Meet-EU

logs

16/10

rozdělil bych meeting na 4 části:

1. pochopení TrmD - mechanismy, které bychom mohli využít
2. přidělení mechanismů pro hlubší zkoumání
3. letmý návrh pipeliny (měli bychom dělat constrains nějaký - predicting toxicity, v jakém prostředí jsou stabilní atd.)
4. rozdělení úkolů na sehnání good DBS (a jak s nima pracovat), nějaký github tips

dělat diagramy on the way (abychom vždy věděli kde jsme a jak to funguje v pipě -> highlevel all the way to lower levels)
možná by bylo dobrý to udělat modular (možná registry pattern -> ale to co chci je, aby on the way jsme to mohli spouštět - integrativně i unit testama; nedodělaný části mohou být dummy classy, ale abychom pořád při každém pushi věděli, že to fachá)

datová analýza databází a vzniklých dat on the way, abchom věděli s čím pracujeme a jestli není něco špatně (možná pak můžeme i charakterizovat screened ligandy atd.; mužem taky říct po získání různých trmds, jak jsou od sebe odlišné průměrně atd - ze sekvence a struktury - jaké jsou outliers atd. - nějaké struktury mohou být hodně jiné -> vyřadit/nevyřadit)

CGBench (benchmark pro Struct-Based Drug Design)

budem vázat na apo/holo struktury?
voda/Mg/prostředí

14/10

- nutný determinovat stabilitu ve specific prostředí
- počítat s H2O
- na začátek meetu jednotlivý papers a pak je prodiskutujem

- měli bychom založit drive nebo aspoň sdílenej dokument https://researchrabbitapp.com/home
- astra zeneca papers
- příští týden už nějakej rough přístup a literaturou
- radit se s týpkem na drug design a používat MetaCentrum

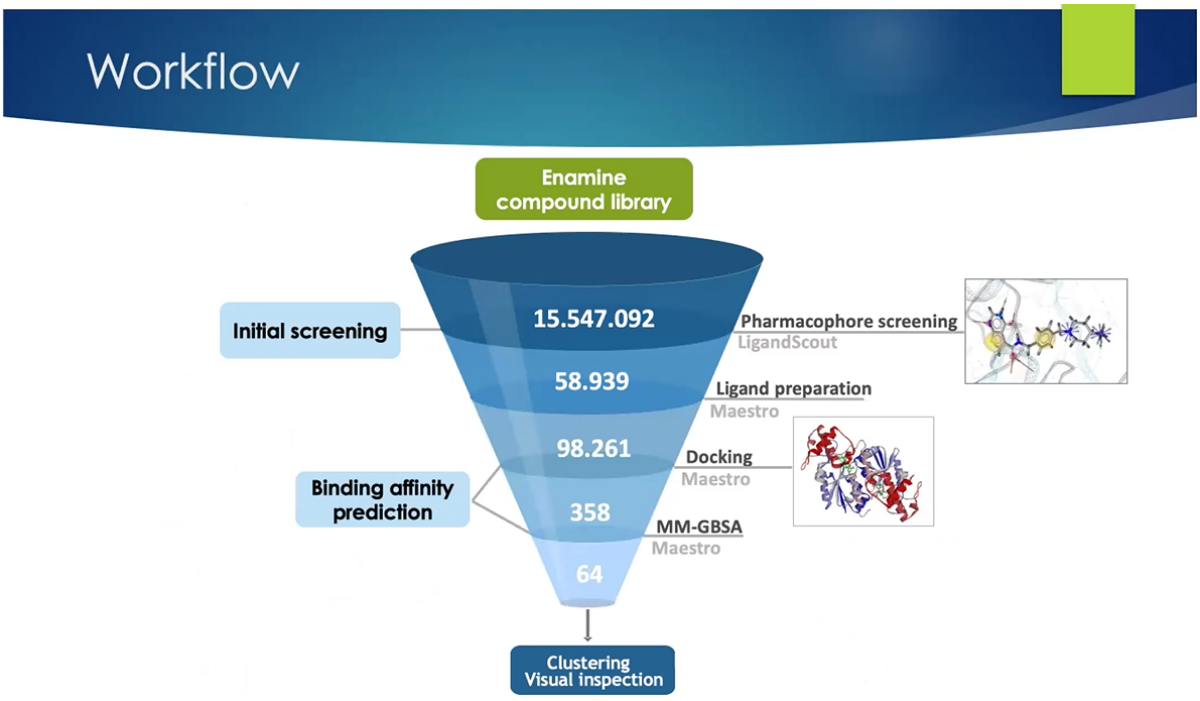
Ideas:

- WHAT we actually want to do is stop the methylation process
- pipeline (MolDynamics a known varianty abychom zjistili víc stable structs (protein není rigidní) -> pocket detection (binding sites) -> known inhibitors (abychom měli library) -> validovat jejich docking na původní strukturu)
- pokud nenajdeme v DBS possible medical compounds strukturu s dobrou afinitou, pak bychom mohli zkusit syntetizovat novou pomocí AlphaProteo mby
- má cenu vůbec prohledávat databáze léků on the market? (protože TrmD je prý resistant na všechny) - SPECS library jsou compounds easy to make
- najít cryptic pockets
- block dimerization? (it is active only when dimere)
- what about having two compounds - one bound to change conformation and then the other to block the function and release the other one so that trm5 is not blocked idk
- block the knotting - but we dont know which structure helps the protein to do the knotting when folding or wthv
- we can find mRNA interactions as well (we can consider gene therapy/vaccination - anything tbh)
- what about trying to remove the Mg ion (this could maybe change the confo and help identify new pockets - but the Mg is only in one crystalographic struct since it has been discovered lately that it has important role; H2O takes the place of removed Mg ion)
- blokovat dimerický interface je taky možnost (nebo prostě jiným způsobem než vazbou v klasickým místě); monomer vs dimer
- možný i protein binder - pořád ale drahé
- dockovat do apo nebo holo struktury je taky možnost divergence
- využít toho, jestli se posttranslačně modifikuje
- predikovat jak by se mohla struktura/sekvence vyvíjet v čase (to by nám ale zase zúžilo šance)

- nejen targetovat aktivní místo, ale zkusit hledat cryptic pockets (třeba po navázání se trochu jinak rozbalí), nebo alosterické místo

Formal pipeline:

1. **target identification** and **validation** (get everything about the biological target, refine and clean)
2. **hit discovery**
   a) SBVS (identify small molecules that can bind; 3D struct, docking, MD)
   b) LBVS (use known active compounds to identify candidates; QSAR, pharmacophore, similarity searching or ML trained on known ligands)
3. **lead optimization** (refine hit compounds; SAR, FED, ADMET, ML tools)
4. ***in silico* toxicity prediction** (ensure pharmacokinetic and toxicity profiles)
5. **molecular dynamics** (evaluate stability and dynamics of the complex; GROMACS, MM/..SA)
6. ***in silico* mutagenesis** (drug resistence mutations, point mutations in pyMOL..)
7. ***in silico* drug repurposing** (match existing drug profiles with the disease's molecular signature; MD, similarity search)
8. **binding-free energy calculations and ranking** (estimate binding affinity; FEP, TI, MM/GBSA, ML methods)
9. ***in silico* clinical trial simulation** (predict outcomes, safety; PBPK, Adverse Drug Reactions)
10. ... not our problem

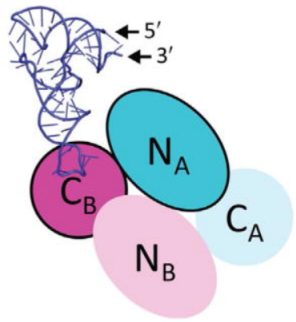


Tools:

- **Molecular Docking**: AutoDock, Glide, DOCK, GOLD
- **Molecular Dynamics (MD)**: GROMACS, AMBER, NAMD
- **ADMET Prediction**: QikProp, pkCSM, ADMET Predictor
- **Structure-based tools**: Schrodinger Suite, MOE, Discovery Studio
- **Ligand-based tools**: RDKit, OpenEye, ChemAxon
- **Machine Learning**: TensorFlow, Scikit-learn, DeepChem
  https://prankweb.cz/viewer?id=8APT&database=v3-conservation-hmm

TrmD

- S-adenosyl methionine (AdoMet)-dependent methyl transferase that synthesizes the methylated m1G37 in tRNA
- methylates G37 of tRNA containing the seq G36pG37
- essential for bacterial growth
- unusual by using a topological protein knot to bind AdoMet
- the structure is overall rigid but the knot has complex dynamics (to transmit the signal of AdoMet binding to promote tRNA binding and methyl transfer)
- mutations in the knot block the signalling and decrease the synthesis of methylated tRNA

- requires Mg2+ in the catalytic mechanism (for methyl transfer it is theorized)
- in addition to Mg2+, TrmD can also use Ca2+ and Mn2+ as an active ion
- N-terminal domain of the monomer:
  - has a trefoil knot (in which AdoMet or AdoHcy is bound) (While protein knots are generally proposed to enhance protein stability (thermal as well), the trefoil knot in TrmD was shown to be required for methyl transfer)
  - active knot is more rigid (on B chain)
- C-terminal domain of the monomer:
  - shows structural similarity to trp repressor
  - binds tRNA
- however only the B chain (monomer) is capable of binding tRNA -> by positioning G37 base within its flexible linker (disordered linker into helical structure) -> dimer binds two AdoMet but only one tRNA (and only AdoMet in chain A is active for methyl transfer)
- substrate levels regulation - ligand bound to A prevents ligand binding to B
- compare TrmD and Trm5 in detail (no big difference apart the topology of two active sites; trm5 je šíleně essential - myslim, že by ani trochu neměla být šance, že se náš compound naváže)
- what is bent active conformation (the one monomer is active when in bent conformation?; active site je dost flexible proto je hodně struktur)
- all tRNA molecules contain posttranscriptional modifications (the methylation suppresses tRNA frameshifting - frameshifting unline missense is commonly lethal)
- weird: >TrmD and Trm5 are therefore considered as an analogous pair of enzymes that share no structural homology, but use the same chemical substrates
- homodimer (and places each active site at the dimer interface), Trm5 is monomeric
- sequence is very similar (conserved) among bacteria
- 

- bent conformation of AdoMet is more rigid than open and extended
- the knot helps AdoMet to take the bent conformation
- G37 and G36 substrates for methylation

> The bent shape offers structural novelty and diversity from the conventional AdoMet analogs, and it should be explored for drug design.

> Identifying Pro codons in bacterial genes that are most dependent on TrmD for translation would be a major step forward to develop a TrmD-specific strategy that will improve human health in the global population.

> Mutational analyses demonstrate that the knot is important for AdoMet binding and catalytic activity, and that the C-terminal domain is not only required for tRNA binding but plays a functional role in catalytic activity.

> We imagine that the overall reaction might proceed as follows; tRNA is first non-specifically bound to enzyme, perhaps involving surface C-terminal sequences. We propose, based on earlier biochemical evidence, that the C-terminal domain may recognize and bind the core (hinge) structure of the general tRNA molecule. Next, tRNA might be positioned in the cleft and, if the identity element G36pG37 is present, G36pG37 might be moved or perhaps “flipped” into the dual pocket catalytic center for subsequent methylation by bound AdoMet. If a G is not found at position 36, then a stable complex of G37 near AdoMet cannot form and catalysis does not proceed.

> common evolution origin and form a single superfamily (of SpoU - TrmH and TrmD)

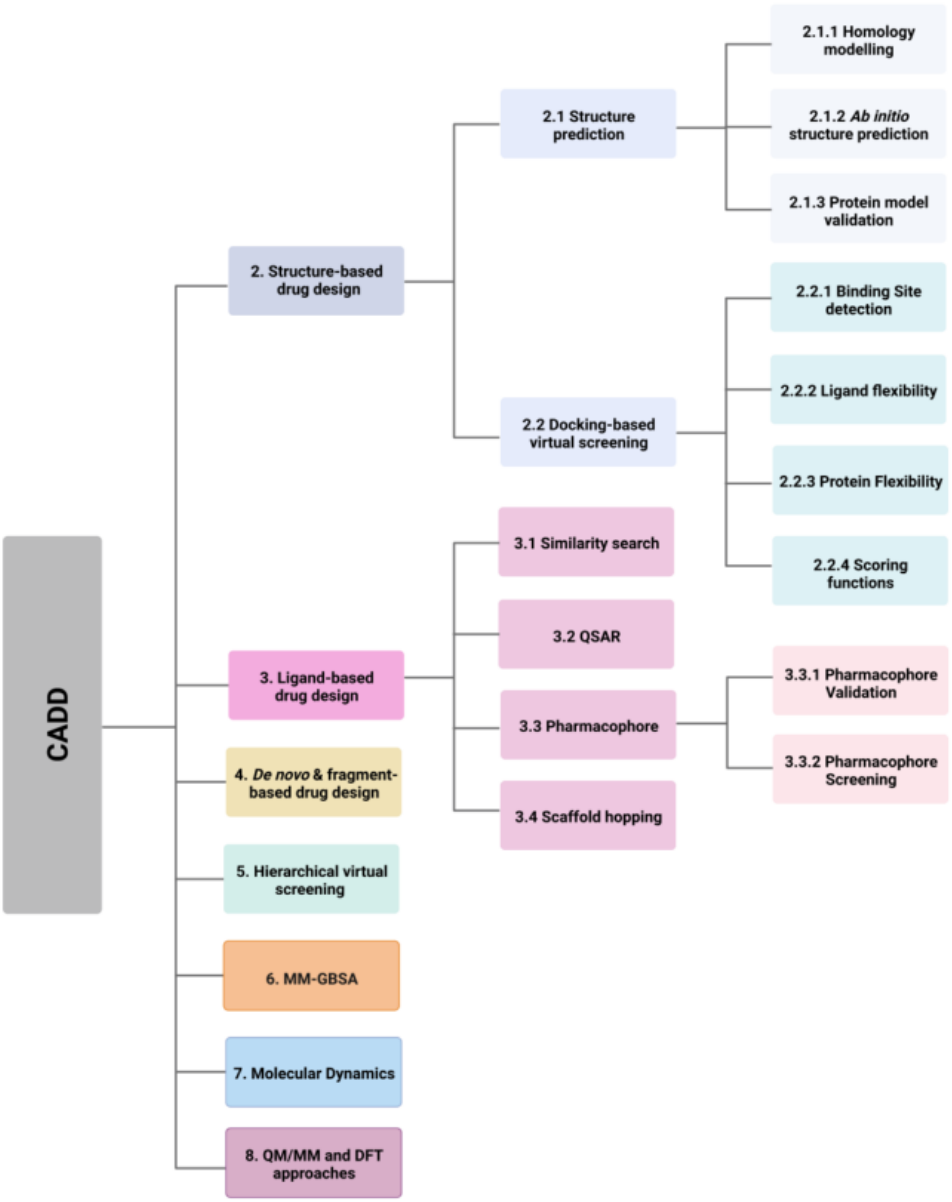
> disruption of the anti-codon stem caused the complete loss of the methyl group acceptance activity, the anti-codon stem is essential for the recognition

Diffusion methods

- metrics: E(3)-equivariant score-based diffusion framework for 3D molecular generation via SDEs, aiming to address the constraints of unified Gaussian diffusion methods

A Guide to In Silico Drug Design

Computer-aided drug design



2. Structure-Based DD

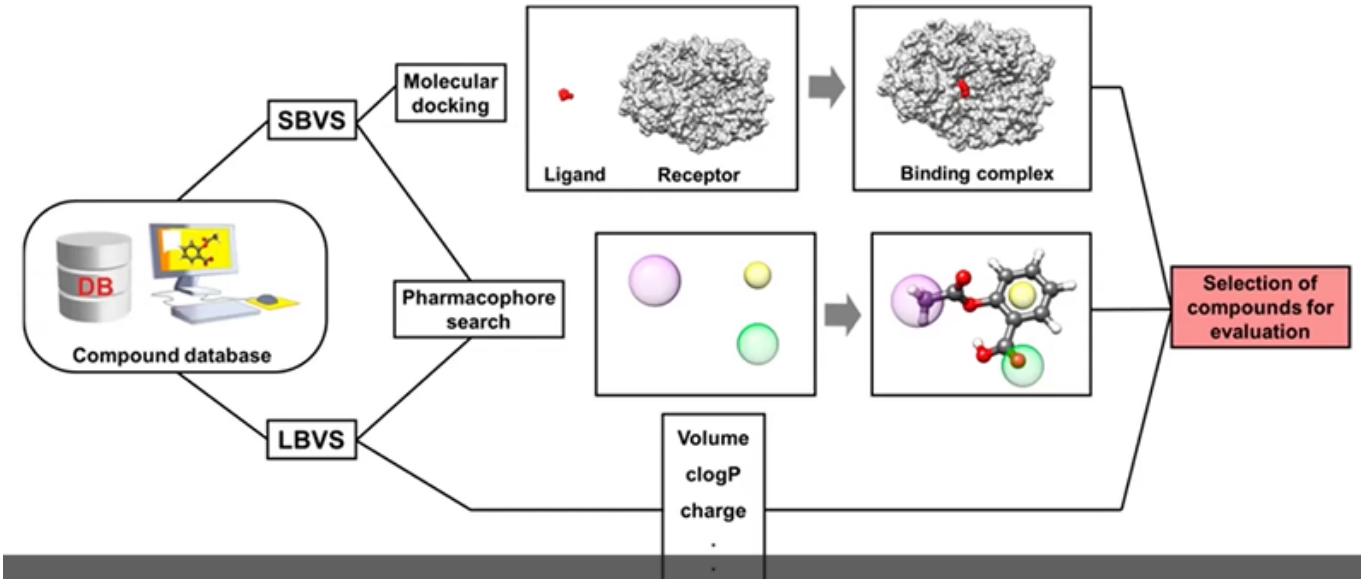
- aim to predict the Gibbs free energy of binding (∆Gbind), the binding affinity of ligands to the binding site, by simulating the interactions between them
- molecular dynamics simulations, molecular docking, fragment-based docking, de-novo drug design

Structure prediction

- ligand-based homology modelling (ligand-based?; e.g. for GPCRs - the receptor is reorganised and refined based upon the ligand binding in order to better accommodate ligands with higher affinity)
- then validation of the predicted sturcture

Virtual screening

3D structures need to be prepared (known protonation states, free atoms etc.)



docking-based

- tries to predict binding modes of both ligand and the receptor (affinity, binding modes, interaciton patterns)
- AutoDock, GOLD, Glide, SwissDock
  (1) pose prediction to envisage how a ligand may bind to the receptor,
  (2) virtual screening to search for novel drug candidates from small molecule libraries and
  (3) binding affinity prediction using scoring functions to estimate the binding affinity of ligands to the receptor

> A good search algorithm should explore all possible binding modes, and this can be a challenging task

> modern docking programs treat both protein and ligand with varying degrees of flexibility in order to address this issue

> Most of the protein structures in the PDB are ligand-bound (holo) structures, which defines the binding pocket and provides us with its geometries. In cases where only ligand-free (apo) structures available, there are traditionally three main types of method to identify potential druggable binding sites

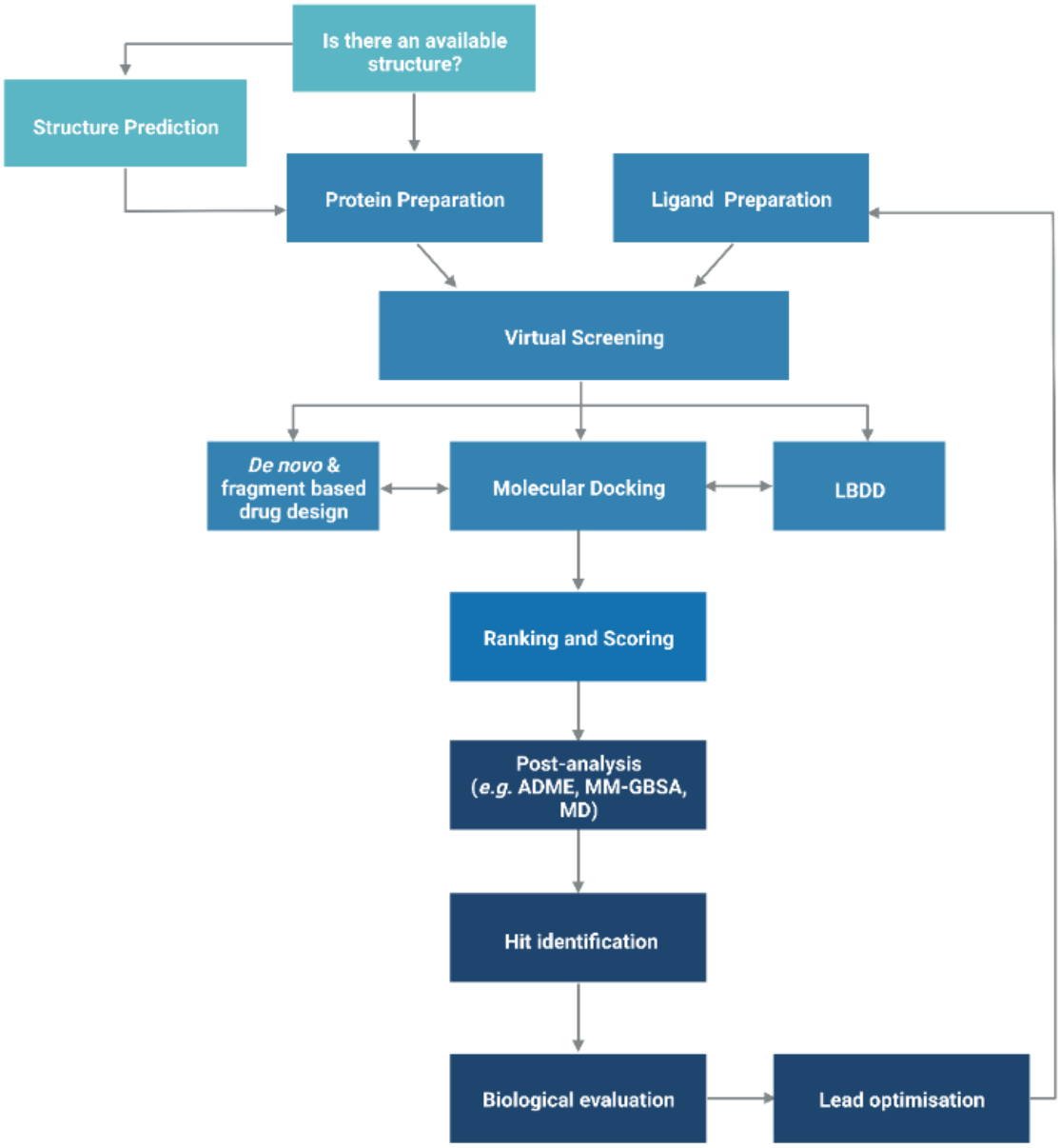
- template-based
- geometry-based
- energy-based
- hybrid methods, such as ConCavity and MPLs-Pred, as well as ML methods, such as DeepSite, Kalasanty, and DeepCSeqSite

Beyond locating the orthosteric binding site, these tools can identify allosteric binding sites to modulate protein function, hot spots on protein surface to alter protein–protein interactions and also analysing known binding sites to design better molecules that complement the binding pocket

In addition, the evaluation of its potential druggability is equally important:

- seq-based approaches
- structure-based (MAP$_{POD}$ score, Dscore, Drug-like Density (DLID), DrugPred, DoGSiteScorer, FTMap, and PockDrug)

> we can predict the druggability of a pocket with ML (SimpleScore)
> they have used DoGSite, along with FTMap, CryptoSite



molecular docking pipeline

3. Ligand-based DD

- if limited structural knowledge - features (chemical info etc.) is drawn from known ligands
- we can find structures and functions of similar proteins/chemicals
- Quantitative Structure-Activity Relationship (QSAR) for correlation of molecular properties
- structure-based and ligand-based pharmacophores (concepts of a minimal compound that is responsible for activity)

- Scaffold hopping (lead hopping) is a technique that identifies iso-functional molecular structures with significantly different molecular backbones

4. De novo and Fragment-based DD

5. Hierarchal Virtual Screening (HLVS)

- unlike docking studies where there are well established scoring functions used to approximate binding affinity and to rank molecules, pharmacophore methods lack a reliable and general scoring system

6. Molecular Mechanical/Generalised Born Surface Area (MM-GBSA)

- force-field based method that computes the free energy of binding from the difference between the free energies of the protein, ligand, and the complex in solution
- he free energy is calculated by using a combination of gas-phase molecular mechanics (MM) energy, electrostatic solvation energy (GB) and non-electrostatic contribution to solvation energy (SA).
- more accurate prediction because it can treat both the ligand and protein as flexible, allowing structural rearrangements required for the induced-fit pose
- MM-GBSA is more computationally expensive compared to conventional docking studies, therefore they are generally implemented after a completed docking study to re-score selected ligands

7. Molecular Dynamics

- time-costly - only for specifically selected molecules
  (1) to provide dynamic structural insights of biomolecules and
  (2) to provide precise energetic information of receptor–ligand complexes, key information in lead identification and lead optimisation
- There are two main hypotheses of the ligand recognition: conformational selection and induced fit mechanism, which may coexist in most cases
- MD combined with ensemble docking is one solution to address receptor flexibility by conducting simulations to explore the conformational space and select

representative conformations as a receptor ensemble into following dockings

- This method is usually integrated in virtual screening workflows to enrich the structural diversity of lead candidates and possible rational binding pose

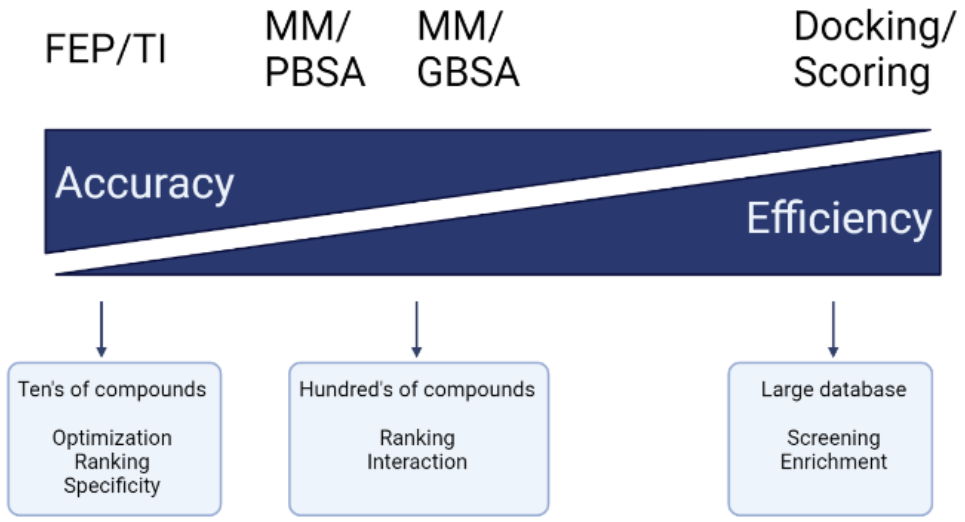


Figure 14. The distribution of the regular free energy calculation methods in accuracy/efficiency scale and their applications in drug discovery.

8. QM/MM and DFT Approaches

- factors are a consequence of the electronic interactions within a system

...

9. Examples in Conclusion

- **drug repurposing** 442 443
- **inverse docking** 444
- **guiding chemical synthesis** 442 443
- **AI** (for molecular docking 445 446, property prediction 447 448, compound retrosynthesis 449 450 451, de novo drug design 452 453)