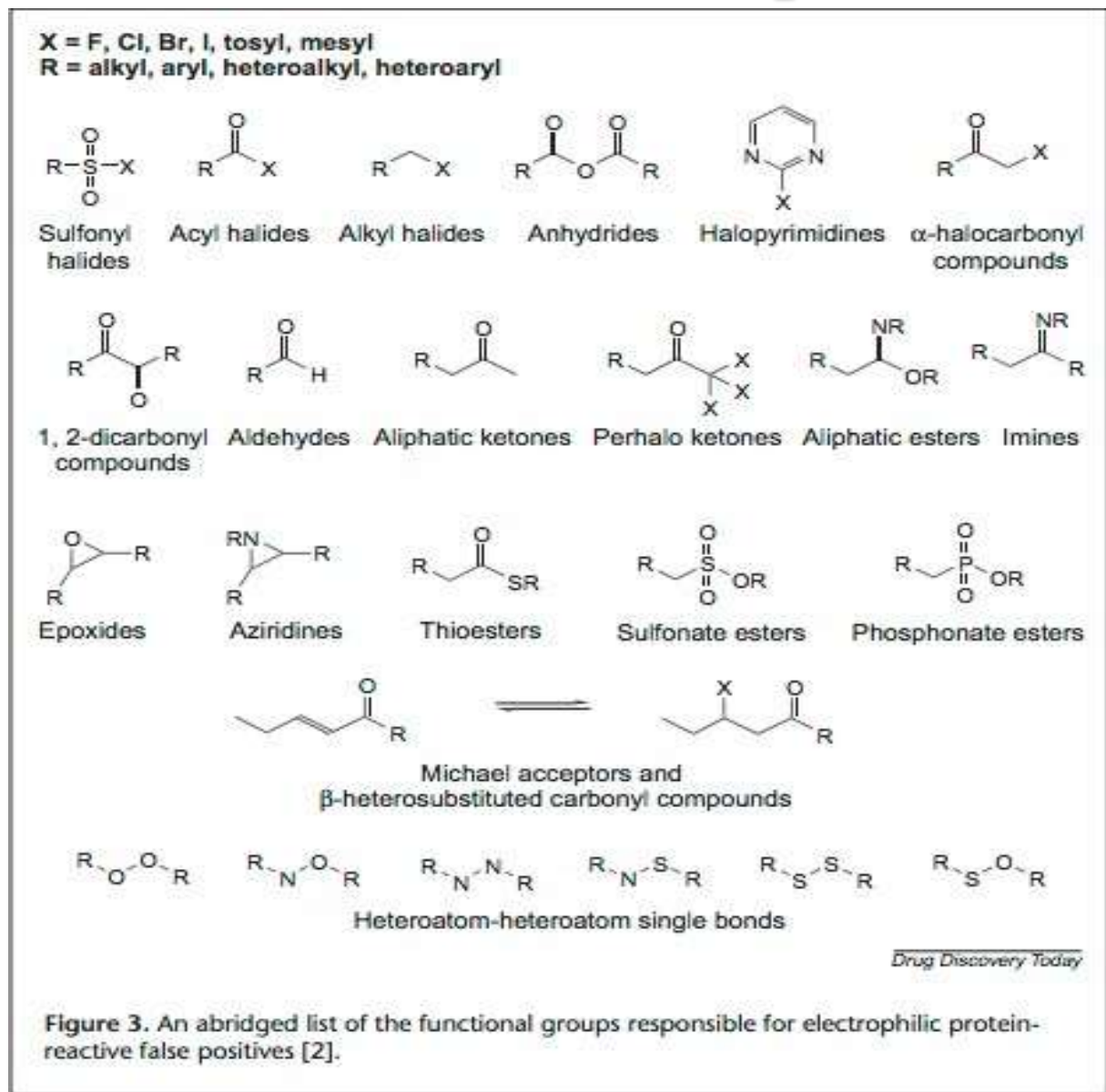


Figure 3. An abridged list of the functional groups responsible for electrophilic protein-reactive false positives [2].

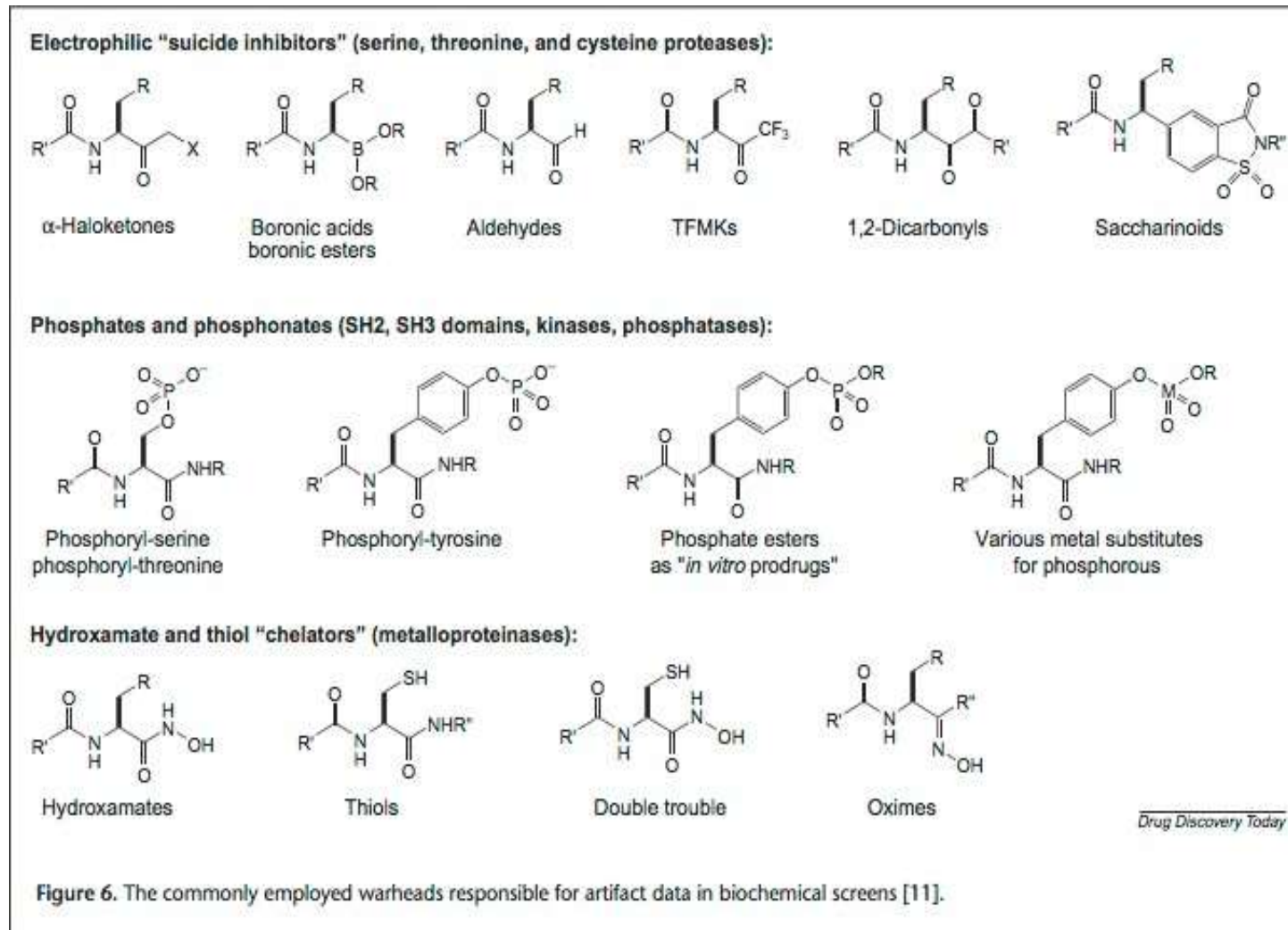


Centurion
UNIVERSITY

Shaping Lives...
Empowering Communities...

Filter out: Synthesis Intermediates, Chelators

'Warhead' agents - functional groups which shows high reactivity to proteins due which there is high attrition rate in drug development.



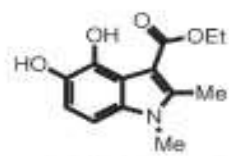


Centurion
UNIVERSITY

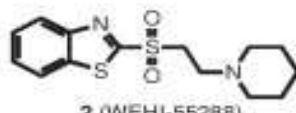
Shaping Lives...
Empowering Communities...

PAINS Filter

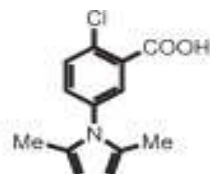
- PAINS = “Pan-Assay Interference Compounds”
- Problematic scaffolds – has cost their Institute time and \$\$



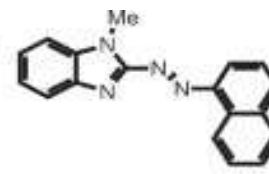
1 (WEHI-14937)



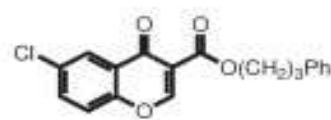
2 (WEHI-55288)



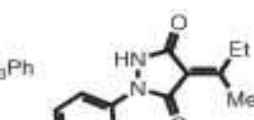
3 (WEHI-60846)



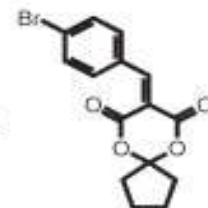
4 (WEHI-84964)



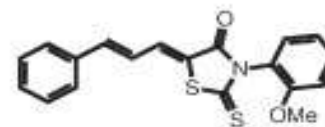
5 (WEHI-66678)



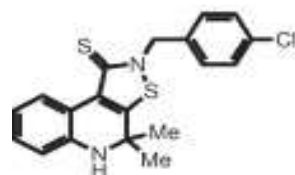
6 (WEHI-66114)



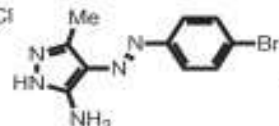
7 (WEHI-73081)



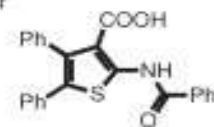
8 (WEHI-76490)



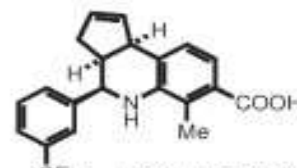
9 (WEHI-81112)



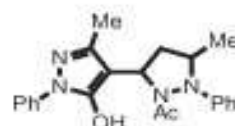
10 (WEHI-91312)



11 (WEHI-96605)



12 (WEHI-97605)



13 (WEHI-98650)



**Centurion
UNIVERSITY**

Shaping Lives...
Empowering Communities...

Rules-of-Thumb for Hit Selection & Lead Optimization

parameter	rules-of-thumb	comment	programs	key references
oral bioavailability ("rule of 5")	MW \leq 500 Da ClogP \leq 5 H-bond donors \leq 5 #(N + O) \leq 10	violation of these limits decreases oral bioavailability	Biobyte ClogP ^{85,86} or ACD LogP v4.0 ¹²	Lipinski (1997) ¹ Wenlock (2003) ¹²
oral bioavailability	Nrot \leq 10 PSA \leq 140 Å ²	violation of these limits decreases oral bioavailability	tPSA ⁶² (nitrogen and oxygen only)	Veber (2002) ¹³
oral bioavailability ("Golden Triangle")	MW \leq 500 variable LogD (LogD range: 0 – 5)	violation of these limits decreases oral bioavailability	experimental LogD	Johnson (2009) ³⁵
toxicity	ClogP \leq 3 PSA \geq 75 Å ²	violation of these limits increases the risk of toxicity	Biobyte ClogP v4.3 ⁸⁵ tPSA ⁶² (nitrogen and oxygen only)	Hughes (2008) ²
toxicity	LLE \geq 5	low ligand-lipophilicity efficiency can lead to increased promiscuity	Biobyte ClogP ⁸⁵	Leeson (2007) ¹⁹ Leach (2006) ²³
membrane permeability	PSA \leq 120 Å ²	violation of this limit decreases membrane permeability	Quanta 3D (nitrogen and oxygen only)	Kelder (1999) ⁶¹
membrane permeability	MW \leq 500 variable LogD (LogD range: 0.5 – 5)	violation of these limits decreases membrane permeability	ACD PhysChem Batch ⁸⁷ or AZlogD ⁸⁸	Bhal (2007) ³⁴ Waring (2009) ³⁶
blood–brain barrier penetration solubility	PSA \leq 70 Å ² Fsp3 \geq 0.4	violation of this limit decreases brain penetration increased fraction of sp3 hybridized carbons (Fsp3) increases solubility	Quanta 3D (nitrogen and oxygen only) Pipeline Pilot 7.5	Kelder (1999) ⁶¹ Lovering (2009) ⁵¹
general "developability"	number of aromatic rings \leq 3	increase in aromatic ring count decreases solubility and increases protein binding	none listed	Ritchie (2009) ⁵²



Centurion
UNIVERSITY
Shaping Lives...
Empowering Communities...

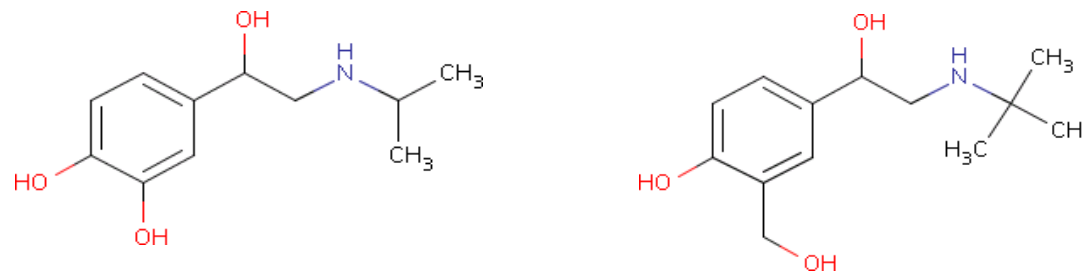
Similarity Searching

What is it ??

Chemical, pharmacological or biological properties of two compounds match.

The more the common features, the higher the similarity between two molecules.

Chemical



The two structures on top are chemically similar to each other. This is reflected in their common sub-graph, or scaffold: they share 14 atoms

Pharmacophore



The two structures above are less similar chemically (topologically) yet have the same pharmacological activity, namely they both are Angiotensin-Converting Enzyme (ACE) inhibitors



Centurion
UNIVERSITY

Shaping Lives...
Empowering Communities...

What is required for a similarity search ?

- A Database SQL or NoSQL (Postgres, MySQL, MongoDB) or flat file of descriptors eg: ChemFP
- Chemical Cartridge to generate fingerprints(descriptors) for molecules (RDKit, openbabel)
- Similarity function to calculate similarity(Jaccard, Dice, Tversky) this can be written in c,c++ or python as a function inside SQL databases.



Centurion
UNIVERSITY

Shaping Lives...
Empowering Communities...

3D based similarity

- Shape-based ROCS (Rapid Overlay of Chemical Structures) Silicos-it.com (Shape it)
- Computationally more expensive than 2D methods
- Requires consideration of conformational flexibility
 - Rigid search - based on a single conformer
 - Flexible search
 - Conformation explored at search time
 - Ensemble of conformers generated prior to search time with each conformer of each molecule considered in turn
 - How many conformers are required?



Centurion
UNIVERSITY

*Shaping Lives...
Empowering Communities...*

THANK YOU