

Docking (dokování)

Zdeněk Kříž

Národní centrum pro výzkum biomolekul
Přírodovědecká fakulta MU, Brno

Struktura a funkce biomolekul – Brno, 4. - 8. září 2006

Co je docking?





... a nyní vážně ...



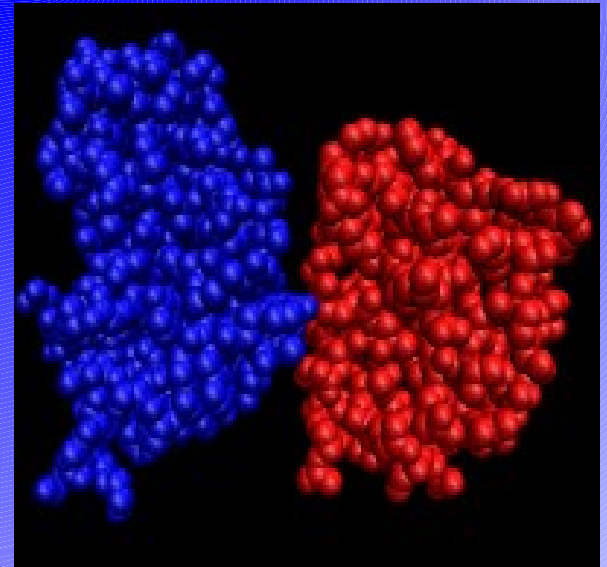
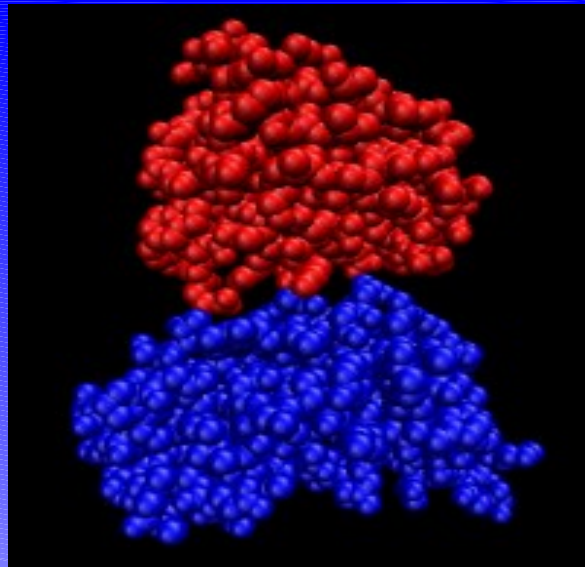
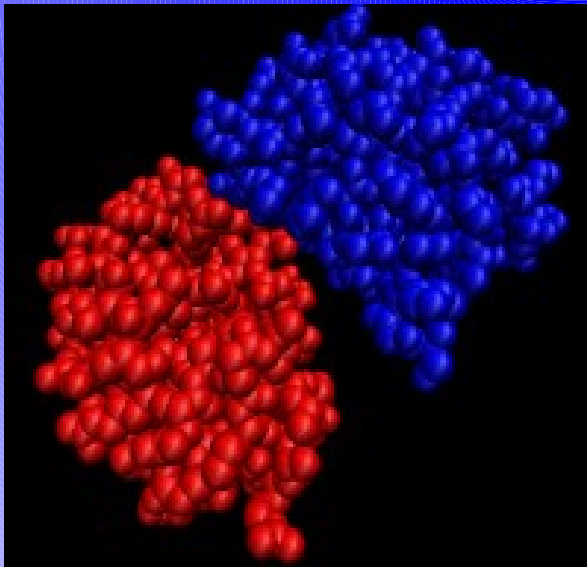
- Metoda, která pro dvě různé molekuly najde:
 - Uspořádání, v jakém spolu mohou tyto dvě molekuly interagovat.
 - Pokud takové uspořádání existuje, najde nejvhodnější orientaci obou molekul z hlediska největší komplementarity a z hlediska maximálního počtu nevazebných interakcí a nejmenší energie celého systému.
- Hlavním cílem je pro existující databáze molekul najít automaticky ty molekuly, které spolu mohou interagovat.

K čemu je dokování užitečné ...

- Molekulární biologie – vysvětlení funkce proteinových komplexů mezi sebou a komplexů proteinů s ostatními molekulami
- Výzkum léčiv – hledání vhodných inhibitorů určitých enzymů a návrh vysoce selektivních a účinných léků.

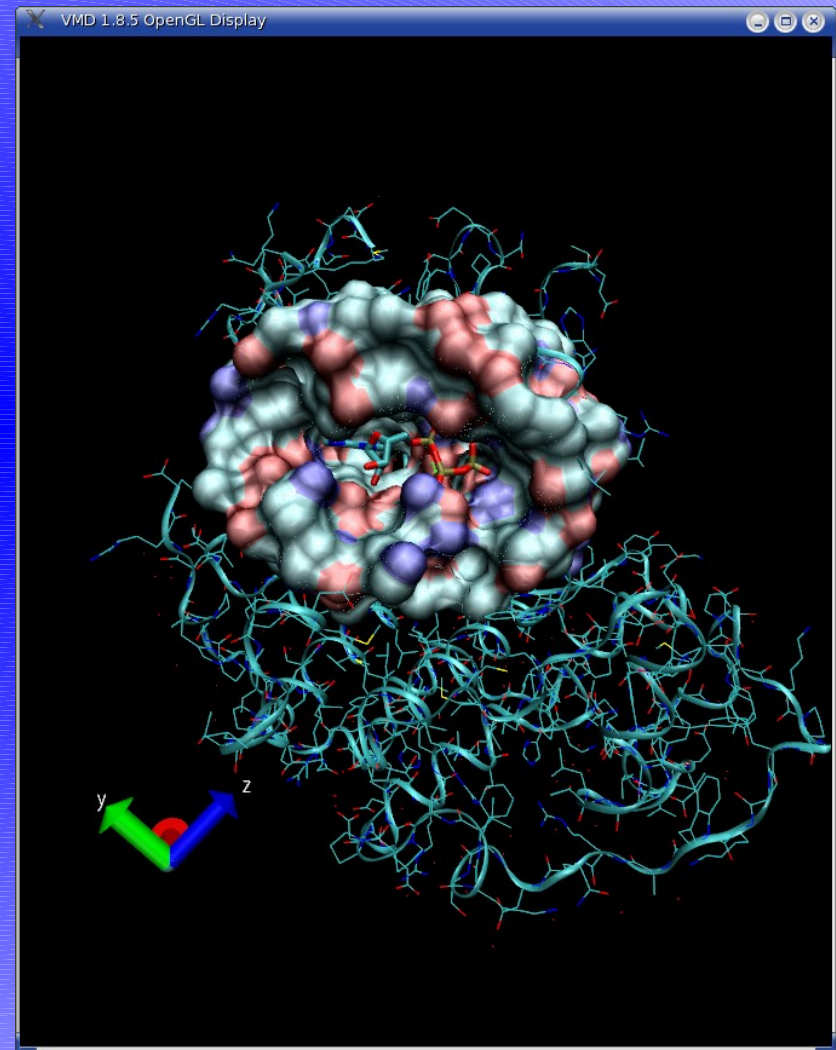
Proč je to tak složité?

- Obě molekuly jsou flexibilní a mohou tedy “měnit” svůj tvar.
- Mnohdy neznáme oblasti interakcí těchto dvou molekul.
- Někdy spolu neinteragují pouze dvě, ale více molekul.
- Existují miliony možných uspořádání, které je třeba prozkoumat.



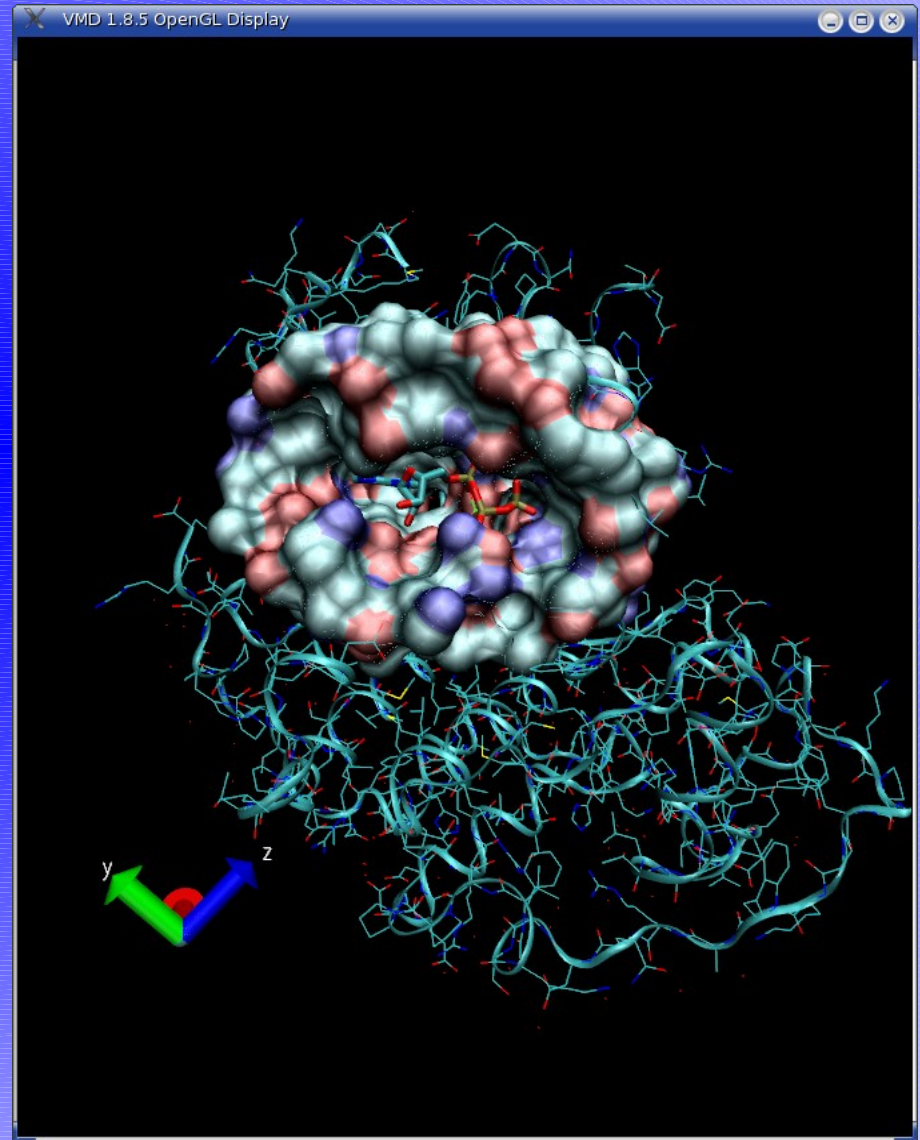
Metody dokování

- Dokování rigidního ligandu do rigidního receptoru
- Dokování flexibilního ligandu do rigidního receptoru
- Dokování flexibilního ligandu do “flexibilního” receptoru
- Dokování dvou proteinů

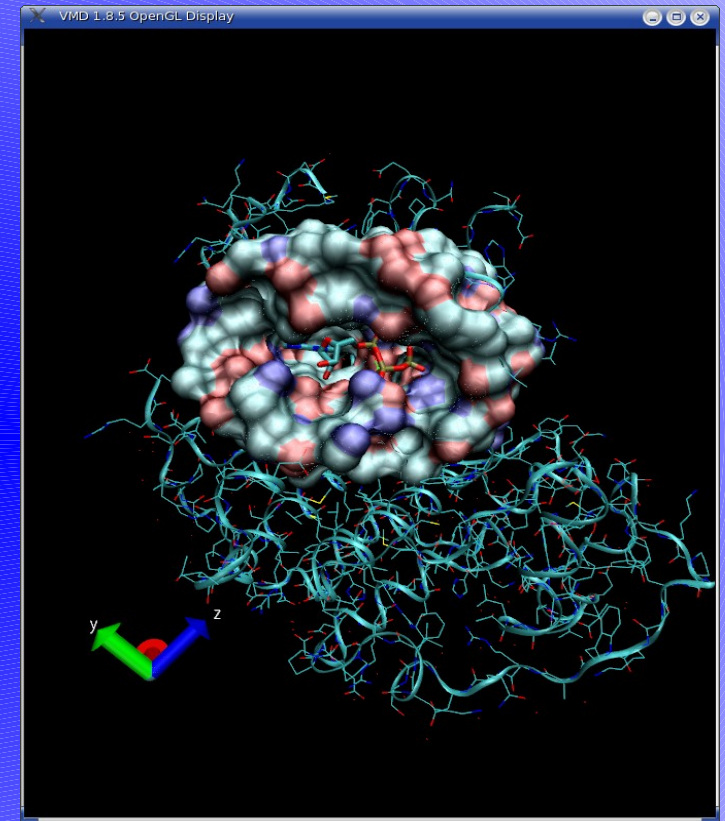
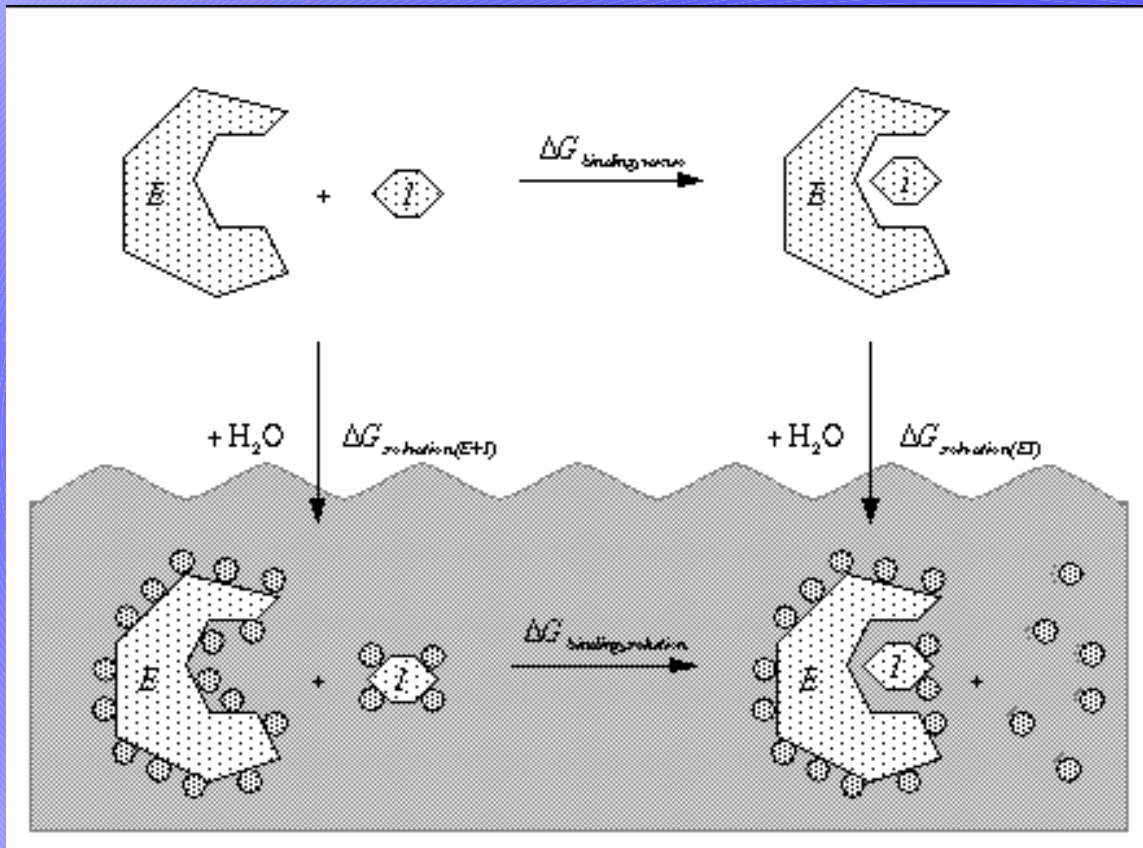


Metody dokování - aproximace

- Používáme předpočítaný grid pro interakce a interakční energie
- Nepočítáme interakce pro celý protein, ale jen pro aktivní místo a ligand
- Ligand uvažujeme jako molekulu složenou z rigidních částí spojenou flexibilními můstky
- Konformační změny neprovádíme systematickými metodami, ale stochastickými – MC, genetický algoritmus
- Nejprve rozhodujeme na základě chemické komplementarity a až potom počítáme interakční energii.

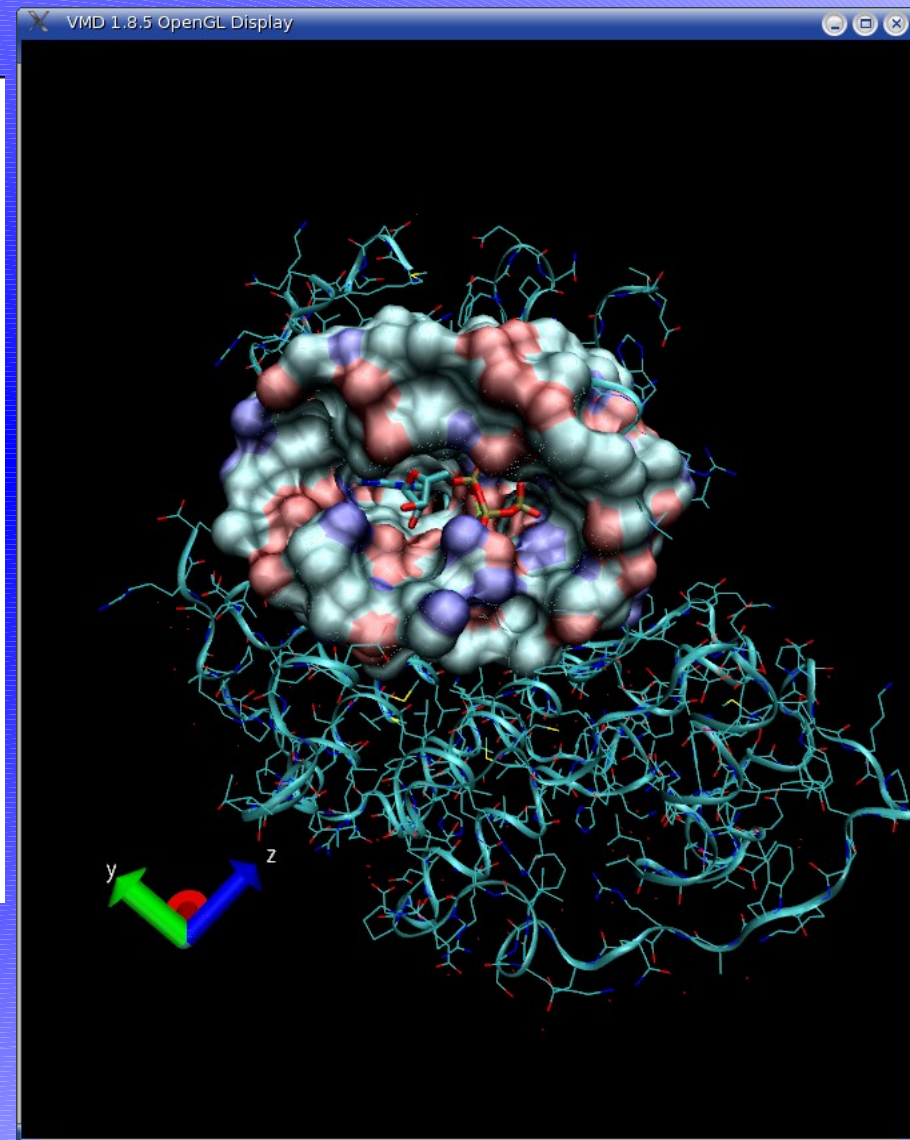
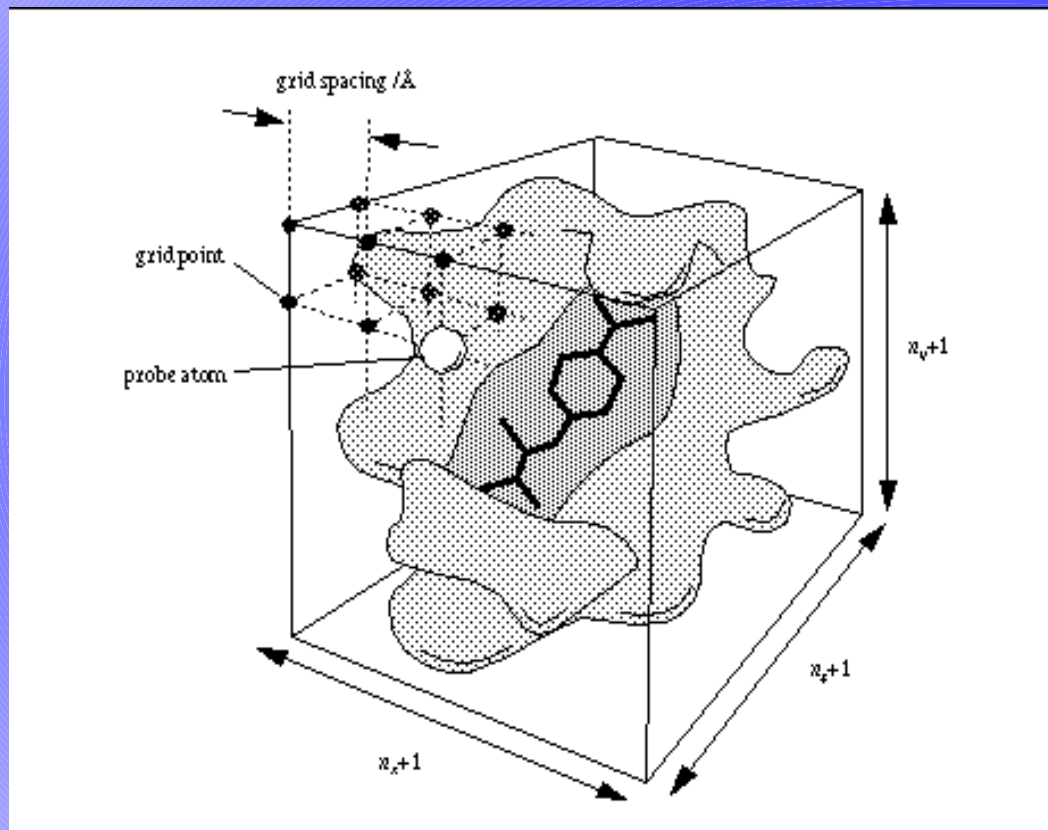


Metody dokování - aproximace

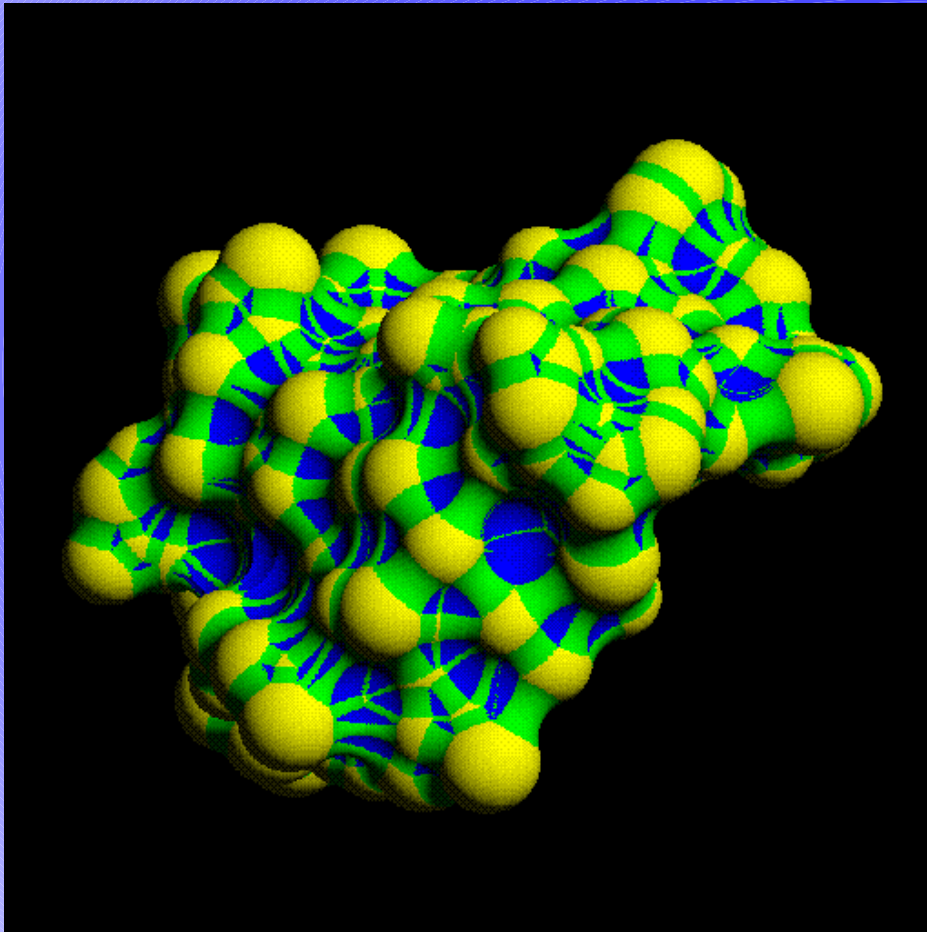


$$\Delta G_{\text{binding/solution}} = \Delta G_{\text{binding/vakuo}} + \Delta G_{\text{solvation}(E/I)} - \Delta G_{\text{solvation}(E+I)}$$

Metody dokování - grid



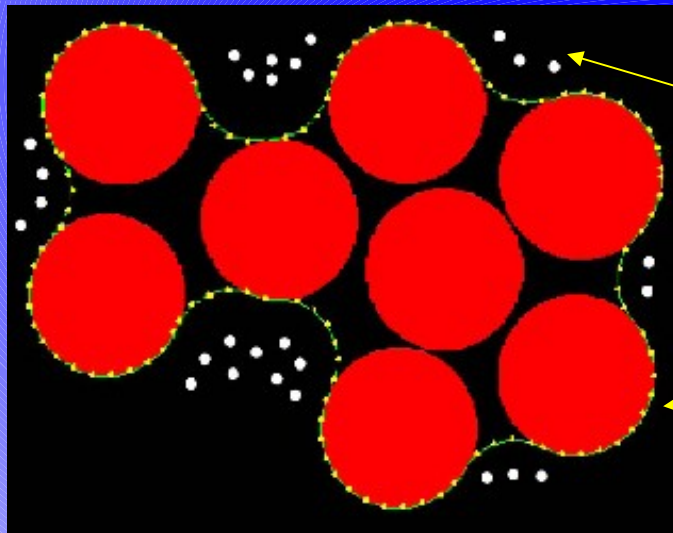
Dokování – povrch molekuly



- Každý atom je reprezentován jako koule o Van der Waalově poloměru
- Válením koule po vdw povrchu vytvoříme Connollyho povrch

Dokování – Lenhoffova metoda

Místo Connollyho povrchu počítá tzv. komplementární povrch pro receptor. Hledá možné pozice pro atomy ligandu na základě vdW povrch receptoru.

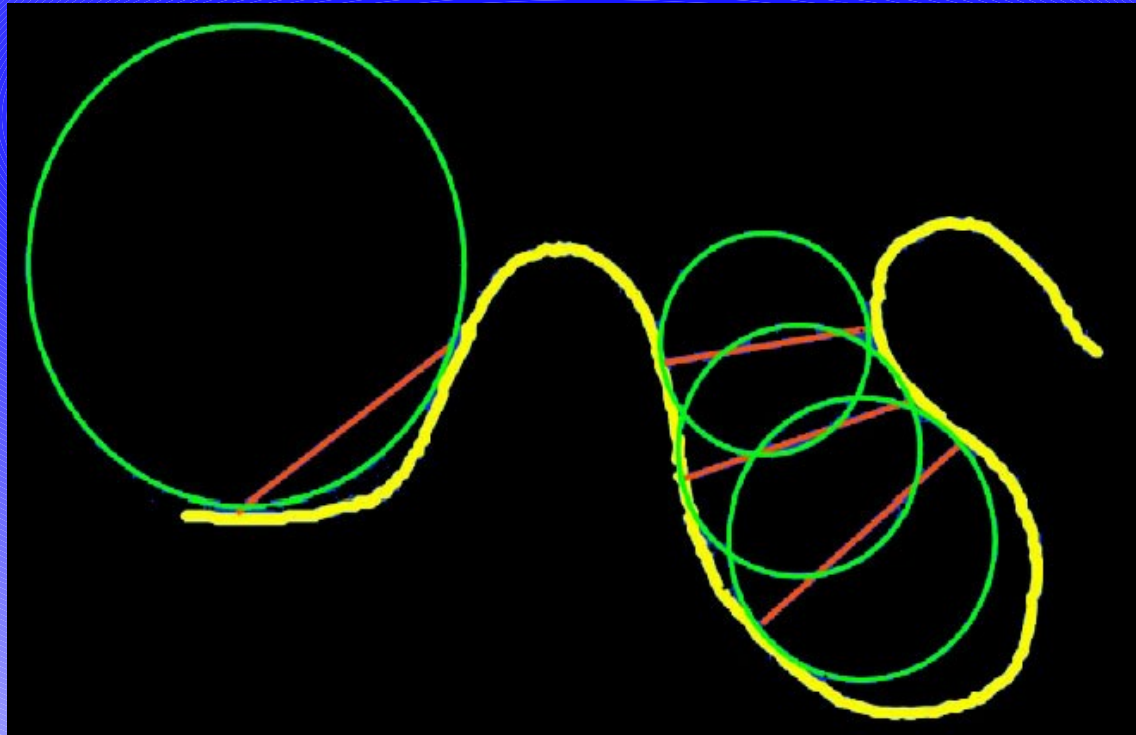


Možné pozice atomů ligandu

van der Waalsovský povrch

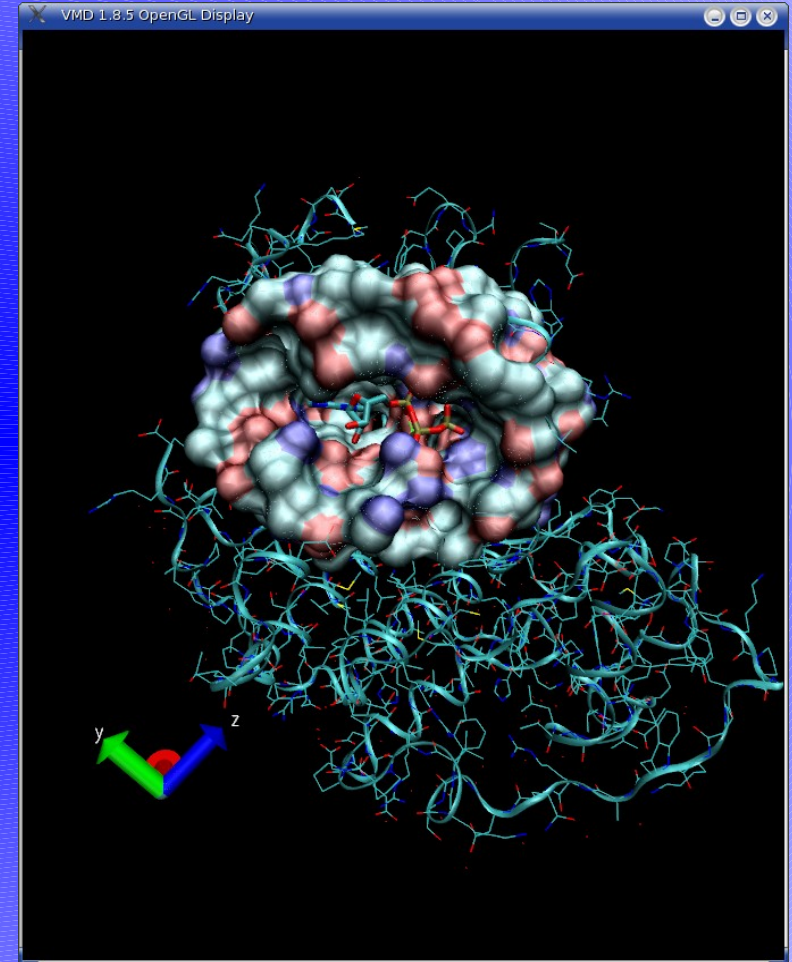
Dokování – Clustered spheres

Hledání možných kavit v proteinu na základě překrývajících se sfér vygenerovaných na povrchu proteinu.



Dokování – používané algoritmy

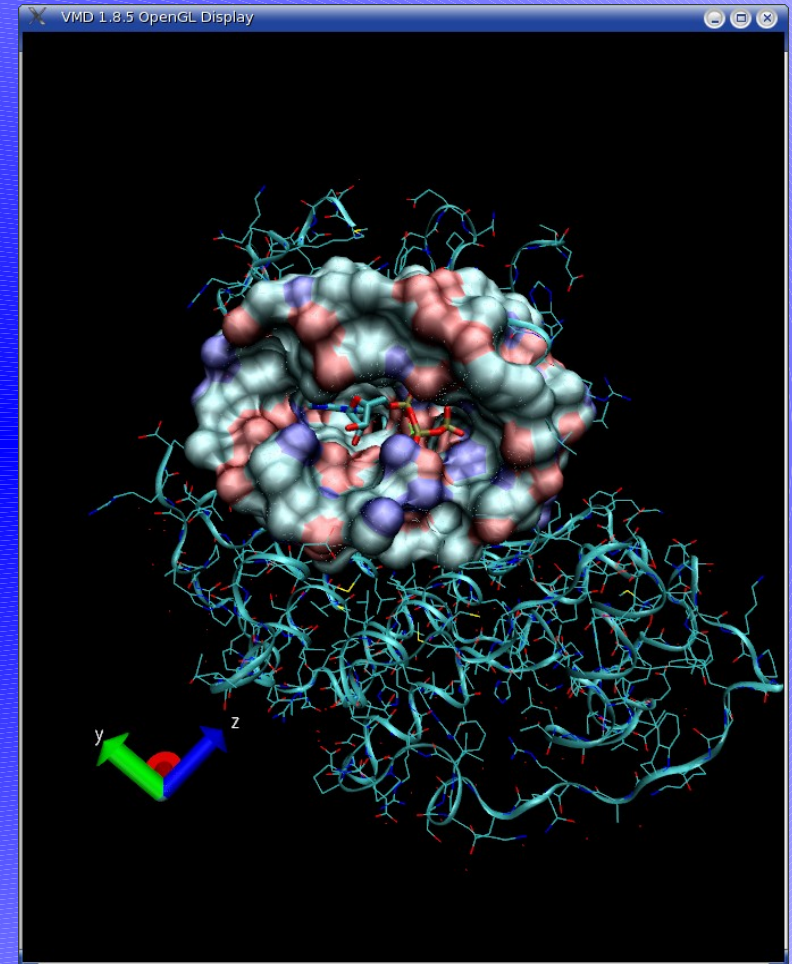
- Manuální dokování
- Postupná stavba ligandu uvnitř receptoru
- Monte Carlo algoritmus
- Evoluční algoritmy
- Předgenerované knihovny konformací
- Molekulová dynamika



Dokování – používané algoritmy

Manuální dokování

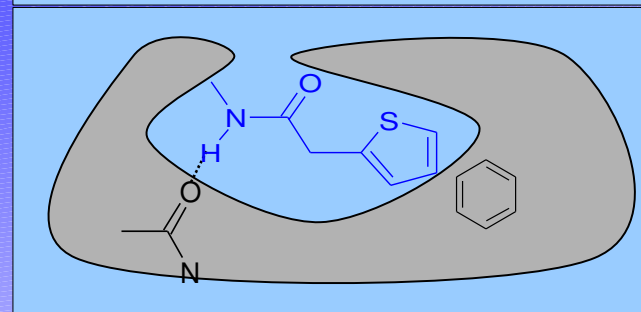
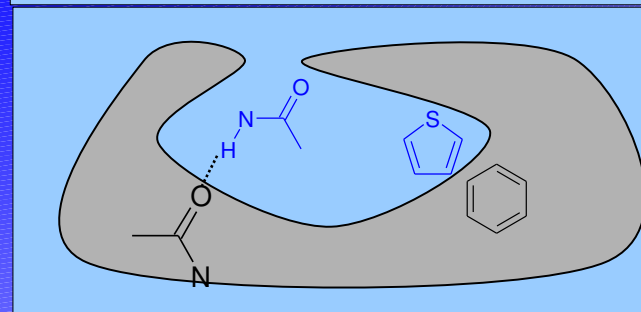
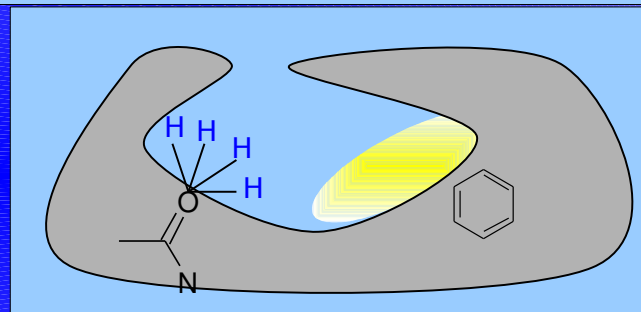
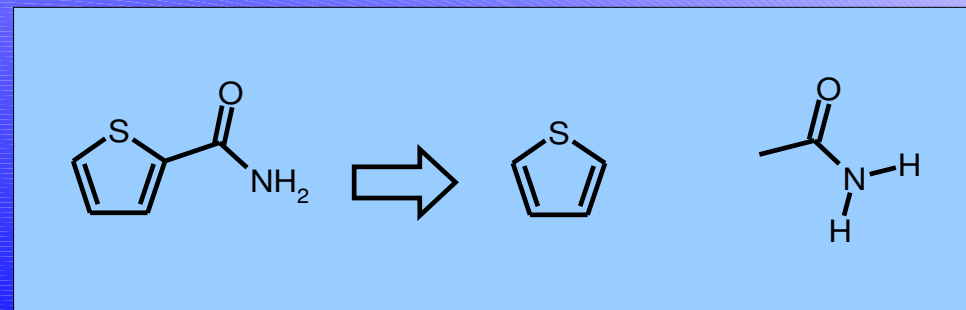
- Interaktivní dokování s využitím předpočítaného gridu pro vdw a elektrostatické interakce.
- Využívá zkušeností člověka a podobnosti se známými ligandy pro konstrukci komplexu.
- Nevýhodou je jednoduché silové pole a také “zkušenost” člověka, který provádí experiment.



Dokování – používané algoritmy

Postupná konstrukce ligandu

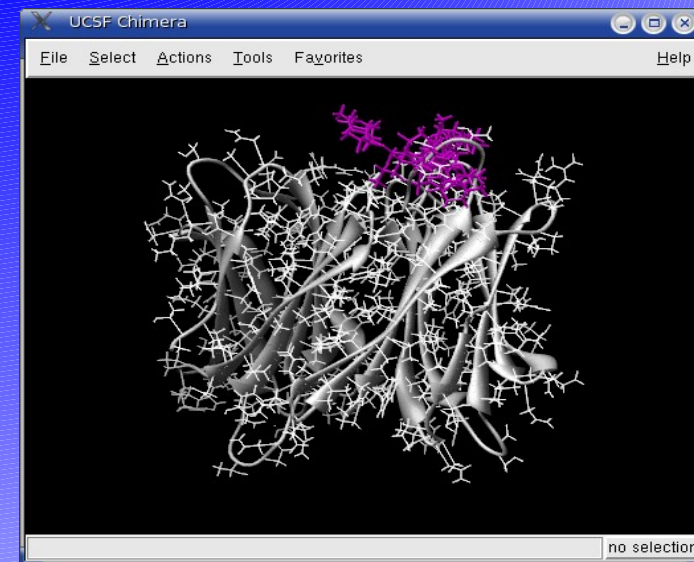
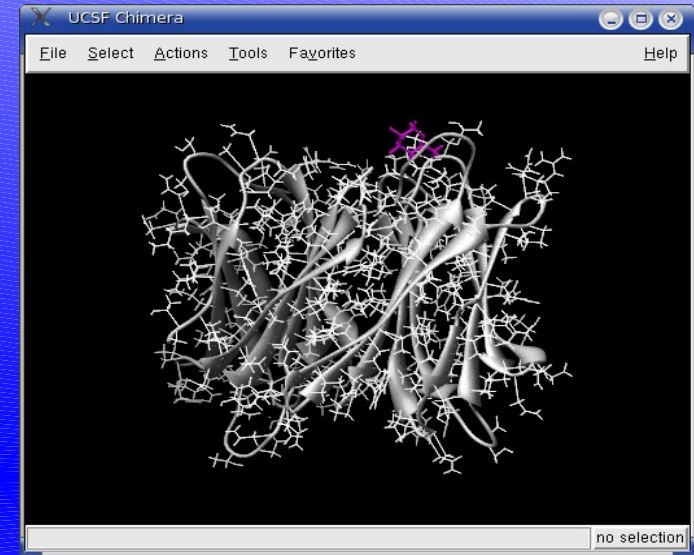
- Ligand je rozdělen na několik rigidních částí, které jsou spojené flexibilními můstky.
- Rigidní části jsou postupně vkládány do receptoru a spojovány.
- Uznávají se pouze ty konstrukce ligandu, které odpovídají vazebným podmínkám pro ligand.



Dokování – používané algoritmy

Monte Carlo

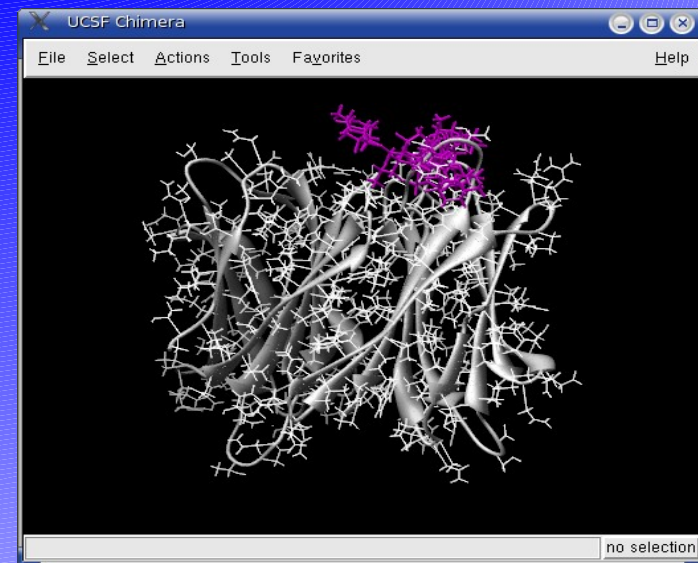
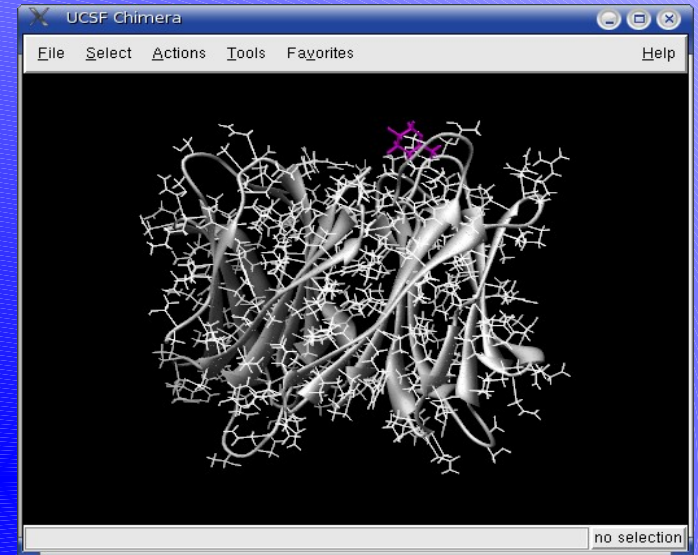
- Ligand je vkládán jako celá molekula.
- Pozice ligandu uvnitř receptoru i konformační změny jsou prováděny náhodně.
- Je třeba provádět velké množství orientací při různých simulačních teplotách, aby bylo nalezeno globální minimum.



Dokování – používané algoritmy

Genetický algoritmus

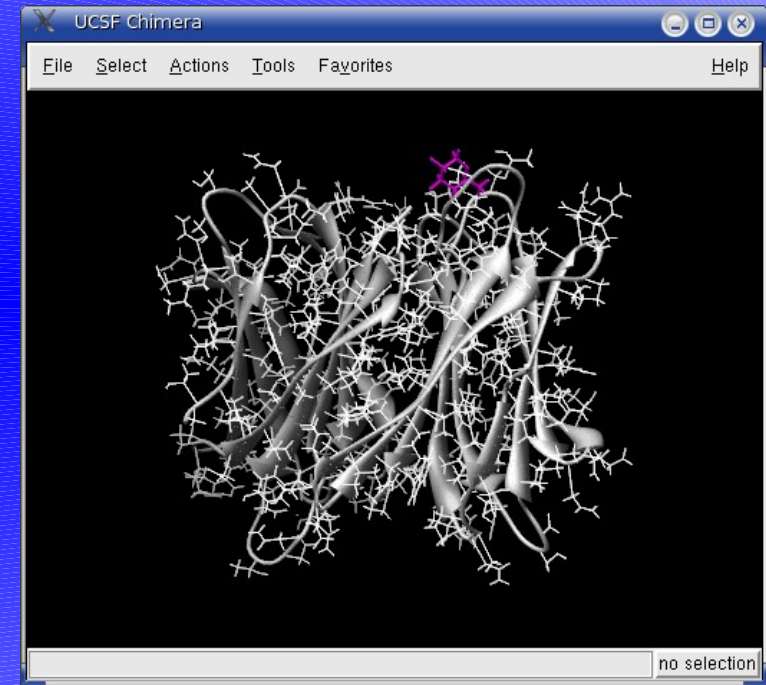
- Ligand je vkládán jako celá molekula.
- Pozice ligandu uvnitř receptoru i konformační změny jsou zapsány pomocí genu a jsou energeticky ohodnoceny.
- Náhodným křížením genů a náhodnými mutacemi je prohledáván konformační prostor.



Dokování – používané algoritmy

Předgenerované knihovny konformací

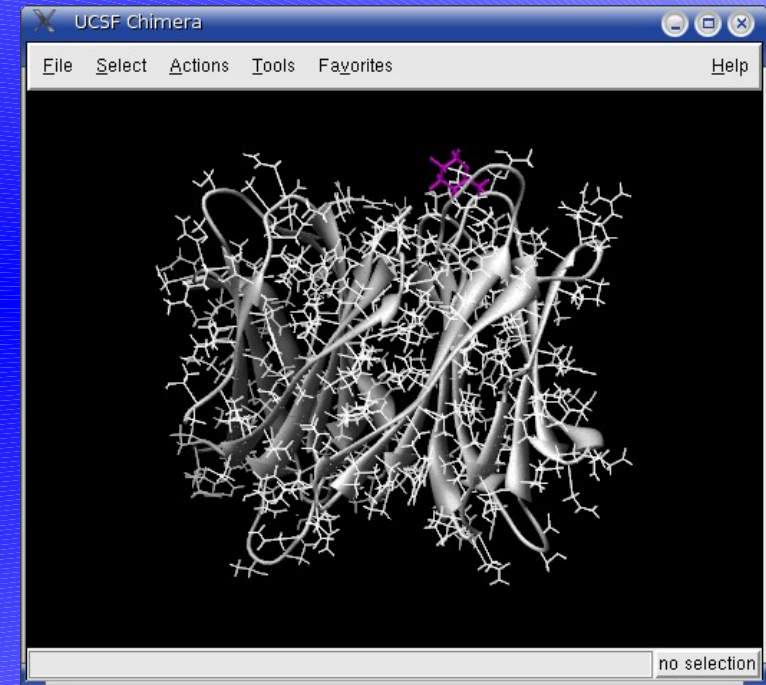
- Ligand je vkládán jako celá molekula.
- Nejdříve je provedeno konformační zmapování pro ligand a vhodné konformace jsou dále použity pro dokování.
- Většinou je použito systematické prohledávání pro jednotlivé konformace.



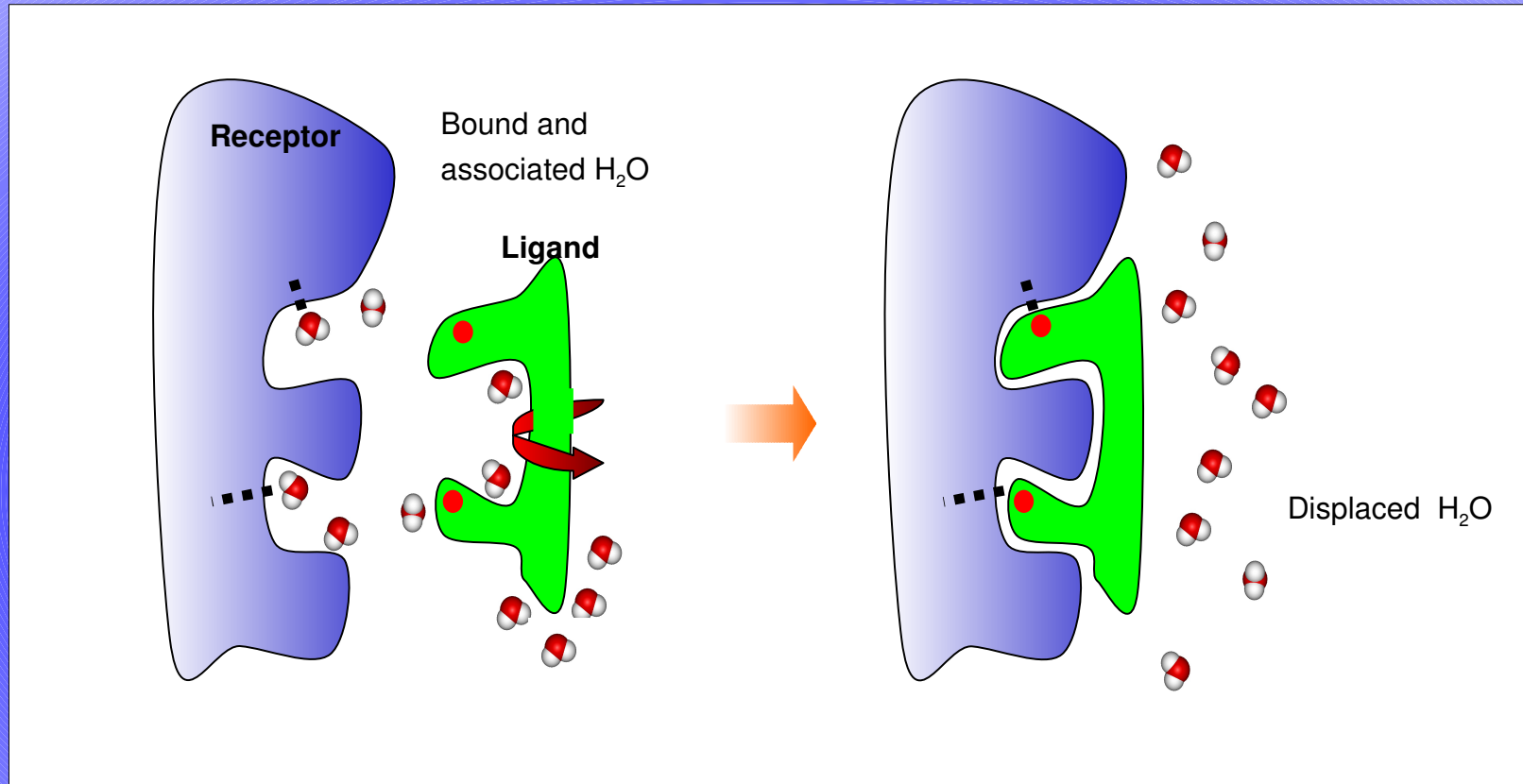
Dokování – používané algoritmy

Molekulová dynamika

- Molekulová dynamika zahrnuje do simulací flexibilitu receptoru i ligandu
- Použití MD výrazně prodlužuje dobu dokovacího experimentu.
- MD ze své definice není schopna překonávat velké energetické bariery.



Co se děje při navázání ligandu do receptoru?



- Dochází ke změně solvatace proteinu i ligandu (přemístění molekul vody).
- Změní se vibrační módy molekul.
- Zmenšuje se entropie systému.

Dokování – skórovací funkce

- Co je úkolem skórovacích funkcí?
 - Správně optimalizovat polohu ligandu uvnitř receptoru, provádět minimalizace systému.
 - Pro řadu ligandu správně určit jejich vazebnou afinitu.
- Co musí umět skórovací funkce?
 - Nalézt správnou polohu ligandu uvnitř receptoru (podle experimentálních dat).
 - Správně ohodnotit skupinu ligandů.
 - Vybrat vhodné a nevhodné ligandy z databáze.
 - Musí být rychlé a odolné chybám v systému (výpočetně stabilní).

Skórovací funkce - přehled

- Funkce, které používají silová pole (first-principles methods) – většinou využívají vdw a elektrostatické termy a solvent je zahrnut do dielektrické konstanty.
- Semiempirické funkce – používají kombinaci předchozích s např. PBSA/GBSA metody pro započítání solvatačního efektu.
- Empirické metody – zahrnují empirické funkce jako např. Hammettovy substituční koeficienty, nebo Hanchovy rozdělovací koeficienty. Dokáží popsat i pH systému.
- PMF metody – (potentials of mean forces) – běžně jsou používány pro metody protein folding, ale nacházejí uplatnění i jako skórovací fce v dokování.

Použití dockovacích algoritmů

- Určení vazebných modů ligandu v proteinu.
- Optimalizace molekul ligandu pro větší selektivitu a afinitu k receptoru
- Prohledávání databází struktur a hledání kandidátů na léčiva

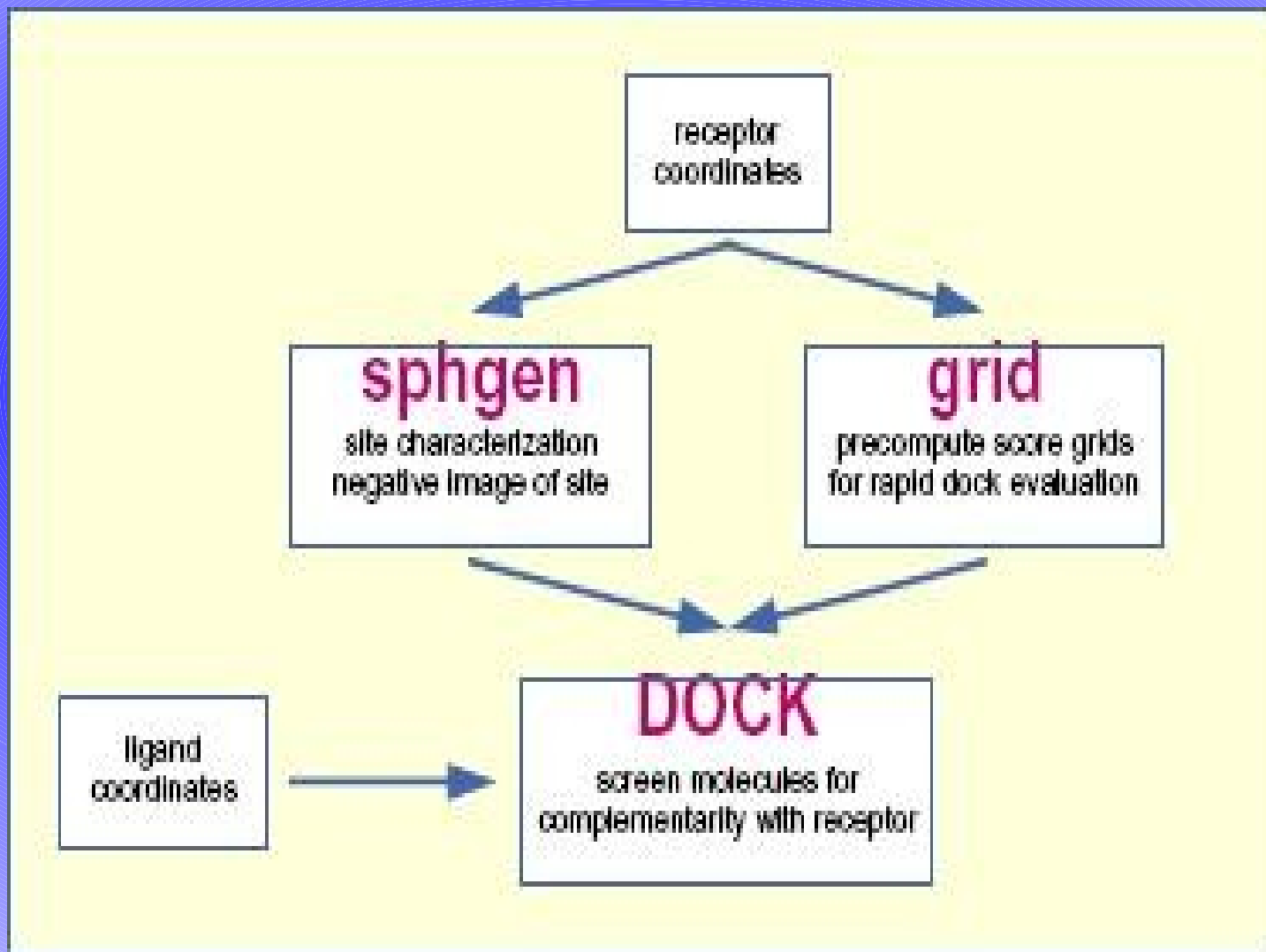
Program DOCK

<http://dock.compbio.ucsf.edu/>

The screenshot shows a Mozilla Firefox browser window displaying the UCSF DOCK website. The browser's address bar shows the URL <http://dock.compbio.ucsf.edu/>. The website header includes the UCSF logo and the text "The Official UCSF DOCK Web-site". A navigation menu on the left contains the following links: [Kuntz Group](#), [Shoichet Group](#), [Overview of DOCK](#), [DOCK 6](#), [Test Sets](#), [Contributed Code](#), [DOCK Related Links](#), [Older Versions of DOCK](#), [Other Useful Links](#), [The MDI](#), and [Contact us](#). The main content area features a 3D molecular model of a protein-ligand complex, with the protein shown as a green surface and the ligand as a stick model. Below the model, a red announcement reads: **NEW and IMPROVED:**
Release of DOCK 6.0!

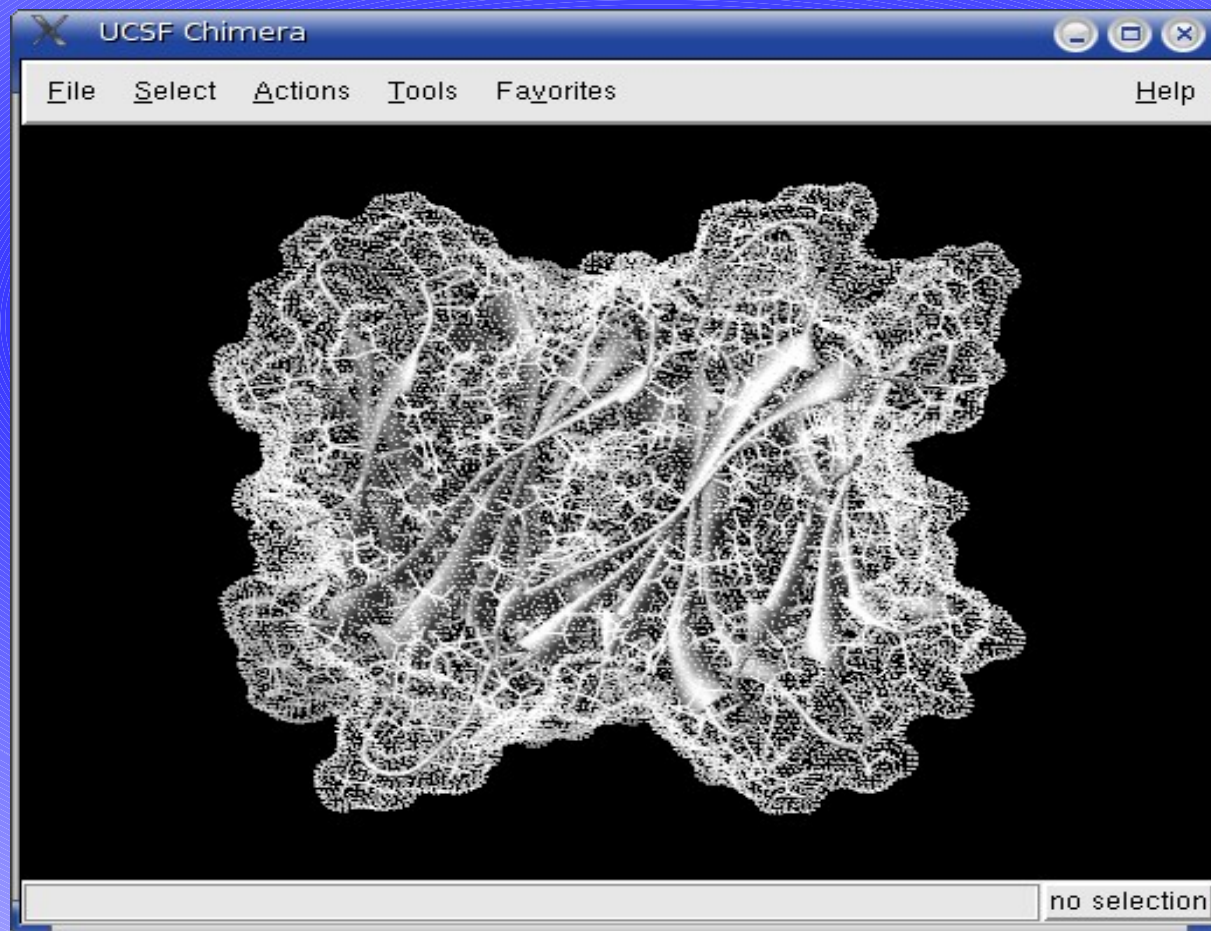
At the bottom of the browser window, the status bar shows the URL http://dock.compbio.ucsf.edu/DOCK_6/index.htm.

Program DOCK



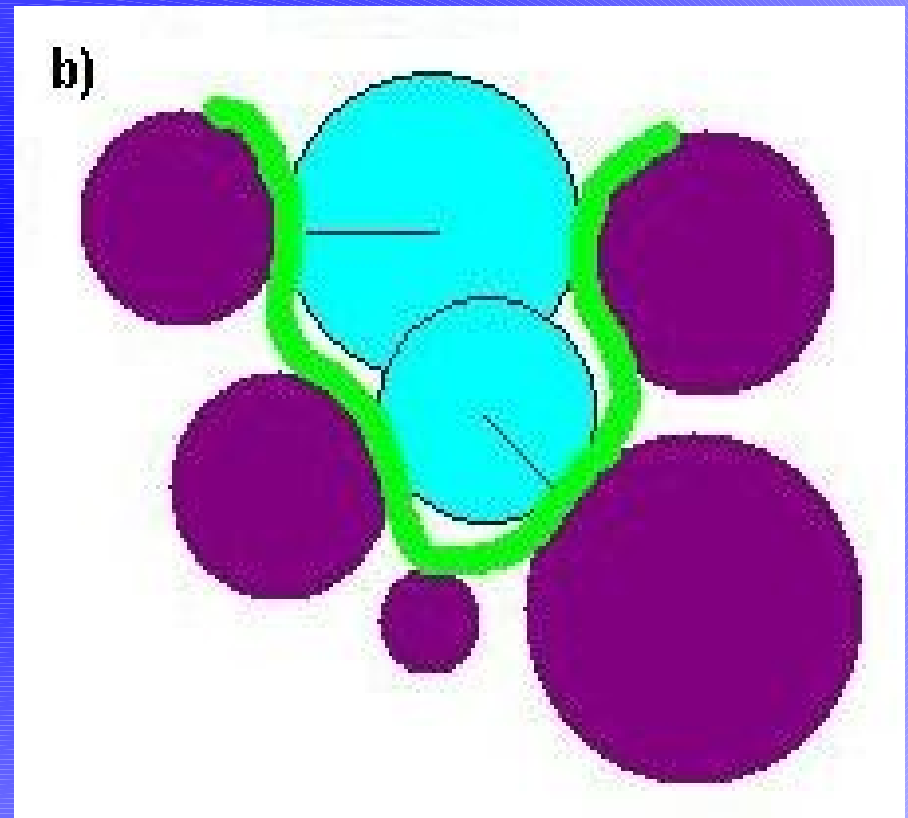
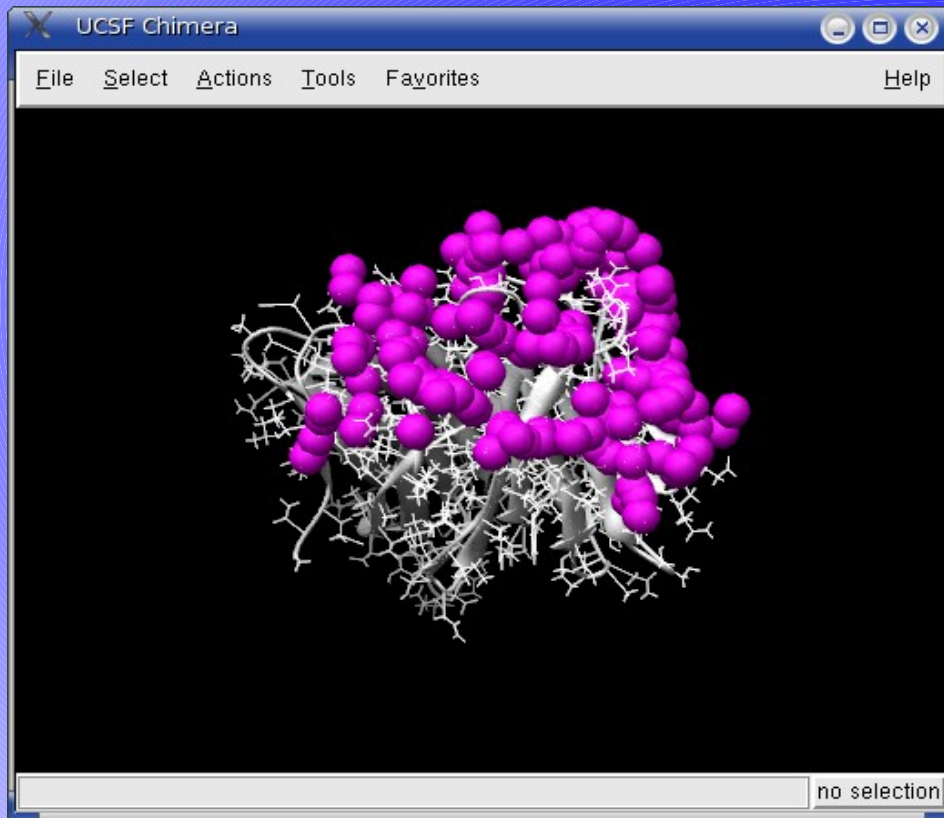
Program DOCK

Generování povrchu molekuly – program dms



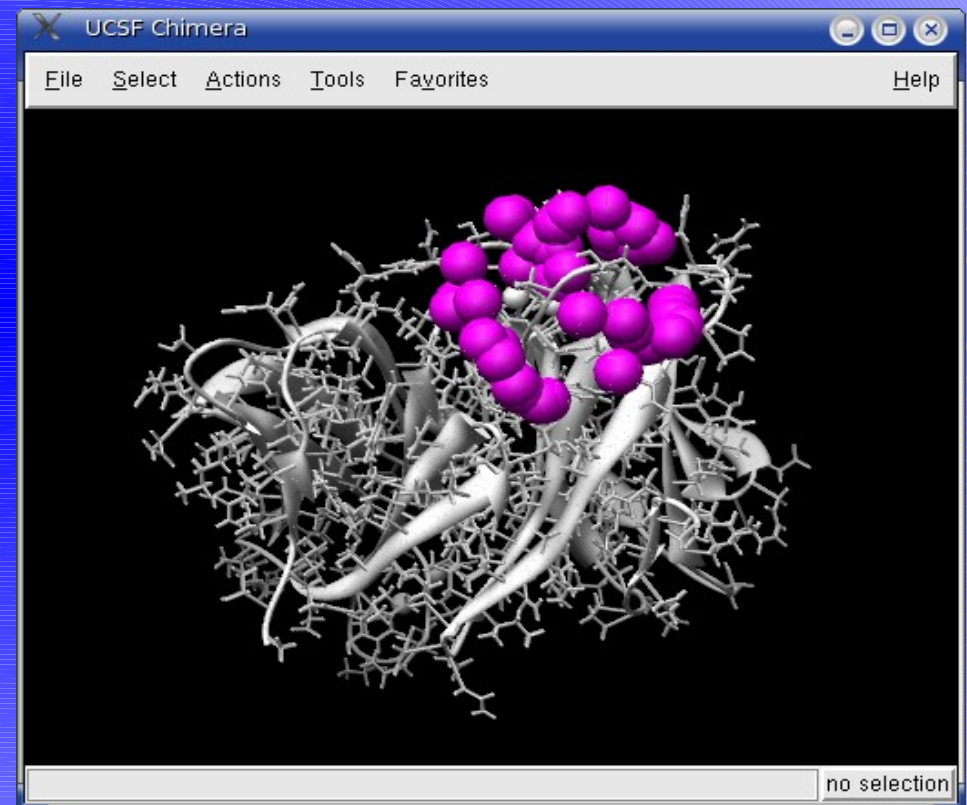
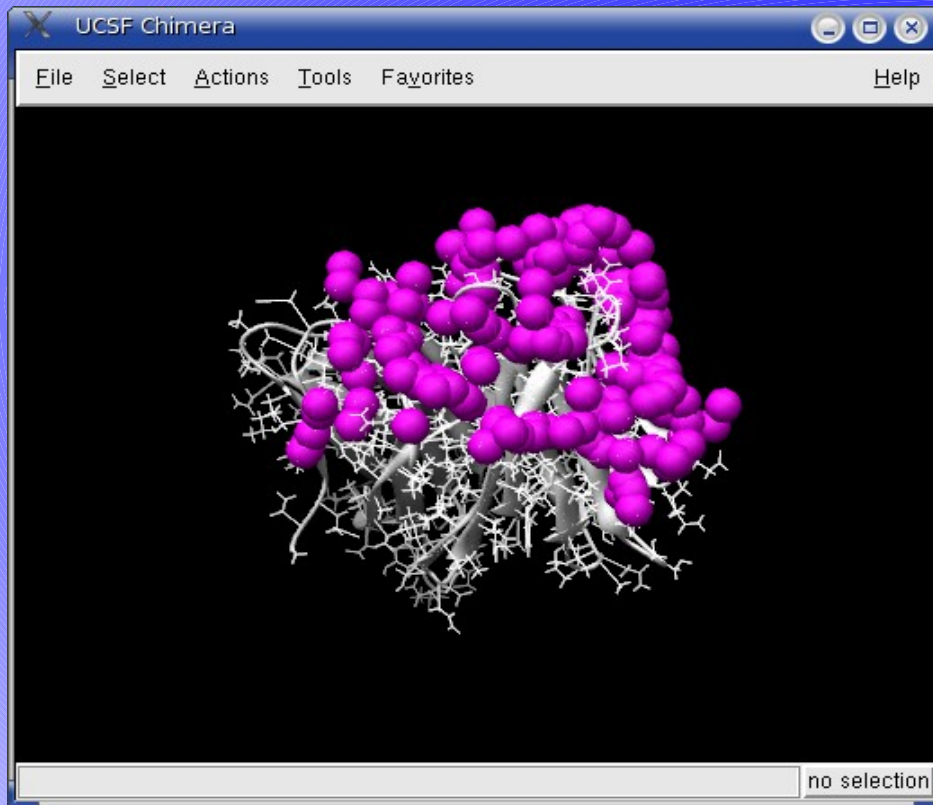
Program DOCK

Generování sfér a hledání kavit v proteinu – program sphgen



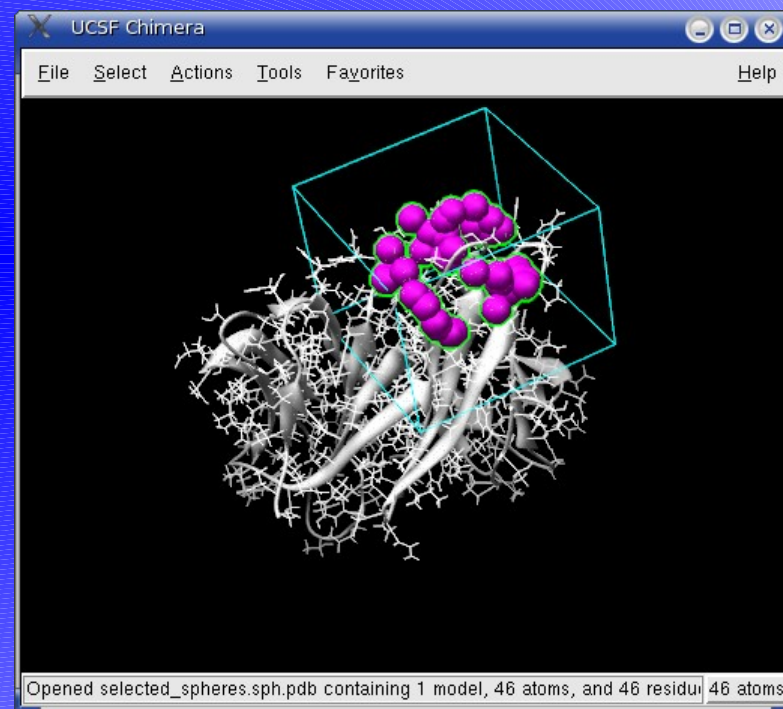
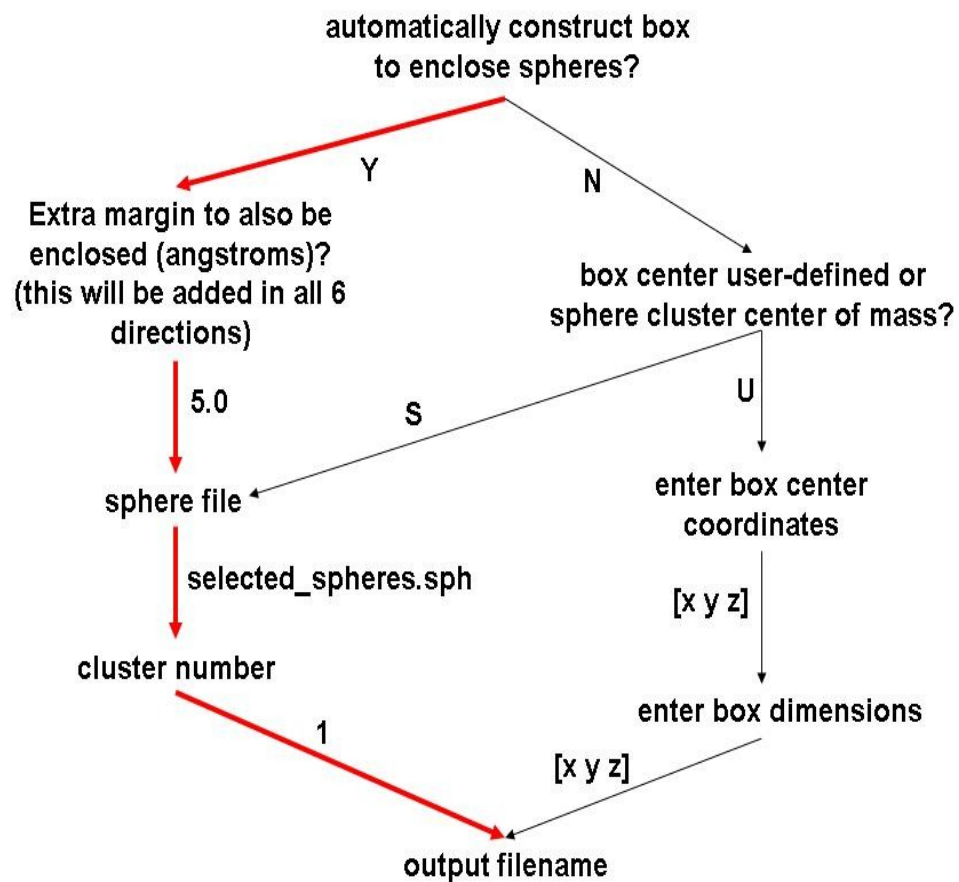
Program DOCK

Výběr sfér – výběr oblasti prohledávání – program sphere_select



Program DOCK

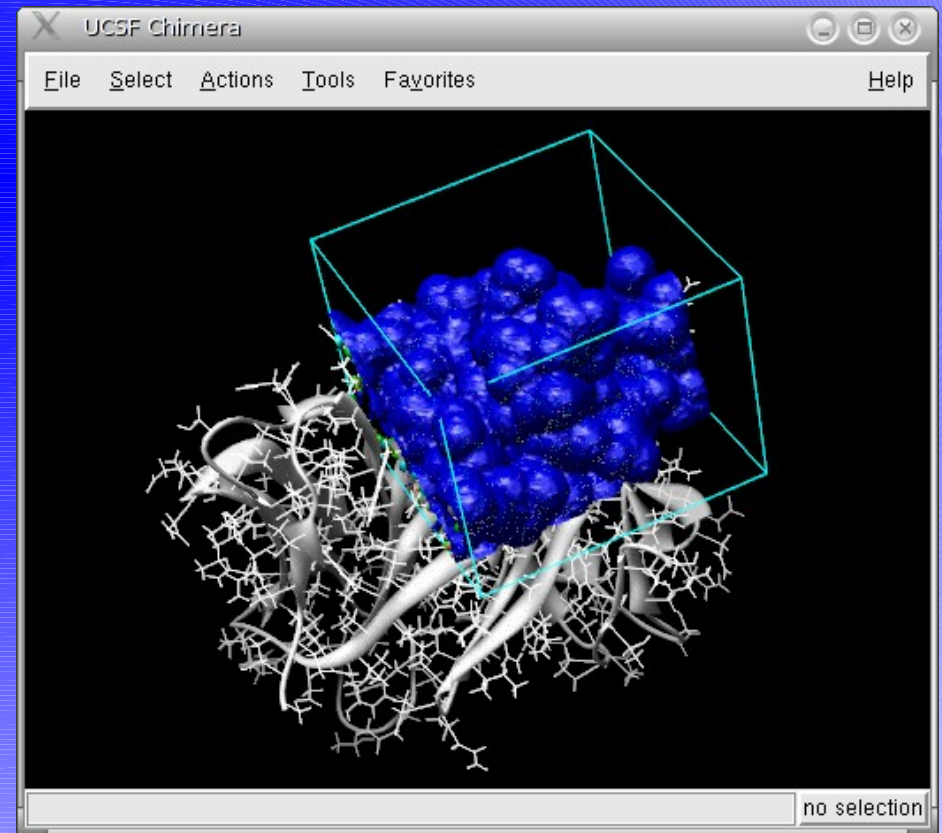
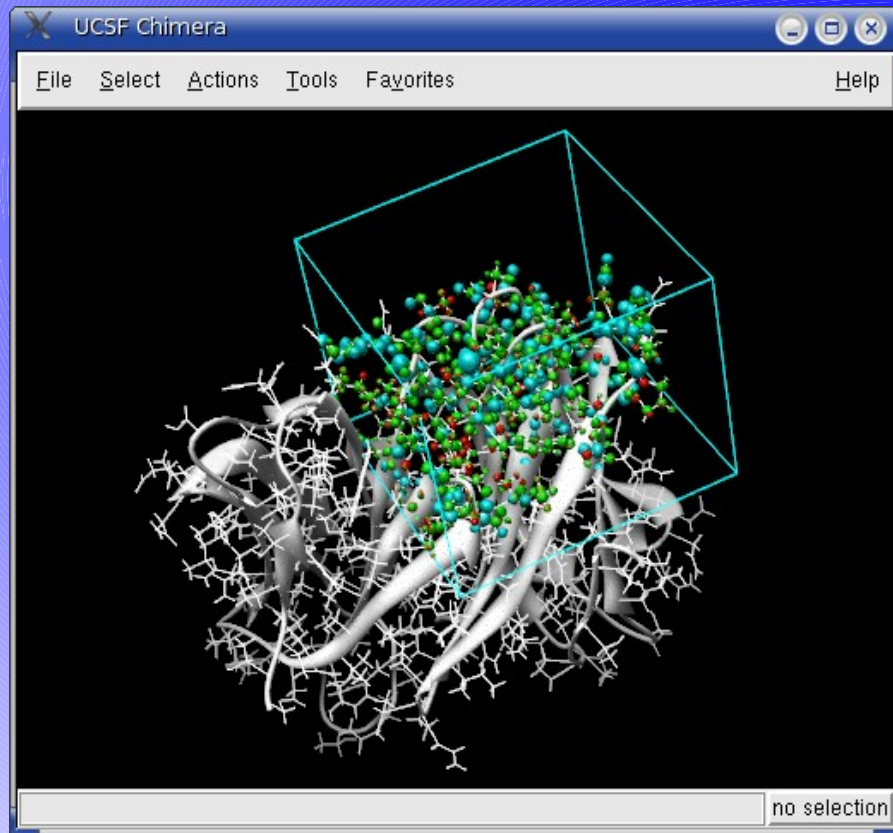
Tvorba simulačního boxu – program show_box



Program DOCK

Počítání gridu pro interakce – program **grid**

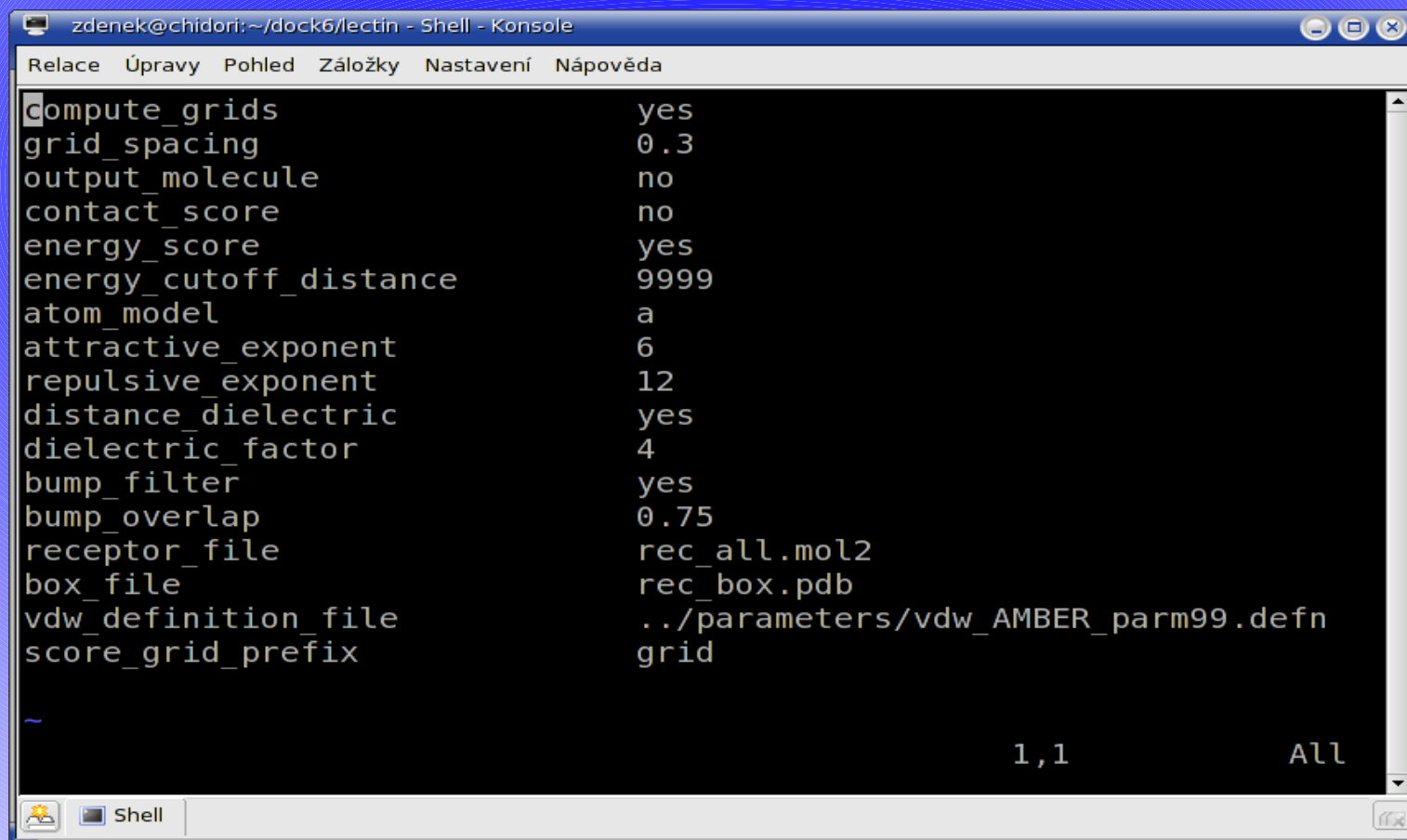
$$E = \sum_{i=1}^{lig} \sum_{j=1}^{rec} \left(\frac{A_{ij}}{r_{ij}^a} - \frac{B_{ij}}{r_{ij}^b} + 332 \frac{q_i q_j}{Dr_{ij}} \right)$$



Program DOCK

Počítání gridu pro interakce – program grid

$$E = \sum_{i=1}^{lig} \sum_{j=1}^{rec} \left(\frac{A_{ij}}{r_{ij}^a} - \frac{B_{ij}}{r_{ij}^b} + 332 \frac{q_i q_j}{Dr_{ij}} \right)$$



The screenshot shows a terminal window titled "zdenek@chidori:~/dock6/lectin - Shell - Konsole". The window contains a list of configuration parameters for the 'grid' program, each followed by its value. The parameters and their values are:

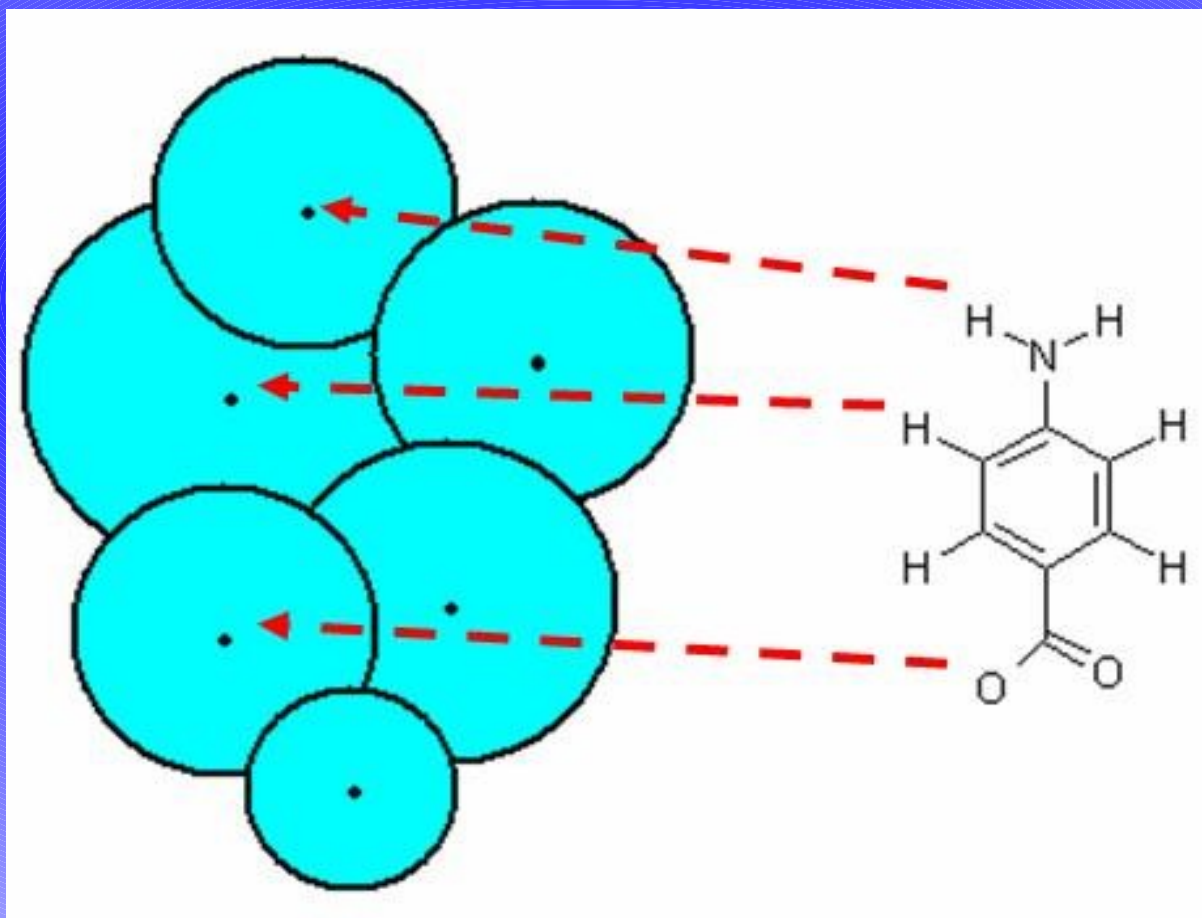
compute_grids	yes
grid_spacing	0.3
output_molecule	no
contact_score	no
energy_score	yes
energy_cutoff_distance	9999
atom_model	a
attractive_exponent	6
repulsive_exponent	12
distance_dielectric	yes
dielectric_factor	4
bump_filter	yes
bump_overlap	0.75
receptor_file	rec_all.mol2
box_file	rec_box.pdb
vdw_definition_file	../parameters/vdw_AMBER_parm99.defn
score_grid_prefix	grid

At the bottom right of the terminal, the coordinates "1,1" and the text "All" are visible.

Program DOCK

Vlastní dokování – program dock6

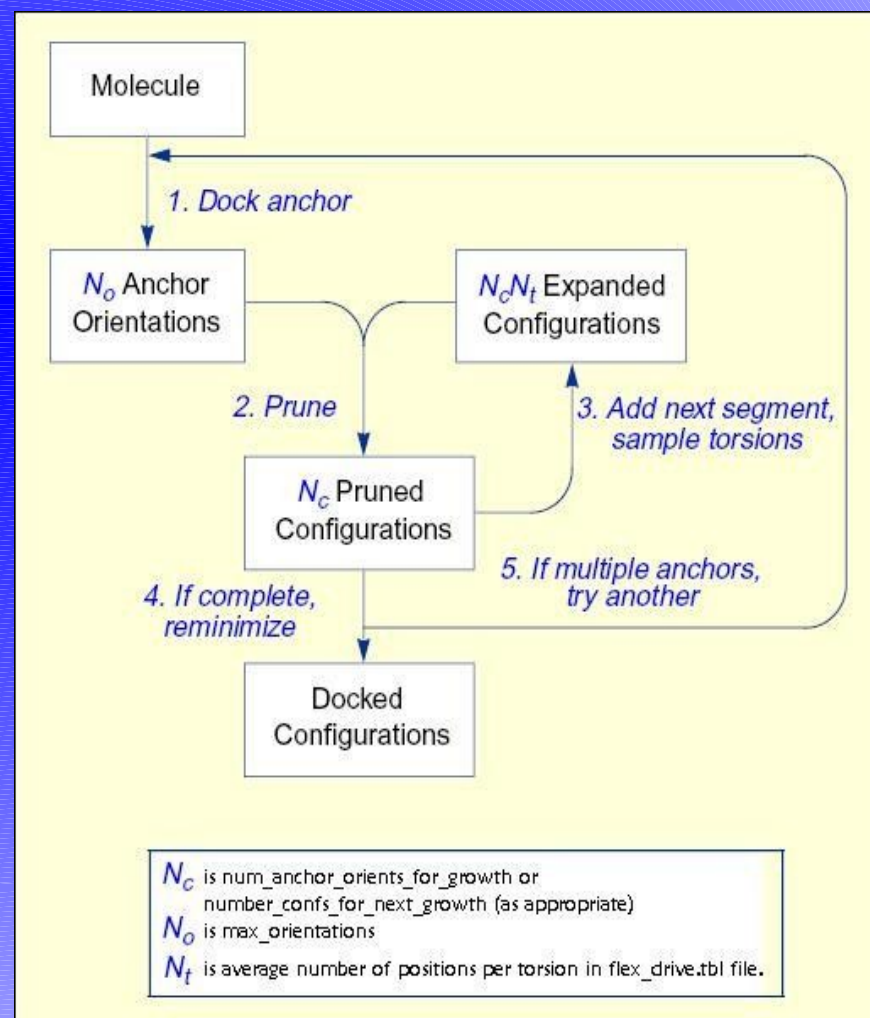
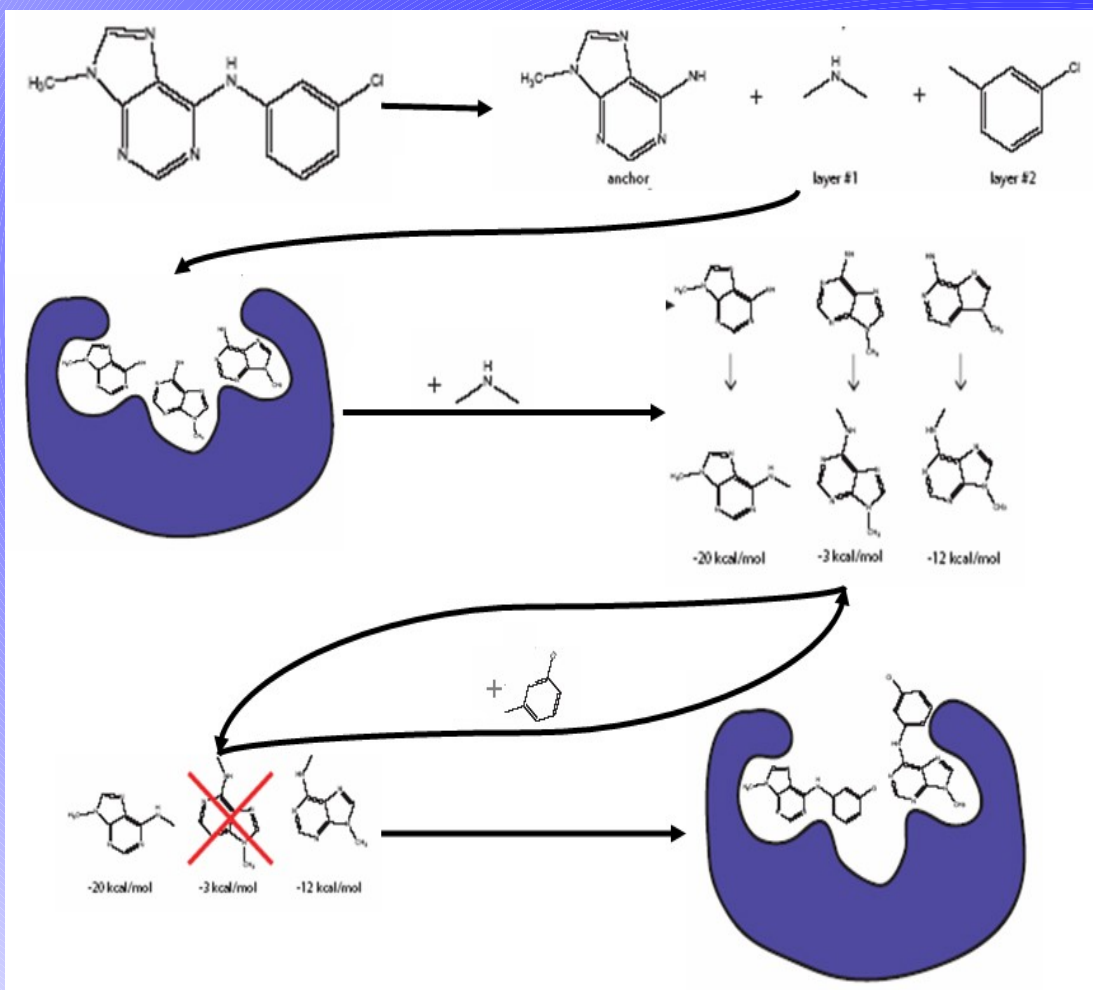
Rigidní dokování



Program DOCK

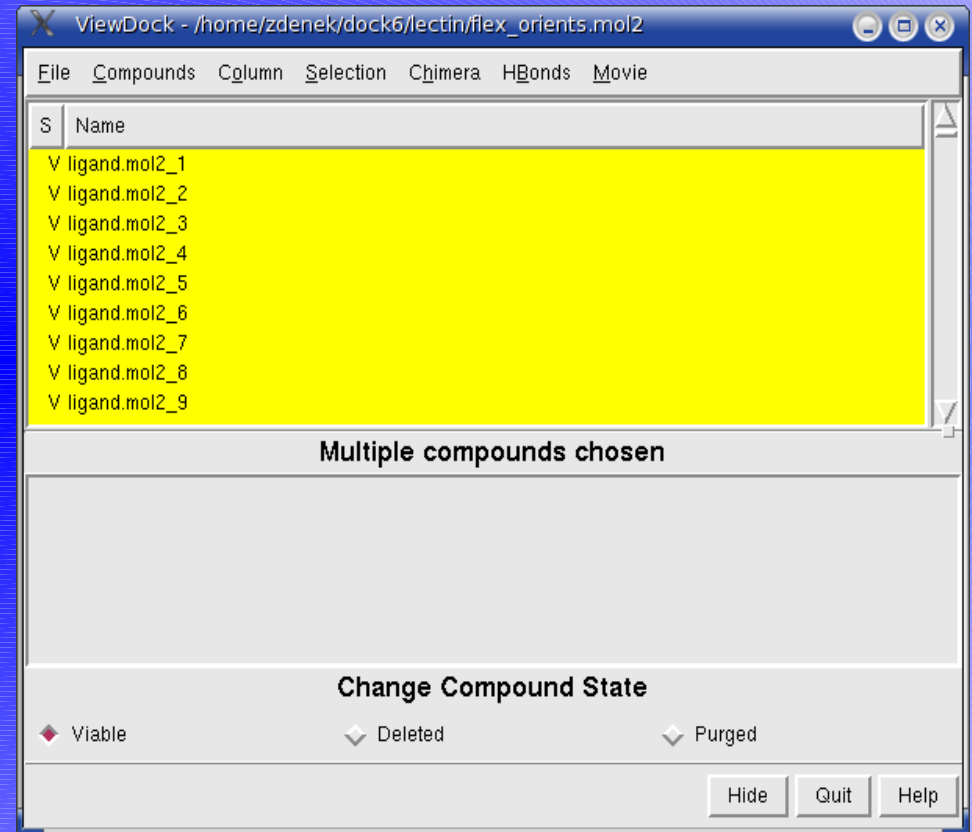
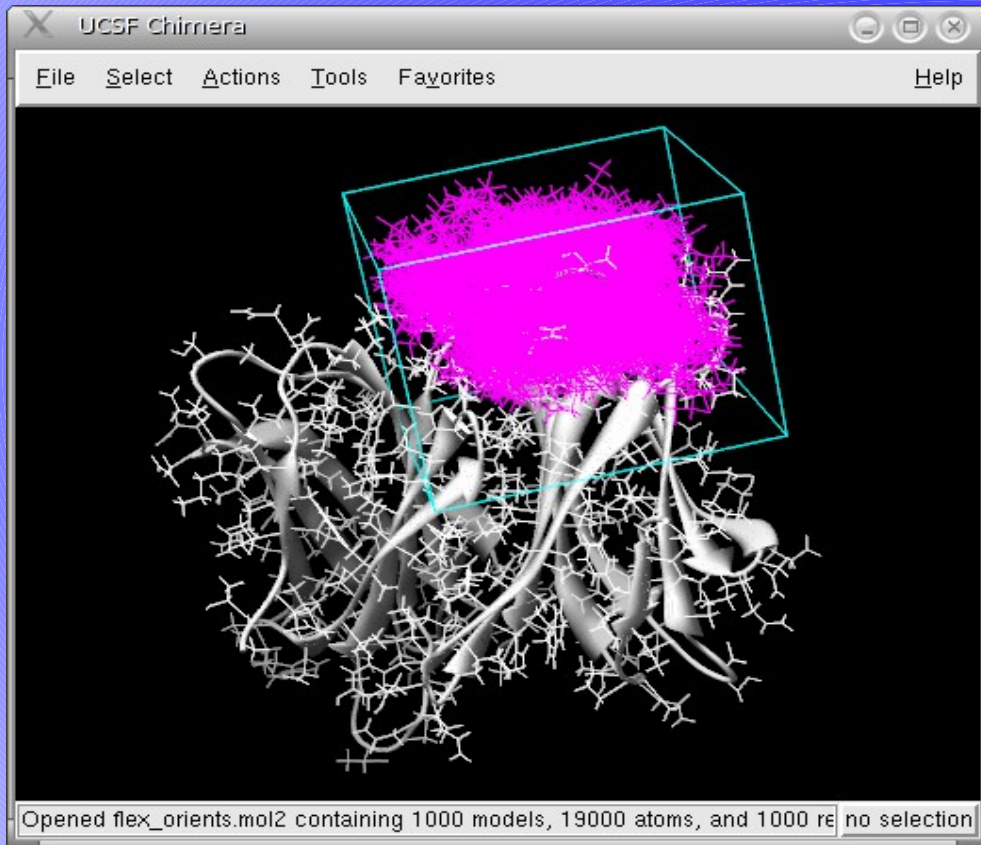
Vlastní dokování – program dock6

Flexibilní dokování



Program DOCK

Výsledky dokování a jejich interpretace



ViewDock - /home/zdenek/dock6/lectin/flex_orients.mol2

File Compounds Column Selection Chimera HBonds Movie

S	Name
V	ligand.mol2_1
V	ligand.mol2_2
V	ligand.mol2_3
V	ligand.mol2_4
V	ligand.mol2_5
V	ligand.mol2_6
V	ligand.mol2_7
V	ligand.mol2_8
V	ligand.mol2_9

Multiple compounds chosen

Change Compound State

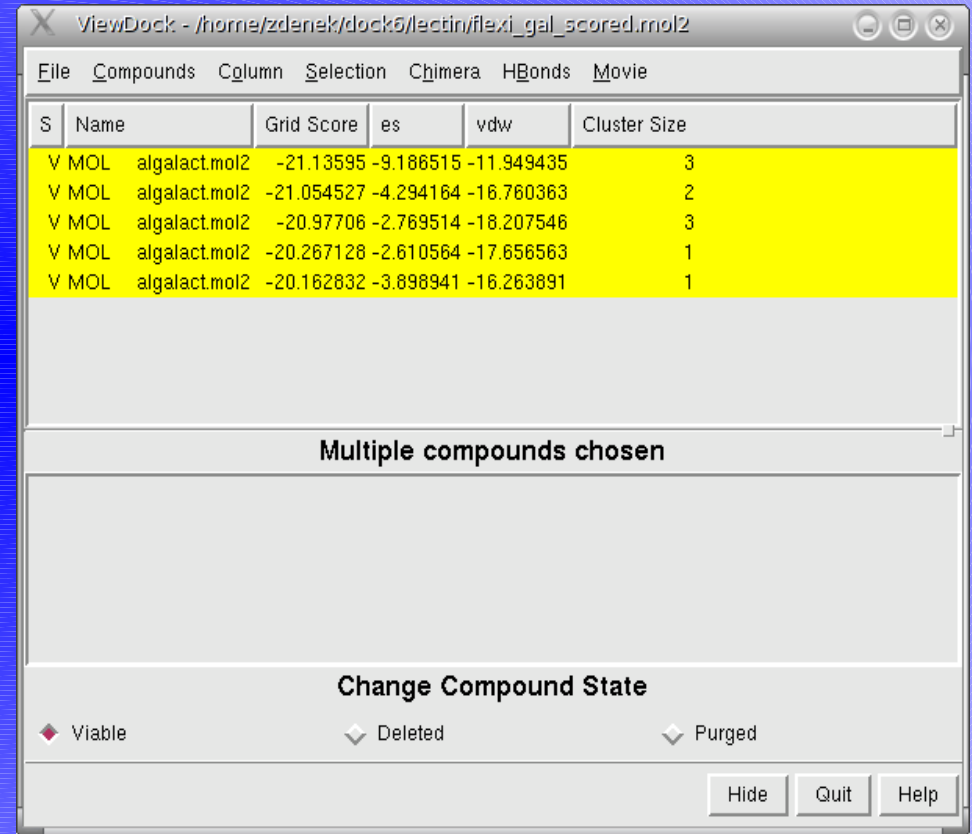
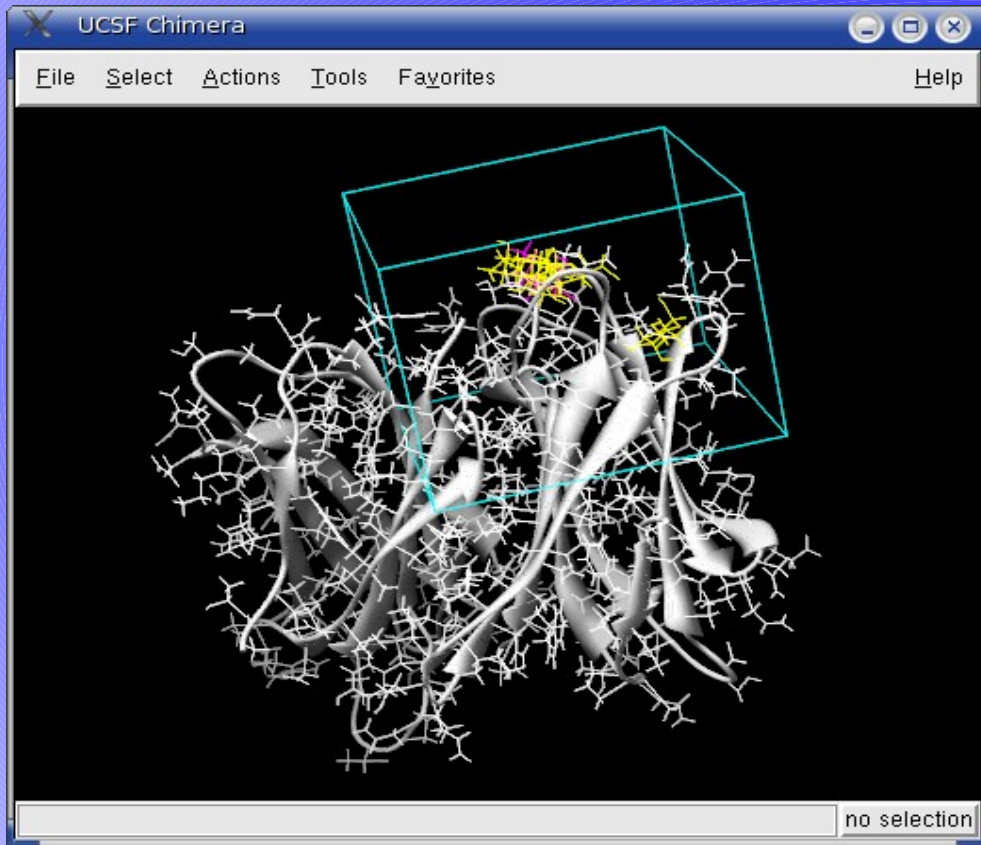
Viable Deleted Purged

Hide Quit Help

The image shows the ViewDock interface. It features a table with columns 'S' and 'Name'. The table lists nine ligand models, each with a 'V' in the 'S' column. Below the table, there are three radio buttons for 'Change Compound State': 'Viable' (selected), 'Deleted', and 'Purged'. At the bottom right, there are three buttons: 'Hide', 'Quit', and 'Help'.

Program DOCK

Výsledky dokování a jejich interpretace



ViewDock - /home/zdenek/dock6/lectin/lexi_gal_scored.mol2

File Compounds Column Selection Chimera HBonds Movie

S	Name	Grid Score	es	vdw	Cluster Size
V MOL	algalact.mol2	-21.13595	-9.186515	-11.949435	3
V MOL	algalact.mol2	-21.054527	-4.294164	-16.760363	2
V MOL	algalact.mol2	-20.97706	-2.769514	-18.207546	3
V MOL	algalact.mol2	-20.267126	-2.610564	-17.656563	1
V MOL	algalact.mol2	-20.162632	-3.898941	-16.263891	1

Multiple compounds chosen

Change Compound State

Viable Deleted Purged

Hide Quit Help

The image shows the ViewDock software interface. It displays a table of docking results for the molecule 'algalact.mol2'. The table has six columns: S, Name, Grid Score, es, vdw, and Cluster Size. Five rows of results are shown, with the first row highlighted in yellow. Below the table, there is a section for 'Change Compound State' with three radio buttons: 'Viable' (selected), 'Deleted', and 'Purged'. At the bottom right, there are three buttons: 'Hide', 'Quit', and 'Help'.

Další dokovací programy

- AutoDock
- RosettaDock
- FlexX
- FTDock
- FlexiDock
- Gold
- ICM-Dock

- CICADA

Děkuji za pozornost