

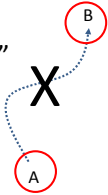
## Modelling of ligand-protein binding

### II. Approximate methods for estimating thermodynamic quantities

1

## Outline

- Calculation of binding free energy by „endpoint” methods
- $\Delta G = G_b - G_A$ 
  - $G/\Delta G$  cannot be accurately calculated
  - Approximate methods:
    - MM-PBSA (Molecular Mechanics Poisson – Boltzmann Surface Area) – not discussed
    - Docking and scoring



2

2

## Docking and scoring

3

3

## Scoring function

- Crude estimation of ligand-protein binding free energy
- Free energy vs. scoring
- Very fast – (several) ligand(s)/second
- Typically a single configuration is considered
- Accompanied by docking
  - Generating the structure of complexes using minimal preliminary information

4

4

## Scoring functions

- Types
  - Force field based
    - Molecular mechanics force field
  - Empirical
    - Sum of localized interactions
  - Knowledge-based
    - Based on the analysis of structural databases (Protein Data Bank, Cambridge Structural Databank)
  - Mixed
    - Combination of the types above

5

5

## Force field based scoring function

- Calculation of gas-phase energy  
( $\leftrightarrow$  free energy in solvent)
- Protein field can be precomputed on a grid  $\rightarrow$  increased computational speed
- Structure optimization possible
- Can be complemented with
  - Solvent effect
  - entropy (?)

6

6

## Empirical scoring functions

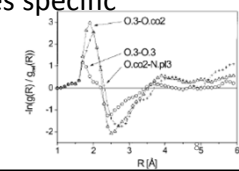
- Intuitive selection of interaction terms
  - Hydrogen-bond
    - Weighed sum of type dependent terms
  - Ionic interaction
  - Hydrophobic interaction
    - Proportional to the contact
- Parameters are fitted to experimental affinities
- „Sees“ only terms included in the model
- Local interactions

7

7

## Knowledge-based scoring function

- Derived from the statistical analysis of experimental structural data
  - $E_i = -kT \ln(p_i)$  – energy  $\sim$  observed frequency
- Protein Data Bank: over 147000 structures in December 2018
- Binding affinity data not required
- Long-range sampling – solvent effect included
- Short-range sampling – emphasizes specific interactions
- Incomplete repulsion



8

## Docking - scoring

- Generating and ranking ligand-protein complex structures
  - Single ligand-protein pair
    - finding binding mode
  - Multiple ligands and a single protein
    - Virtual screening
      - Binding mode identification
      - Ranking ligands by docking score
- Without preliminary structural information (in principle)
- Application in pharmaceutical research – see later

9

9

## Approximations of docking-scoring

Selected approximations:

- Protein is rigid or has limited flexibility
- Protonation state
- Interaction with and structure of water
- Entropy
- Temperature
- ...

10

10

## Protein flexibility – docking-scoring

- Role of protein flexibility in ligand binding
  - Selection of protein conformation advantageous for ligand binding
    - Population shift
  - Induced fit
    - Binding to a protein conformation not available for the free protein
  - No strict distinction between the above two mechanisms

11

11

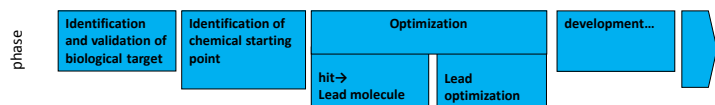
## Protein flexibility – docking-scoring

- Docking with taking into account protein flexibility
  - Using multiple static protein structures
    - Experimental structure – complexes with various ligands, NMR
    - Structures generated by computation (MD, MC)
    - Increased computational requirements
  - „Soft” protein structure
    - Single averaged structure derived from several structures and containing damped interactions
      - Unable to describe large movements
      - Increased binding pocket
      - Mutually exclusive binding sites appear simultaneously
  - Protein conformations generated upon binding (eg. MD)

12

12

## Application of docking-scoring



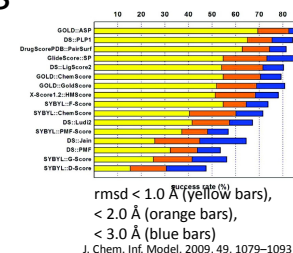
- Virtual screening
  - Identification of chemical starting point
- Docking – Binding mode identification
  - hit to lead

13

13

## Docking

- Protein structure
  - X-ray crystallography
  - homology model
- Ligand structure
  - Model
- Complex structure
  - Fitting the ligand into the protein binding pocket - docking
  - Ranking of binding modes using scoring functions
    - Limited protein flexibility
    - Efficient exploration of ligand conformational space
- RMSD of docked ligand < 2 Å – 70-80% in favourable cases



14

14

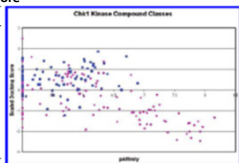
## Ranking

- Docking compounds into a protein and ranking the complexes (ligands) by scoring functions
- Studying similar compounds – lead optimization
- **weak correlation between score and experimental affinity**

Best Correlation Coefficient  $r$  between the -log Affinity (pAffinity) and Docking Score

program	Chk1	FXa
Dock4	-0.33	-0.31
DockIt	-0.49	-0.19
FlexX	-0.57	-0.31
Flex+	-0.44	-0.38
Fred	-0.14	0.01
Glide	-0.47	-0.08
Gold	-0.42	-0.05
LigandFit	-0.45	-0.13
MOEDock	-0.29	0.00
MVP	-0.26	0.10

J. Med. Chem. 2006, 49, 5912



Correlation Between the Scores and Experimental Binding Affinities

method	Pearson R	Spearman $\rho$
code 1	0.76 (0.80-0.71)	0.74 (0.79-0.68)
code 2	0.72 (0.77-0.66)	0.73 (0.78-0.67)
code 3	0.67 (0.72-0.60)	0.68 (0.74-0.61)
code 4	0.64 (0.70-0.58)	0.64 (0.70-0.56)
code 5	0.63 (0.69-0.56)	0.64 (0.71-0.57)
code 6	0.62 (0.68-0.55)	0.61 (0.68-0.53)
code 7	0.62 (0.68-0.55)	0.61 (0.68-0.53)
code 8	0.61 (0.67-0.54)	0.59 (0.66-0.51)
code 9	0.61 (0.67-0.53)	0.60 (0.67-0.52)
code 10	0.60 (0.66-0.52)	0.60 (0.67-0.52)
code 11	0.59 (0.66-0.52)	0.57 (0.64-0.49)
code 12	0.57 (0.63-0.49)	0.57 (0.65-0.49)
code 13	0.56 (0.63-0.48)	0.60 (0.67-0.52)
code 14	0.56 (0.63-0.48)	0.54 (0.62-0.45)
code 15	0.56 (0.63-0.48)	0.56 (0.63-0.47)
code 16	0.53 (0.60-0.45)	0.53 (0.61-0.44)
code 17	0.35 (0.44-0.25)	0.37 (0.46-0.27)

J. Chem. Inf. Model. 2011, 51, 2115

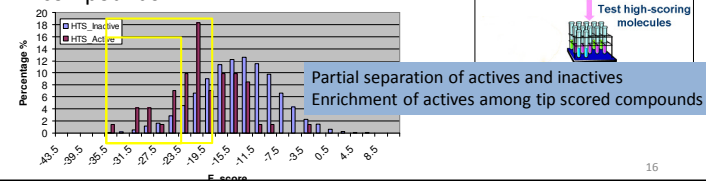
15

15

## Virtual screening

- Identification of chemical starting point
- Computation:
  - Docking a large number of structurally diverse compounds
  - Ranking the complexes (compounds) by score

- Experimental testing of top scored compounds



16

16

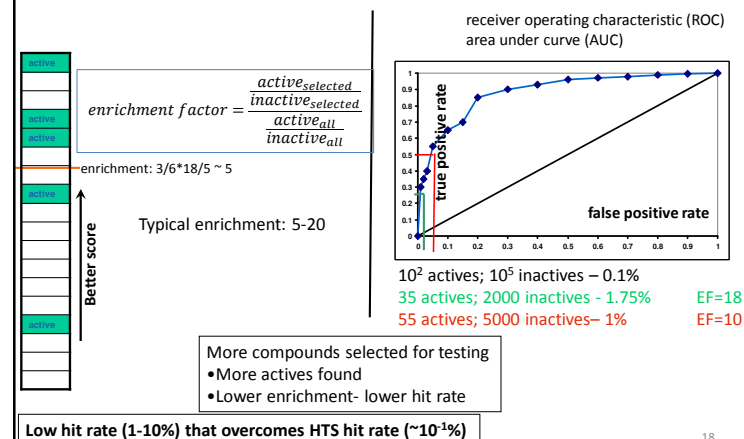
## Identification of chemical starting point and virtual screening

- High throughput screening (HTS) - experimental
  - Finding compounds with the required effect on a target protein
  - Biochemical/biophysical methods
    - receptor binding
    - Enzyme inhibition
    - ...
  - Testing  $10^5$ - $10^6$  compounds
  - Number of hits:  $\sim 10^2$
  - Hit rate: 0.1% ( $10^2/10^5$ )
- Virtual screening
  - Objective: increase HTS hit rate by computational (cheap) prescreening
  - Docking and scoring  $\sim 10^6$  compounds
  - Experimental testing of top  $\sim 10^3$  compounds; typical hit rate: 1-10 %

17 GF1

17

## Efficiency of virtual screening



18

18

## Summary

- Endpoint methods – Approximate schemes for estimating  $\Delta G$ 
  - MM-PBSA – not discussed
    - fast (lesser extent)
    - Fair correlation with experimental values
    - Well fitted to improved virtual screening results
  - Docking - scoring
    - Very fast
    - Good quality binding mode prediction
    - Weak correlation between score and experimental affinity
    - Virtual screening is an established tool in chemical starting point identification
    - Intensively applied in pharmaceutical research

19

19

