

KFC/STBI

Strukturní bioinformatika

05_docking

Karel Berka

Docking

- minule jsme probrali databáze makromolekul i malých molekul
- Dnes se podíváme, jak to spojit – tj. jak zjišťovat, která látka se bude vázat na který protein/DNA ...
- Řeší to tzv. **ligand docking**

Syllabus

- virtual screening
- docking
 - proč?
 - hledání molekul, které se váží do aktivních míst
 - čeho?
 - knihovny léčiv
 - jak?
 - generování konformací ligandů
 - skórovací funkce
 - a hlavně rychle!

Virtuální screening

- ekvivalent biologického screeningu
 - Které látky testovat?
 - Jaké knihovny látek prozkoumávat?
 - Které látky skutečně koupit?
- Častokrát se zkouší metody o různé komplexitě a náročnosti

Virtuální screening

- 4 druhy
 - znám jednu aktivní látku
 - hledám podobné (ligand-based virtual screening)
 - znám několik aktivních látek
 - hledám podobnosti – 3D farmakofor
 - znám jak aktivní, tak neaktivní látky
 - trénink neurální sítě a pak její aplikace na další látky
 - znám 3D strukturu proteinu
 - protein-ligand docking

Hledání základní struktury

- tzv. **lead compound**
- molekuly se v průběhu vývoje komplikují
 - různé vedlejší řetězce
- hledání nových strukturních motivů, na kterých se dá stavět je obtížnější

Shrnutí dockingu

- if znám 3D struktura proteinu:
 - použít informace o struktuře proteinu k tvorbě ligandu
 - předfiltrovat možné látky vůči podobnosti lékům
 - dokovat a ohodnotit skórovací funkcí
 - vybrat nejlepší struktury
 - snad i slavit úspěch...
- a teď si projdeme jednotlivé kroky podrobněji

Filtry pro látky

QSAR

quantitative structure activity relationship

použití deskriptorů

k odhadnutí funkce ze struktury

Filtr: Lipinskiho pravidlo pěti

- podobnost léčiv
- Ústně podávané léčivo
- max 1 porušení z pravidel:
 - Max 5 donorů vodíkových vazeb (NH, OH)
 - Max. 10 akceptorů vodíkových vazeb (N, O)
 - MW < 500
 - octanol-voda rozdělovací koeficient - $\log P < 5$

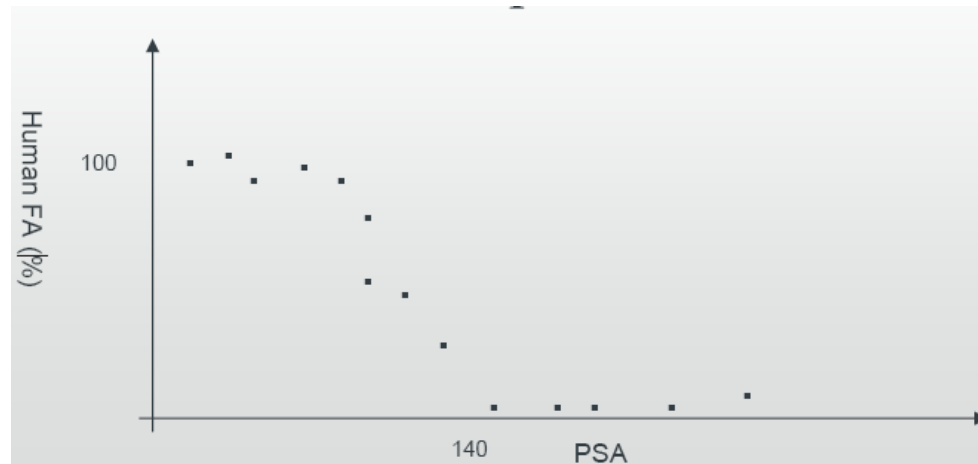
ADMET vlastnosti léčiv

Adsorpce, Distribuce, Metabolismus, Exkrece a Toxicita

- Molekula se musí nejen pevně vázat na protein, ale i
 - Musí být dostupná z podané lékové formy (orální nejlépe)
 - dostat se k receptoru (Blood-brain barrier, rozpustnost ve vodě, aktivní prostupnost membránami)
 - zůstat v těle dost dlouho k provedení účinku (CYP P450s!)
 - ale ne moc dlouho (vedlejší účinky)
 - Nesmí zabít pacienta (ani její metabolit) Molecule must not only bind tightly, but must:
- Problém – první fáze se nesmí testovat na lidech
 - nedostatek exp. dat
 - Caco-2 (tumor) cell lines jako model

ADMET – QSAR příklad

- Polární povrch molekuly (PSA)
 - Orální absorpce a průchod do mozku
 - Predikce průchodu do Caco-2 buněk
 - Vztah mezi PSA a procentem absorbované látky pro člověka

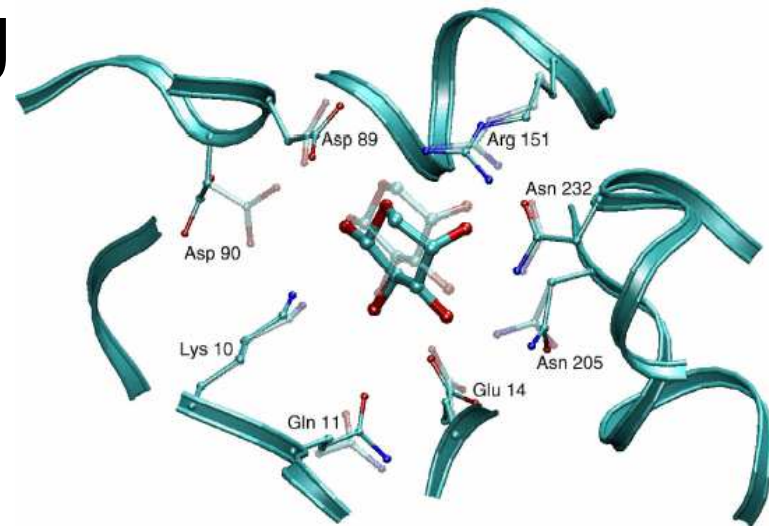


Docking

Hledání pózy
skórovací funkce

Structure-based virtual screening

- Protein-ligand docking
 - predikce 3D struktury komplexu ligandu s proteinem
- postup
 - 1) navrhnutí struktur
= geometrie
 - 2) ohodnocení a stanovení pořadí
= energie



Problém VS: Flexibilita

- První krok
 - prohledání možných rotačních a translačních možností ligandu
- Flexibilita ligandu
 - tvorba ensemble všech možných konformací ligandu
 - vkládání postupně do proteinu
 - protein je nejčastěji rigidní

Hledání

- Algoritmy
 - Monte Carlo
 - Genetické algoritmy
 - Postupný růst

- 70 % přesnosti v redokovacích studiích

Příprava na docking

- identifikace relevantní struktury proteinu
- příprava struktury proteinu
- příprava ligandu
- rozhodnout co s dalšími jevy
 - voda
 - flexibilita receptoru

Příprava receptoru

- Kam?
 - identifikace vazebného místa
- Dobrá struktura
 - nízké rozlišení (přesnost)
 - nízké B-faktory (flexibilita)
 - nízké R-free (pravdivost)
- Flexibilita
 - dokování do více struktur – vybrat nejlepší
 - dokování do struktury s největším vazebným místem
 - použít flexible docking

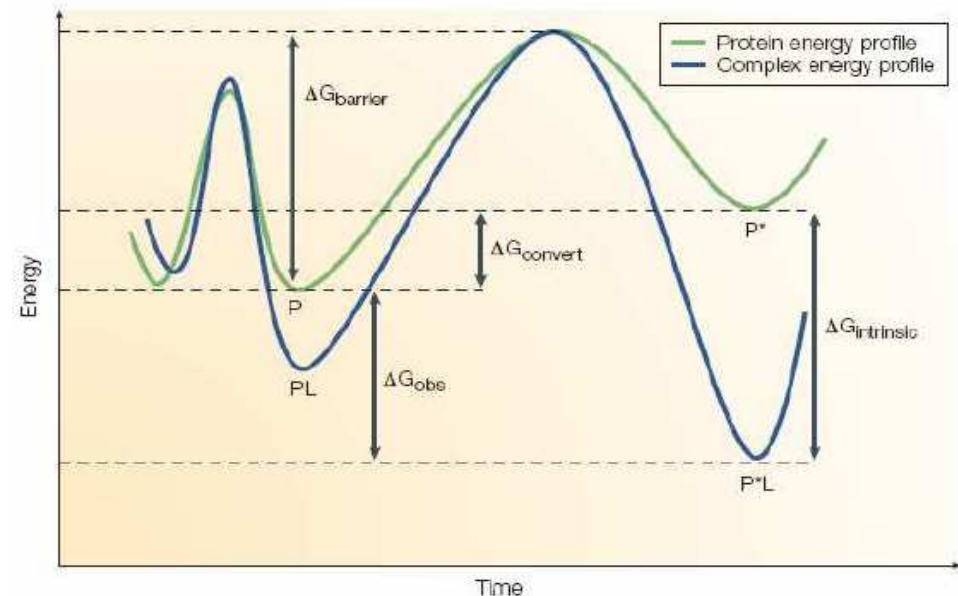
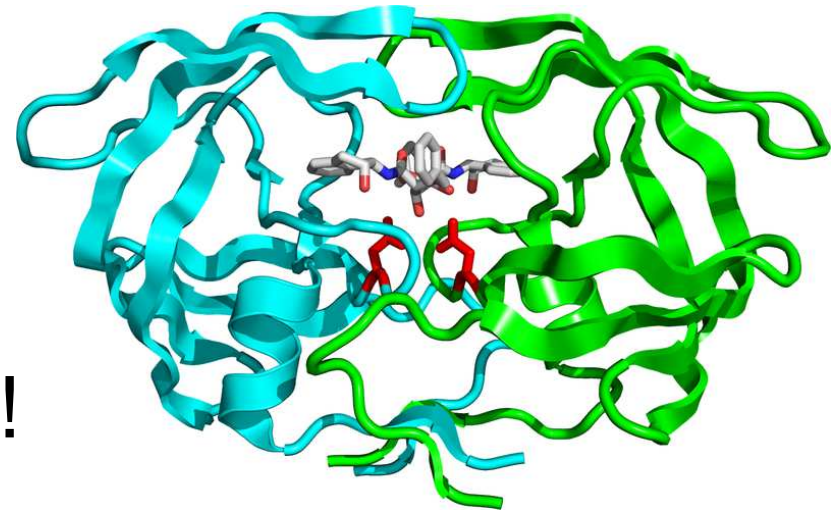
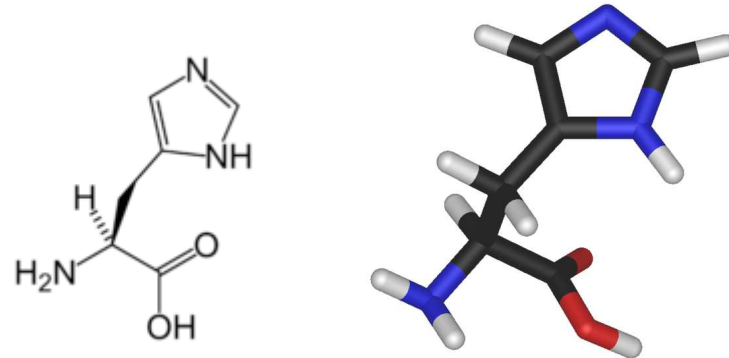


Figure 1 | **Protein mobility and ligand binding.** A protein is considered to exist in two conformations (P and P*) with an energy difference $\Delta G_{\text{convert}}$. The ligand (L) can bind the protein (P) to give a complex (PL), or bind to P* to give a complex (P*L). Although P* has a higher free energy, it might offer greater scope for interaction with L. For instance, P* might represent a conformer in which the binding site has opened and exposed hydrophobic patches. This is energetically unfavourable, but offers the potential for favourable interactions with the hydrophobic moiety of a suitable incoming L, thereby giving rise to a large, favourable interaction $\Delta G_{\text{intrinsic}}$. The resulting complex (P*L) has a lower energy than that of the complex PL. The observed affinity of L for the protein conformational ensemble is governed by ΔG_{obs} . Slow binding kinetics might well be observed, as P* is a higher-energy conformer than P and an energy barrier ($\Delta G_{\text{barrier}}$) must be surmounted before optimal binding to L can take place.

Protein flexibility

Příprava receptoru

- protonace residuí
 - His (pKa ~ 6.04)
 - záleží na okolí (HIV proteáza)
- tautomerizace
- rotamery
- výběr změní výsledky!



Příprava ligandu

- Náboj a tautomerizace
 - Vytvořit všechny a pak je dokovat?
 - jak pak vybrat nejvyšší skóre
 - každý stav bude mít jinou relativní energii.
 - Ask an expert! (organičtí chemici)
- flexibilita
 - rotace kolem C-C vazeb, ale ne kolem C=C
 - úhly jsou fixovány (což může být problém)

Skórovací funkce

1. Ohodnocení jednotlivých póz v průběhu dockingu – **objective function**
 2. Identifikace nejnižší volné energie
 3. Seřadí **vazebné volné energie** mezi různými ligandy
- Pro všechny kroky nemusí být vždy použita stejná funkce

Energetika vaznosti

- Vazebná konstanta

$$K_d = [P...L] / [P][L]$$

odpovídá volné energii:

$$\Delta G_{\text{bind}} = -RT \ln K_d = -RT \ln 1/IC_{50}$$

Volná energie je kombinace enthalpie a entropie

$$\Delta G_{\text{bind}} = \Delta H_{\text{bind}} - T\Delta S_{\text{bind}}$$

Molekulární interakce - chemie

Enthalpie:

- Elektrostatika
 - (dielektrická konstanta)
- Vodíkové vazby
- van der Waals
 - (disperze a repulze)
- Desolvatace

Entropie

- Konformace
- Solvatace
 - (hydrofobní efekt)

Scoring function

- Funkce, která ohodnotí sílu vazby

$$\Delta G_{bind} = \Delta G_{solvent} + \Delta G_{conf} + \Delta G_{int} + \Delta G_{rot} + \Delta G_{t/r} + \Delta G_{vib}$$

- Empirické

$$\Delta G_{bind} = \Delta G_0 + \Delta G_{hb} \sum_{H-bonds} f(\Delta R, \Delta \alpha) +$$

$$\Delta G_{ionic} \sum_{ionic_interactions} f(\Delta R, \Delta \alpha) + \Delta G_{lipo} |A_{lipo}| + \Delta G_{rot} NROT$$

- Parametrizovány proti známým vazebným volným energiím
- Většinou jsou nepříliš spolehlivé, ale rychlé

Typy scoring function

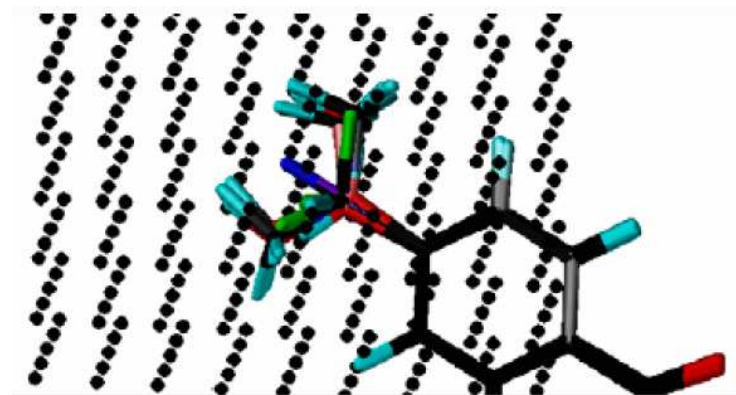
- First-principles
 - molecular mechanics force fields
 - DOCK, Autodock 1.0
- semiempirical
 - nutnost parametrizovat pro daný cíl na základě exp. dat (K_d , IC_{50})
- empirical
 - různé členy (Hbonds, hydrophobic contact)
 - nutné trénování, rychlé
 - GOLDscore, ChemScore
- knowledge based
 - statistické – vycházejí ze znalostí komplexů protein-ligand
 - Boltzmann hypotéza – co se často vyskytuje, to má silnou vazebnou volnou energii

Force-field scoring functions

- Silové pole pro docking

$$E = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dih}} + E_{\text{coulomb}} + E_{\text{vdw}} + E_{\text{solv}}$$

- časté použití gridu – šetří čas
 - vyhledání v tabulce je jednodušší, než znovu počítat interakční energie
 - receptor je často jeden a dokovaných látek je mnoho



Empirical scoring function

- Rozložení vazebné energie do chemicky odpovídajících členů
- Zachycuje specifické interakce
 - H vazby, π - π stacking

Linearizace členů je oblíbená

$$DG_{bind} = DG_{solvent} + DG_{conf} + DG_{rot} + DG_t + DG_r + DG_{vib}$$

Chemscore

- Original Chemscore function for binding free energies

$$\Delta G_{binding} = \Delta G_o + \Delta G_{hbond} S_{hbond} + \Delta G_{metal} S_{metal} + \Delta G_{lipo} S_{lipo} + \Delta G_{rot} H_{rot}$$

- S_{hbond} – hydrogen bonding
- S_{lipo} – lipophilic interactions
- S_{metal} – acceptor-metal interactions
- H_{rot} – loss of conformational entropy on ligand binding

J. Comput. Aided Mol. Des. 11, 425-445, 1997

Chemscore

Chemscore for docking

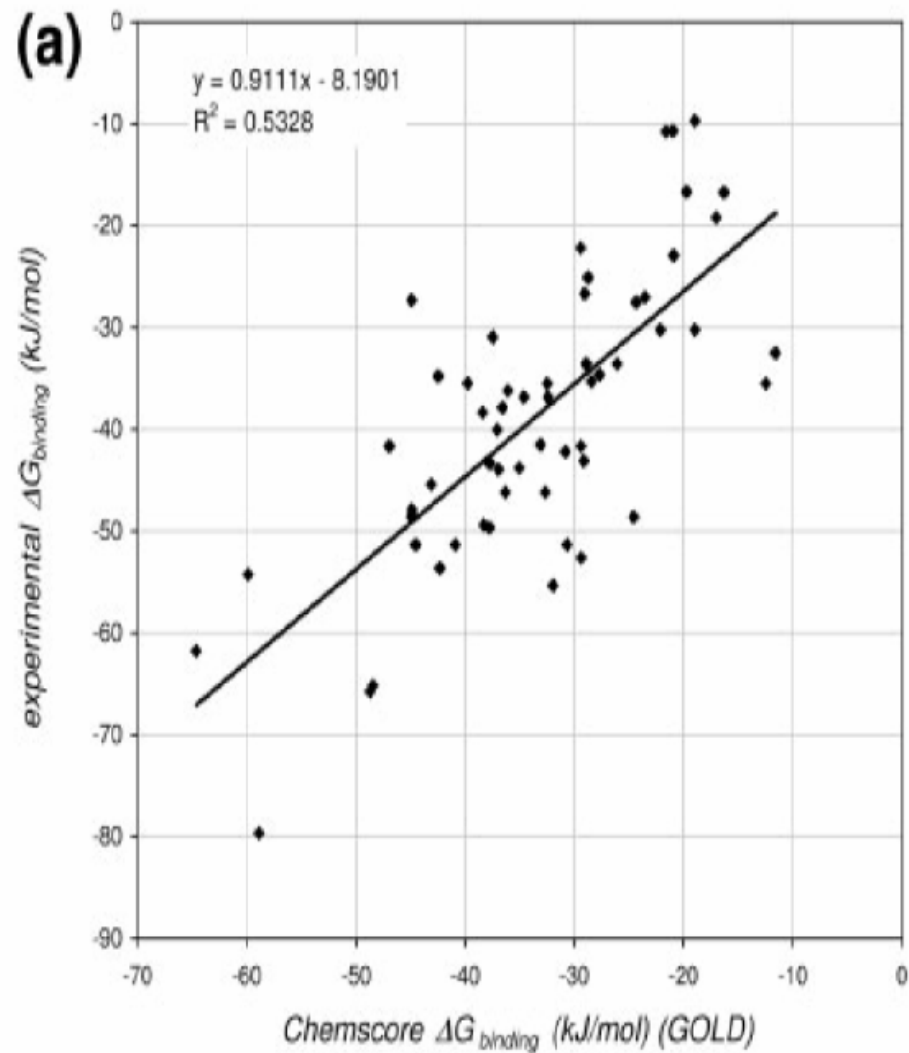
- Add more terms – clash, ligand internal, protein-ligand covalent

$$\Delta G'_{binding} = \Delta G_{binding} + E_{clash} + E_{int} + E_{cov}$$

- Complex functional forms – look them up!
- Parameters carefully rederived

Proteins 52, 609-623, 2003

Chemscore úspěšnost



Empirical troubles

- Záleží na testovacím setu
- chybějící interakční členy
 - metal-ion
- Parametrizovány na úspěch
 - only molecules which bind are used, so all binding ability over-emphasised

Knowledge-based functions

Korelace strukturních dat z komplexů s volnou energií vazby

- Use a rigorous statistical mechanical result:

$$A = -kT \ln g(r)$$

- This equation holds for an ensemble of particles at equilibrium
- ne nutně pak pro proteiny

Drugscore

DRUGSCORE

$$\Delta W_{i,j}(r) = W_{i,j}(r) - W(r) = -\ln \frac{g_{i,j}(r)}{g(r)}$$

$$g(r) = \frac{\sum_i \sum_j g_{i,j}(r)}{i*j}$$

Short-range (6 Å) contributions only – ignoring solvation

Programy

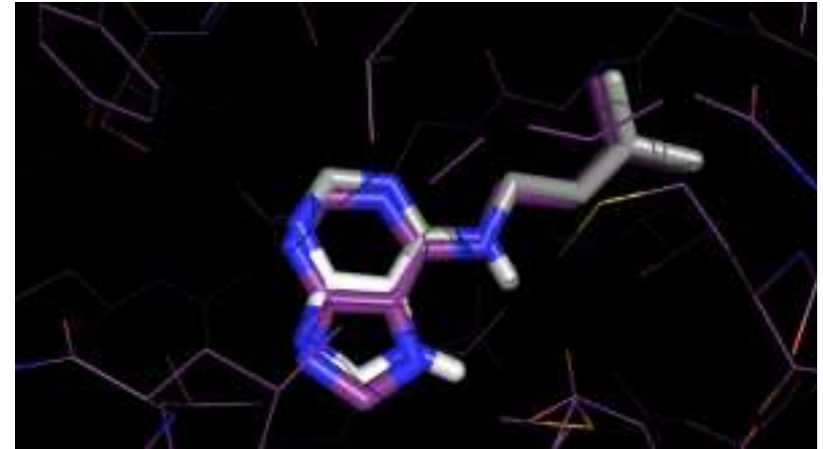
- DOCK
- AUTODOCK
- GOLD
- FlexX
- Glide

Úspěch?

hodnocení kvality

Kvalita

- Redocking zpátky do krystalu
 - RMSD < 2Å
 - flexible ligand docking ~70%
 - nutno otestovat, který docking program dává pózu dobře
- test sety – validace programů
 - GOLD test set, Astex set
 - decoys – ZINC, DUD (podobné fyz.chemické vlastnosti, odlišné struktury)
- VS: obohacení
 - Enrichment factor
 - (BED)ROC křivky
 - ΔG_{eff}

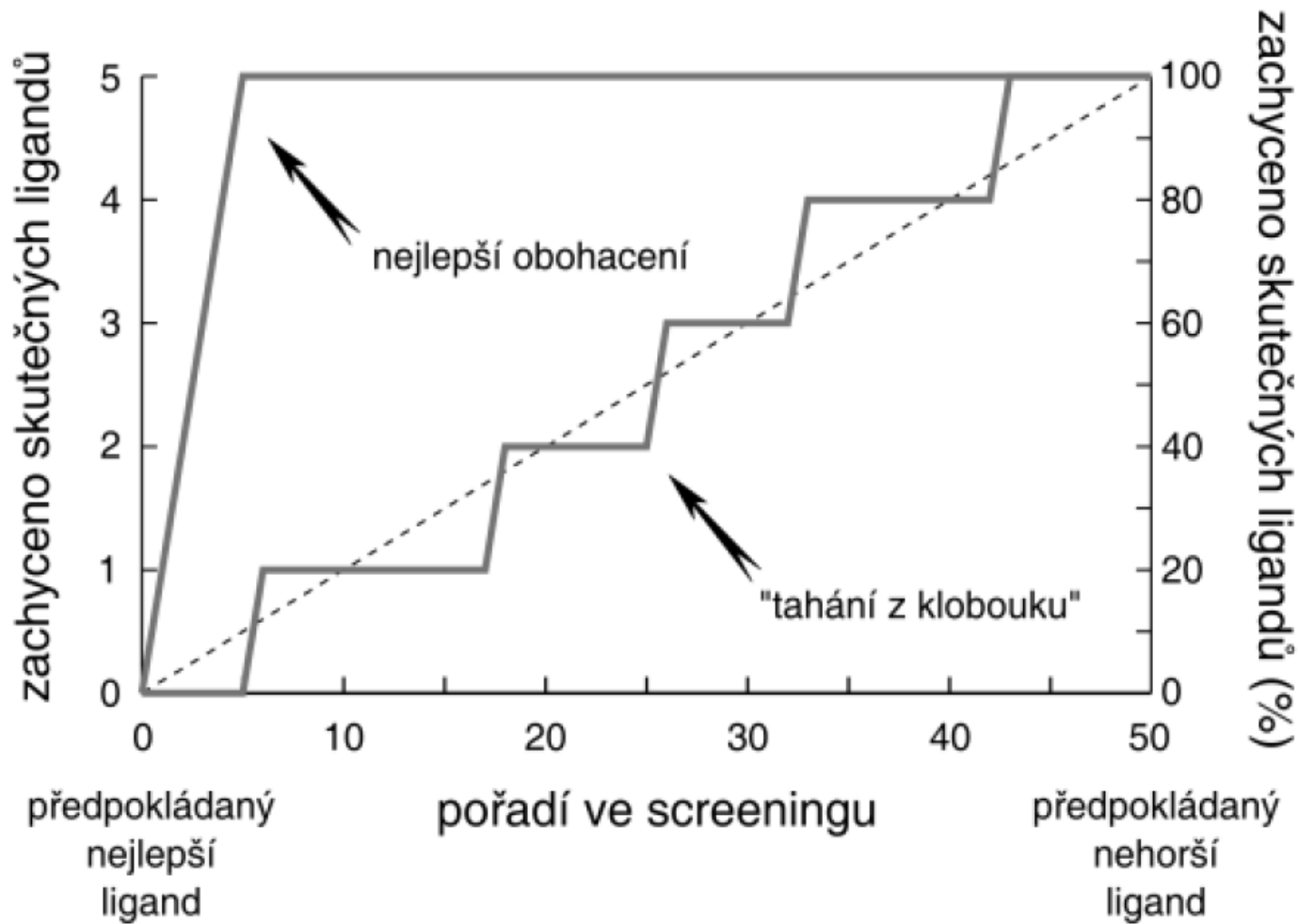


$$EF = \frac{a/n}{A/N}$$

- top (např. top10)
a – aktivní
n - celkem
- total

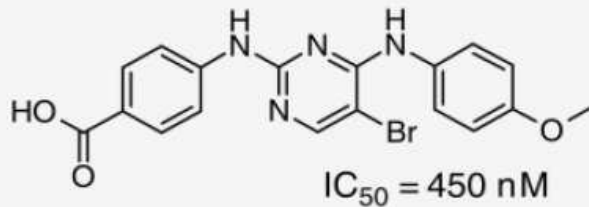
$$\Delta G_{\text{eff}} = \Delta G_{\text{eff}} / N_{\text{nonHatoms}}$$

Kvalita virtuálního screeningu



Příklady použití

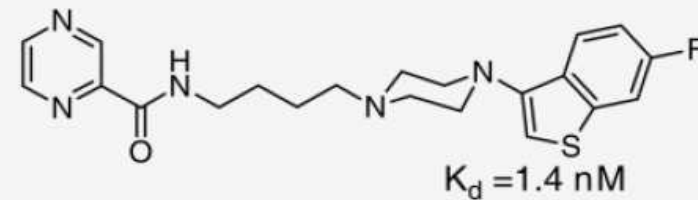
Checkpoint-kinasa 1



560 000 slouč.
↓ výběr léčivům podobných slouč.
199 000 slouč.
↓ FlexX-Pharm docking
250 sloučenin
↓ ruční výběr
103 testováno, 36 úspěšně

J. Med. Chem. 47, 1962 (2004).

α_{1A} adrenergní receptor

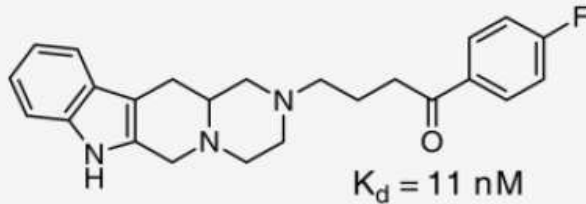


mnoho sloučenin
↓ výběr léčivům podobných slouč.
22 950 slouč.
↓ GOLD docking
300 sloučenin
↓ statistická analýza
80 testováno, 37 úspěšně

J. Med. Chem. 48, 1088 (2005).

Příklady použití

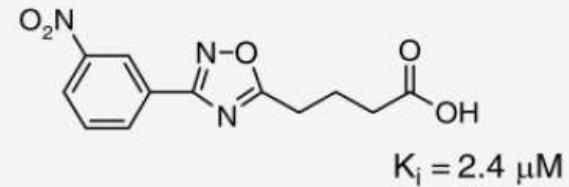
Dopaminový D3 receptor



250 251 slouč.
↓ 3D-farmakofor
6 727 slouč.
↓ Cerius² docking
2 478 sloučenin
↓ výběr podle podobnosti
20 testováno, 8 úspěšně

J. Med. Chem. 46, 4377 (2003).

Aldosareduktasa

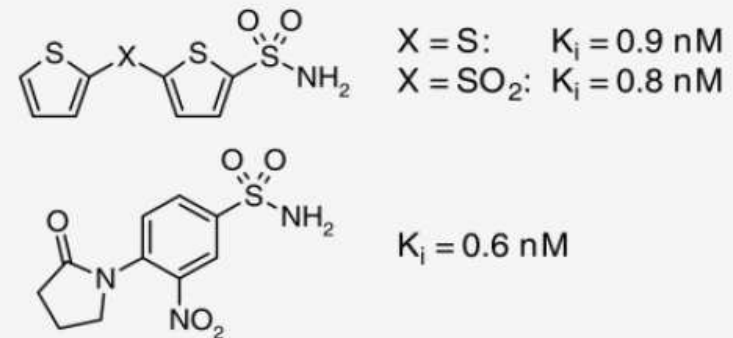


259 757 slouč.
↓ výběr léčivům podobných slouč.
12 545 slouč.
↓ 3D-farmakofor
1 261 sloučenin
↓ FlexX docking
9 testováno, 1 úspěšně

Proteins 55, 814 (2004).

Karbonáthydrolyasa

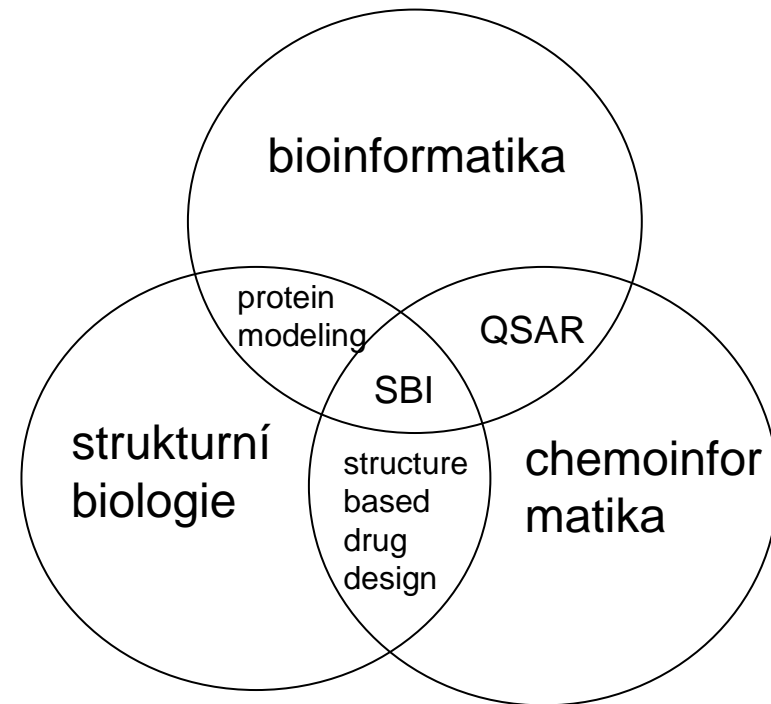
98 850 slouč.
↓ výběr podle vazby Zn^{2+}
5 904 slouč.
↓ 2D a 3D-farmakofor
3 314 sloučenin
↓ FlexS/FlexS docking
13 testováno, 3 úspěšně



J. Med. Chem. 45, 3588 (2002).
Angew. Chem. Int. Ed. Engl. 40, 389 (2001).

Drug discovery

- Drug targets
 - 70 % drug targets – 10 protein families (CATH, SCOP)
 - 50 % - 4 rodiny: GPCR, nuclear receptors, ligand, and voltage-gated ion channels
 - CATH: 130 druggable domains
 - tvorba homologních modelů ze sekvence



Protein-protein docking

řádově složitější,
méně dat

Protein-protein docking

- rigidní/flexibilní



- kontrola – mutační studie
- komplementarita ploch

Programy

- **CLUSPRO** 3D-DOCK DOT
- BIGGER HADDOCK MOLFIT
- PatchDock **RosettaDOCK** ZDOCK

- CLUSPRO

- cluspro.bu.edu

- RosettaDOCK

- rossettadock.greylab.jhu.edu

- CAPRI Critical Assessment of PRotein Interactions

- <http://capri.ebi.ac.uk>

- nejlepší – CLUSPRO – 5%

Poděkování

- Jonathan Essex, University of Southampton, UK